Étude des ressources génétiques du noyer en vue de la mise en œuvre d'une sélection assistée par marqueurs

Anthony Bernard

▶ **To cite this version:**

Anthony Bernard. Étude des ressources génétiques du noyer en vue de la mise en œuvre d'une sélection assistée par marqueurs. Biologie végétale. Université de Bordeaux, 2020. Français. NNT : 2020BORD0128 . tel-03118714

HAL Id: tel-03118714
https://theses.hal.science/tel-03118714

Submitted on 22 Jan 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE PRESENTÉE

POUR OBTENIR LE GRADE DE

DOCTEUR DE L'UNIVERSITÉ DE BORDEAUX

École Doctorale : Sciences de la Vie et de la Santé

Mention : Sciences, Technologie, Santé

Spécialité : Biologie végétale

Anthony BERNARD

Sous la direction d'Elisabeth DIRLEWANGER

Étude des ressources génétiques *Juglans* en vue de la mise en place d'une sélection assistée par marqueurs

Soutenue le 28 septembre 2020

Président : Michel Hernould, Professeur, Université de Bordeaux, UMR BFP, Bordeaux

Rapporteur·e·s :
Evelyne COSTES, Directrice de Recherche, UMR AGAP, Montpellier
Didier PELTIER, Professeur, Université d'Angers, IRHS, Angers

Examinateurs·trices :
Laurent BOUFFIER, Chargé de Recherche, UMR BIOGECO, Cestas-Pierroton
Valérie SCHURDI-LEVRAUD, Maître de Conférences, Université de Bordeaux, UMR BFP, Bordeaux
Vincent SEGURA, Chargé de Recherche, UMR AGAP, Montpellier

Invité·e·s :
Fabrice LHEUREUX, Ingénieur Responsable Filière Noix, CTIFL, Centre Opérationnel de Lanxade, Prigonrieux
Elisa MARGUERIT, Maître de Conférences, Bordeaux Sciences Agro, UMR EGFV, Bordeaux

« Carpalim d'une coquille de noix grosliere faisoit un beau, petit, joyeulx et harmonieux moulinet à aesle de quatre belles petites aisses d'un tranchouoir de vergne. »

François Rabelais, *Le Quart Livre des faicts et dicts héroïques du bon Pantagruel*, 1552

« Lait et poisson est poison. Et après le poisson noix est contrepoison. »

Laurent Joubert, *Erreurs populaires*, 1600

« Une noix fait grand bien, Deux c'est trop, Trois c'est la mort. »

Garnerius, 1612

« Pour teindre les cheveux en noir, prenez jus d'écorces de noix vertes, une livre, poudre de litharge, trois onces, meslez le tout avec lessive commune et en lavez les cheveux.
De cettuy, les femmes limosines et les hommes aussi s'en servent tous les jours. »

Guyon, *Cours de Médecine en français*, tome II, 6ème édition, 1673

« Les teinturiers se servent de la racine, de l'écorce, de la feuille & du brou des noix pour teindre les étoffes en fauve, en café & en couleur de noisette.
Ils emploient à cette fin la racine avant que l'arbre soit en sève, l'écorce lorsque la sève est en mouvement, les feuilles lorsque les noix sont à demi formées, & le brou dans le temps des cerneaux. »

Le Grand Vocabulaire françois, 1771

« Ces beaux noyers qui m'apportaient je ne sais quel contentement d'âme.
Comme ils rendaient la cour du presbytère agréable et hospitalière !
Que leurs rameaux étaient frais et magnifiques ! »

Johann Wolfgang von Goethe, *Les Souffrances du jeune Werther*, 1774

« Le tabac blanc ouvre à la nuit ses tubes étroits de parfum et ses corolles en étoile.
Un rayon, en touchant le noyer, l'éveille : il clapote, remué jusqu'aux basses branches par une mince rame de lune.
Le vent superpose, à l'odeur du tabac blanc, l'odeur amère et froide des petites noix véreuses qui choient sur le gazon. »

Colette, *La Maison de Claudine*, 1922

« La noix.
Ce serait un soir de décembre,
De la neige avec un vent froid,
Nous serions au fond d'une chambre,
Et nous grignoterions des noix. »

Louis Calaferte, *Fruits*, éd. Hesse, 1992

« Il avait atteint les grands arbres, […] les plus beaux étaient deux noyers : il se souvint des statues de la bibliothèque. La plénitude des arbres séculaires émanait de leur masse, mais l'effort par quoi sortaient de leurs énormes troncs les branches tordues, l'épanouissement en feuilles sombres de ce bois, si vieux et si lourd qu'il semblait s'enfoncer dans la terre et non s'en arracher, imposaient à la fois l'idée d'une volonté et d'une métamorphose sans fin. Entre eux, les collines dévalaient jusqu'au Rhin ; ils encadraient la cathédrale de Strasbourg très loin dans le crépuscule heureux, comme tant d'autres troncs encadraient d'autres cathédrales dans des champs d'Occident.

Et cette tour dressée dans son oraison d'amputé, toute la patience et le travail humain développés en vagues de vignes jusqu'au fleuve n'étaient qu'un décor du soir autour de la séculaire poussée du bois vivant, des deux jets drus et noueux qui arrachaient les forces de la terre pour les déployer en ramures.

Le soleil très bas poussait leur ombre jusqu'à l'autre côté de la vallée, comme deux épais rayons.

Mon père pensait aux deux saints, à l'Atlante ; le bois convulsé de ces noyers, au lieu de supporter leur fardeau du monde, s'épanouissait en une vie éternelle en leurs feuilles vernies sur le ciel et leurs noix presque mûres, en toute leur masse solennelle au-dessus du large anneau des jeunes pousses et des noix mortes de l'hiver. »

André Malraux, *Les Noyers de l'Altenburg*, éd. Gallimard, 1996

Fruits de *Juglans nigra* L., le noyer noir, sur le site du centre INRAE Nouvelle-Aquitaine-Bordeaux

Remerciements

Mes ami·e·s nommé·e·s « docteur·e·s » avant moi m'ont fait vivre leurs thèses par procuration. J'ai pu constater le fait qu'ils·elles sont tou·te·s, sans exception, passé·e·s par de nombreux sentiments et états d'esprit au cours de ces trois années : beaucoup d'excitation, de passion, de dévouement, d'abnégation, de doutes, de remises en question, de stress intense et finalement, de joie et de satisfaction. Je savais donc à quoi m'attendre en me lançant à corps perdu dans ce projet et je dois dire que je n'ai pas fait figure d'exception. Mais je crois que la passion de la recherche scientifique et la satisfaction du travail accompli prennent le dessus sur tout le reste au moment venu de la rétrospective de ces trois années. Ce petit préambule s'adresse donc à quiconque souhaitant faire une thèse mais se disant « Ah, fais-je le bon choix ? Vais-je y arriver et tenir le rythme ? Et quid du stress de la rédaction et des articles ? ». Mes réponses à ces questions sont très simples : « Oui, oui et oui ! ». Vous ne regretterez rien et vous en sortirez plus confiant·e, grandi·e et certainement encore plus passionné·e…

Donc ce moment que j'attendais tant arrive enfin, celui du témoignage de ma gratitude envers toutes les personnes qui ont contribué de près ou de loin à ce que je travaille dans les meilleures conditions. Cet espace de libre expression est d'une importance primordiale pour moi. Ainsi, je tiens à adresser mes premiers et sincères remerciements à mes deux encadrant·e·s sans qui cette thèse n'aurait jamais vu le jour : **Elisabeth Dirlewanger** et **Fabrice Lheureux**.

Elisabeth, merci d'avoir appuyé ma candidature auprès de Fabrice aux prémices de la construction du projet. Ton appui dès l'instant où je t'ai soumis l'idée de postuler à ce sujet de thèse m'a donné grande confiance et je n'ai jamais eu envie de reculer. Merci pour ta grande disponibilité à chacun de mes questionnements. Tu as su me pousser dans le meilleur de moi-même, notamment grâce à ton idée de la rédaction de la review. Tu as été une directrice de thèse totalement dévouée pour tes étudiant·e·s thésard·e·s et ton dynamisme au quotidien a été une chance pour nous. Merci pour tout ce temps consacré à la lecture, relecture, re-relecture des articles et de ce manuscrit. Et ces « voyages » en Turquie et en Californie restent inoubliables ! Enfin, merci pour la confiance que tu m'as confiée depuis ces longues années, depuis mon arrivée en tant que stagiaire sur la cerise il y a huit ans déjà, où j'ai pu découvrir le monde de la recherche sur les espèces fruitières.

Fabrice, merci d'avoir accepté ma candidature au premier entretien. J'ai senti rapidement que vous m'accorderiez votre confiance et de l'autonomie dans la conduite de ce projet. Vous avez su porter ce projet d'une main de maître et je vous exprime aujourd'hui toute la reconnaissance que j'ai envers vous. Je n'aurais pas rêvé d'un meilleur sujet de thèse et j'ai ressenti chaque jour votre grande motivation pour ce sujet. Les questions qui incombent à l'avenir des filières des fruits à coque vous animent et votre implication n'a fait que me motiver encore plus en retour ! Je n'oublierai pas ces matinées sous la pluie à Toulenne et ces heures de correction de rapports de nos supers stagiaires !

Il faut néanmoins signaler qu'une thèse de ce type, en partenariat avec l'INRAE et le CTIFL, n'aura pas toujours été simple à mettre en œuvre… Nous avons tous les trois vécu des difficultés pour satisfaire certaines exigences mais je crois que nous avons relevé le défi. Un·e thésard·e Cifre doit savoir jouer parfois (souvent !) le rôle de médiateur·trice puisque les enjeux ne sont pas les mêmes de chaque côté. Néanmoins, je pense que le « contrat » est réussi et j'espère que cette thèse mènera à d'autres partenariats de ce genre concernant l'arboriculture fruitière ; elle en a grandement besoin dans ce contexte actuel de changement climatique ! Ainsi, j'en profite pour glisser ces quelques mots même si j'ai peu d'espoir qu'ils soient lus ou aient un impact un jour : « Activons-nous, tou·te·s ensemble, avant qu'il ne soit trop tard ! La France doit rester un pays acteur de la R&D pour nos filières agronomiques… Les producteurs·trices nous attendent, alors rassemblons nos forces, INRAE, CTIFL, et autres, et mettons-nous au travail ! »

Je souhaite remercier ensuite l'ensemble de l'équipe A3C et plus particulièrement **Teresa Barreneche**. Tu m'as beaucoup appris en termes de relativisation des résultats et tu m'as transmis ton savoir-faire dans ce passionnant domaine qu'est la génétique des populations. Tu as été très présente pour moi et en dehors de ton savoir-faire, j'ai également beaucoup apprécié ton savoir-être. Tu m'as beaucoup soutenu aux moments où j'en avais besoin. Tes anecdotes personnelles ont beaucoup égayé les pauses café et tu m'as donné envie d'aller découvrir le Mexique ! Merci également au super trinôme **Lydie Fouilhaux**, **Jacques Joly** et **José Quero-García**. Voir tou·te·s ces stagiaires « cerise » défiler chaque année m'a rappelé de bons souvenirs partagés avec vous. Un grand merci également à **Bénédicte Wenden** pour son talent dans la création de figures et le traitement de données sous « R ». Oui, c'est un talent à ce niveau-là ! Je remercie aussi **Mathieu Fouché** pour sa constante bonne humeur et son calme (impressionnant !) face à toute situation, **Laurent Richard** pour son bon café (remplacé d'ailleurs par une vulgaire machine expresso automatique, ces robots

sont partout !), **Xavier Lafon** pour son humour douteux, **Hélène Christmann** pour son implication auprès de l'ADAS (excellente matinée « jeux » pour les 50 ans de l'association !) et **Loïc Le Dantec** pour ses connaissances en bio-informatique. Un énorme merci à **Noémie Vimont**, super collègue de bureau qui m'a vite manqué à son départ (tu as fait une excellente thèse, bravo !) et à **Rémi Beauvieux**, avec qui j'ai projeté « d'aller élever des chèvres dans la montagne ! ». Enfin, merci à **Camille Branchereau**, dernière thésarde arrivée à A3C (toutes ces heures à papoter avec toi ont été extrêmement agréables et bonne chance pour ta dernière année !), **Armel Donkpegan**, post-doctorant vraiment doué en génomique et bio-informatique (tu m'as évité des heures de travail de transformation de fichiers et grâce à toi, j'ai décidé de suivre un MOOC sur « Python3 » !) et **Tina Ternjak**, notre thésarde slovène reine des prunes !

Je tiens à féliciter le magnifique travail réalisé par les personnes qui m'ont précédé à l'INRAE sur le noyer. Je félicite donc **Francis Delort** (et merci pour votre expertise à plusieurs reprises), **Alain Bonnet** et **Patrick Reynet** qui nous a malheureusement quittés en 2018. J'offre mon hommage à **Éric Germain**, certainement un « grand Monsieur » d'après celles et ceux qui vont ont connu. Sans lui, il n'y aurait pas eu de ressources génétiques et donc, pas de thèse…

Je remercie également **Marine Delmas** du CRB et l'équipe de l'UEA de Toulenne avec **Sylvie Carrié**, **Michel Corradi**, **Marie-Laure Greil**, **Martine Saint-Marc**, **David Alletru**, **Julien Parmentier**, **Jean-Yves Chevrollier** et **François Gazel**. Merci **Thibault Théodore** pour ton implication dans la construction de la base de données pour les ressources génétiques

Ensuite, j'adresse mes sincères remerciements à tout le personnel du CTIFL. Merci à **Aude Moronvalle** pour son expertise en pathologie, à **Julie Castaing** et **Marie-Lisa Brachet** pour leur soutien et leur aide en biologie moléculaire, à **Véronique Bara**, **Gwendoline Batista**, **Claire Descamps** et **Amanda Roulleau** pour leur joie de vivre. Vous avez toutes les quatre constamment le sourire aux lèvres et c'est toujours un plaisir de discuter avec vous ! Merci à **Catherine Teigné** pour les ressources documentaires et un grand merci aux « gens du domaine et des serres » : **Alain Galon**, **Gabriel Lecocq**, **Romuald Courtais**, **Marie-France Longequeue**. Merci à **Annick Rialland** et à **Amina Lamgalas** pour leur assistance au niveau administratif et merci aux collègues du premier étage, **Abdeloifi Bechti**, **Henri De Taffin**, **Joël Favareille** et **Christine Reynier**. Merci également à **Marie-Pierre Crouzet** pour son aide concernant l'utilisation du texturomètre et à **Elise Vaud** pour avoir traité les nombreux

lots de noix en urgence ! La rapidité de ton adaptation m'impressionne encore… Merci finalement aux « super stagiaires » **Joris Molina**, **Nicolas Jammes**, **Célestine Foussard** et à l'apprentie **Morgane Vincent**, qui a passé de nombreuses heures avec moi pour mesurer les noix (« 10,6, 10,4, 12,4, etc. »). Merci **Julie Crabier** pour ton aide précieuse durant la dernière année et pour ta bonne humeur ! Ce sont elles·eux qui font avancer la science !

Merci au personnel de la Station de la Noix de Creysse, particulièrement **Eloïse Tranchand** pour ta gentillesse et ta bienveillance. Merci pour les photos. Merci au personnel de la SENuRA, dont **Marianne Naudin** pour le partage des données phénologiques.

Je remercie du fond du cœur toutes les merveilleuses personnes qui j'ai rencontrées à l'Université de Davis en Californie. D'abord, quel magnifique pays. Entre le fourmillement culturel de San Francisco, la grandeur de Los Angeles et la beauté inouïe du Yosemite Park et de la Death Valley, vous n'avez rien à envier aux français ! Many thanks to **David Neale**, **Charles Leslie**, **Brian Allen**, **Sara Montanari**, **Gina Sideli** and especially **Annarita Marrano** for her great knowledge about GWAS. Thank you **Naomi** and **Akira** for the wonderful and unforgettable Thanksgiving evening! A kiss to **Fluffy**, even if I am allergic!

Un grand merci aux membres de mon comité de suivi de thèse ; **Benjamin Brachi**, Chargé de Recherche au sein de l'UMR BIOGECO à l'INRAE de Nouvelle-Aquitaine que j'ai beaucoup sollicité concernant la GWAS et **Charles-Eric Durel**, Directeur de Recherche au sein de l'IRHS, que je suis allé rencontrer à l'INRAE Pays de la Loire-Angers. Avoir deux grands chercheurs tels que vous disponibles pour discuter de génétique quantitative m'a grandement rassuré ! Merci aussi à **Vincent Segura**, Chargé de Recherche au sein de l'UMR AGAP à l'INRAE Occitanie-Montpellier de m'avoir apporté votre grande expertise concernant les modèles statistiques appliqués à la GWAS !

Les financeurs·ceuses possèdent un rôle primordial dans tout projet de recherche. Merci donc au CTIFL, porteur du projet, et à son partenaire, l'INRAE de Nouvelle-Aquitaine-Bordeaux. Merci à l'ANRT via son dispositif Cifre d'avoir financé la moitié du doctorat et à la région Nouvelle-Aquitaine d'avoir financé une partie du projet « INNOV'noyer ». Merci au pôle de compétitivité AgriSudOuest Innovation d'avoir labellisé le projet. Merci enfin à l'Académie d'Agriculture de France de m'avoir octroyé la bourse Dufrenoy via le Crédit Agricole d'Ile-de-France et à l'Initiative d'Excellence de l'Université de Bordeaux d'avoir accepté de financer une partie de mon projet de mobilité à l'Université de Davis.

Merci à l'Université de Bordeaux, à l'Ecole Doctorale Sciences de la Vie et de la Santé et à ma tutrice de thèse **Emmanuelle Bayer**, au CNRS UMR 5200, pour sa gentillesse, son écoute et son regard extérieur durant le comité de suivi de thèse et sur les derniers mois de thèse.

Merci également à **Laurence Garmendia-Auckenthaler**, Adjointe au Chef de Département Biologie et Amélioration des Plantes, pour avoir trouvé une solution de financement en CDD durant mon dernier été de thèse. Grâce à toi, j'ai passé un été serein et j'ai pu donner un coup de main à l'équipe A3C sur d'autres travaux concernant le cerisier ; que du bonus ! En conséquence, un grand merci au département BAP pour cette aide précieuse…

Mes prochains remerciements sont adressés aux membres de mon jury de thèse qui ont accepté de donner de leur temps pour relire ce présent manuscrit et assister à ma soutenance. Un grand merci donc à mes deux rapporteur·e·s : **Evelyne Costes**, Directrice de Recherche au sein de l'UMR AGAP à l'INRAE Occitanie-Montpellier et **Didier Peltier**, Professeur à Université d'Angers. Merci à mes trois examinateurs·trices : **Laurent Bouffier**, Chargé de Recherche au sein de l'UMR BIOGECO à l'INRAE Nouvelle-Aquitaine, **Vincent Segura**, précédemment cité et **Valérie Schurdi-Levraud**, Maître de Conférences de l'Université de Bordeaux, qui a très largement favorisé à ma passion pour la génétique en Master ! C'est un honneur pour moi que tu fasses partie de mon jury et une « validation » de mon travail de ta part est très importante ; la boucle est bouclée ! Un grand merci, enfin, à **Michel Hernould,** Professeur à l'Université de Bordeaux, président de mon jury. Je n'oublierai jamais votre « avis favorable » à intégrer votre Master suite à ma Licence Pro ! J'espère que vous serez toutes et tous convaincu·e·s par mon travail et autant intéressé·e·s par ce sujet que j'ai pu l'être ! Enfin, un immense merci à **Elisa Marguerit**, Maître de Conférences à Bordeaux Sciences Agro, qui a accepté d'être invitée.

Pour terminer, je voudrais remercier chaleureusement mes ami·e·s qui ont su être présent·e·s à chaque moment important. Merci **Sophie Gelin**, la meilleure amie que l'on puisse avoir… Merci **Cécile Hastoy**, **Tamara Billon-Goedertier**, **Marie Brault**, **Romain Jacquet**, **Xavier Arnoux**, **Guillaume Lundi**, **Jérémy Humez**, **Charlotte Degraeve-Guilbault** et **Iason Tsarmpopoulos**. Ce groupe qui vient d'être cité est le groupe des « Loulous » et ces personnes sont d'une extrême importance dans ma vie ! Même si je ne suis pas quelqu'un d'extraverti et qui communique beaucoup sur ses sentiments ou émotions, sachez que je vous aime et que vous êtes tou·te·s formidables ! Merci aussi **Julie Pernin**, tu

me manques beaucoup et on devrait se voir plus maintenant ! Merci aussi à **mes parents** et à **mon frère** d'être toujours « fiers de moi » alors que je ne fais rien d'extraordinaire ! Le plus grand des « merci » revient à mon compagnon **Sylvain Chevalier**. Tu as subi la thèse mais tu n'es toujours pas capable d'expliquer ce que je fais ! Merci pour ton soutien, ta patience et ton amour… Merci aussi à tous nos ami·e·s en commun… Enfin, peu de gens comprendront mais… Merci **Blomma**… et **Bianca**, petite dernière venue !

En quelques chiffres anecdotiques, cette thèse, c'est plus de 34 000 noix ramassées, 85 000 mesures individuelles et 54 000 km parcourus pour me rendre au CTIFL, soit plus qu'un tour du monde ! Malgré tout, si c'était à refaire, je resigne tout de suite ! Je regrette déjà que ce moment plaisant à rédiger doivent se terminer. Comme vous pouvez le voir, une thèse, c'est avant tout un travail d'équipe, de collaboration et d'entraide. Nous avons notamment tou·te·s connu la situation de stagiaire, parfois stressante, donc je tenais à mettre en avant le travail des stagiaires tout au long du manuscrit. J'espère que ces petites touches ça et là de compliments dans ce manuscrit concernant leurs travaux n'en rendront la lecture que plus agréable et non pas moins sérieuse ! Et j'encourage aussi chacun·e d'entre vous à adopter l'écriture inclusive !

Je le reconnais, cette thèse a porté sur un sujet très large, ce qui est inhabituel et peut paraître déroutant. Il ne s'agit pas d'une thèse « classique », dans le sens où elle n'apporte pas une nouvelle pierre à l'édifice représenté par des longues années de recherches dejà menées au sein d'un laboratoire. Il était question de « dégrossir » un sujet large, basé sur un matériel végétal unique. Mais c'est aussi l'originalité de ce sujet qui a su me convaincre. J'espère donc que le format et la longueur du manuscrit sauront être appréciés ! Il n'est certes pas aussi relaxant à lire qu'un roman d'été d'Aurélie Valognes, mais j'ai tenté de le rendre le plus didactique et agréable que possible…

Je tiens aussi à signaler que cette thèse s'est terminée dans un contexte particulier, celui de la pandémie de Covid-19 qui a touché le monde au premier trimestre 2020. Les mesures de confinement, essentielles pour limiter autant que faire se peut les contaminations, ont eu un impact évident sur ce travail. J'ai mentionné dans les paragraphes concernés les conséquences de ce confinement.

Je n'oublierai jamais ces trois années, ça c'est certain !

Mille mercis à tou·te·s et à ceux·celles que j'aurais pu oublier…

Productions scientifiques

1. Publications

Bernard A, Lheureux F, Dirlewanger E **(2017)** Walnut: past and future of genetic improvement. *Tree Genetics & Genomes* 14:1. DOI 10.1007/s11295-017-1214-0.

Bernard A, Barreneche T, Lheureux F, Dirlewanger E **(2018)** Analysis of genetic diversity and structure in a worldwide walnut (*Juglans regia* L.) germplasm using SSR markers. *PLoS ONE* 13(11):e0208021. DOI 10.1371/journal.pone.0208021.

Bernard A, Barreneche T, Delmas M, Durand S, Pommier C, Lheureux F, Tranchand E, Naudin M, Dirlewanger E **(2019)** The walnut genetic resources of INRA: chronological phenotypic data and ontology. *BMC Research Notes* 12:662. DOI 10.1186/s13104-019-4678-1.

Bernard A, Marrano A, Donkpegan A, Brown PJ, Leslie CA, Neale DB, Lheureux F, Dirlewanger E **(2020)** Association and linkage mapping to unravel genetic architecture of phenological traits and lateral bearing in Persian walnut (*Juglans regia* L.). *BMC Genomics* 21:203. DOI 10.1186/s12864-020-6616-y.

Bernard A, Barreneche T, Donkpegan A, Lheureux F, Dirlewanger E **(2020)** Comparison of structure analyses and core collections for the management of walnut genetic resources. *Tree Genetics & Genomes* 16:76. DOI 10.1007/s11295-020-01469-5.

Bernard A, Hamdy S, Le Corre L, Dirlewanger E, Lheureux F **(2020)** 3D characterization of walnut morphological traits using X-ray computed tomography. *Plant Methods* 16:115. DOI 10.1186/s13007-020-00657-7.

Bernard A, Crabier J, Donkpegan A, Marrano A, Lheureux F, Dirlewanger E **(2020)** Genome-wide association study reveals candidate genes involved in fruit traits variation in Persian walnut (*Juglans regia* L.). *Frontiers in Plant Science* (submitted September, 16th 2020).

Donkpegan A, Barreneche T, Quero-García J, Bernard A, Fouche M, Branchereau C, Le Dantec L, Wenden B, Christmann H, Dirlewanger E **(2020)** Genome-wide association study for traits related to fruit quality in sweet cherry germplasm collection (*Prunus avium* L.) mapped on genotyping-by-sequencing. (in preparation).

Bases de données:

Bernard A, Dirlewanger E, Delmas M, Barreneche T **(2018)** INRAE's Walnut Genetic Resources Observation. *Portail Data INRAE, V5.0*. DOI 10.15454/58SYJV.

Bernard A, Germain E **(2018)** Walnut Trait Ontology. *Portail Data INRAE, V3.0*. DOI 10.15454/AV5RT2.

Bernard A **(2019)** INRAE's Walnut Genotyping Resources. *Portail Data INRAE, V3.1*. DOI 10.15454/XPKII8.

2. Communications

Bernard A « Projet INNOV'noyer, insuffler un vent de nouveauté dans la filière nucicole française ». *Médiafel*, **25/10/2017** [En ligne] http://www.arboriculture-fruitiere.com/articles/technique-fruit/insuffler-un-vent-de-nouveaute-dans-la-filiere-nucicole-francaise.

Lheureux F, Verhaeghe A, Bernard A, Dirlewanger E, Delmas M « Recherche et développement sur la noix, des projets pour un nouvel élan ». *InfosCtifl*, n°342, **juin 2018**, p. 22-31.

Bernard A, Molina J, Dirlewanger E, Lheureux F **(2018)** The French walnut improvement program: preliminary investigations. *XXX International Horticultural Congress* (IHC 2018), Istanbul, Turkey, August 12-16, 2018 (Poster).

Bernard A, Barreneche T, Lheureux F, Dirlewanger E **(2018)** SSR genetic diversity assessment of the INRAE's walnut (*Juglans* spp.) germplasm collection. *XXX International Horticultural Congress* (IHC 2018), Istanbul, Turkey, August 12-16, 2018 (Oral presentation).

Bernard A « La collection de noyers de l'INRA, un réservoir de diversité essentiel pour l'avenir de la filière nucicole ». *Hommes & Plantes*, n°106, **novembre 2018**, p. 38-45.

Bernard A « The INRA's Walnut Germplasm Collection, a Reservoir of Diversity Essential for the Future of French Walnut Industry ». *Nutfruit Magazine*, edition 77, n°2, **July 2019**, p. 28-29.

Bernard A, Marrano A, Donkpegan A, Brown PJ, Leslie CA, Neale DB, Lheureux F, Dirlewanger E **(2020)** Association and Linkage Mapping of Walnut (*Juglans regia* L.) Phenological Traits. *International Plant & Animal Genome XXVIII*, San Diego, CA, USA, January 11-15, 2020 (Oral presentation).

Bernard A, Vincent M, Molina J, Foussard C, Jammes N, Lheureux F, Crabier J, Delmas M, Barreneche T, Dirlewanger E « Caractérisation et sélection de matériel végétal innovant du noyer, de la génomique à la sélection variétale ». *InfosCtifl*, n°364, **septembre 2020**, p. 31-41.

3. Encadrement de stagiaires

Joris Molina
- Licence Professionnelle Génomes et Biotechnologies pour l'Amélioration des Plantes, à l'Université de Toulouse III Paul Sabatier.
- Stage de mars à août 2017.
- Etude de la sensibilité et/ou de la tolerance de différentes accessions de noyer à *Colletotrichum acutatum*.

Morgane Vincent
- Niveau L3 dans le cadre de son apprentissage en entreprise, au CTIFL, à l'école d'ingénieurs Agrocampus Ouest option Horticulture.
- Stage de mars à août 2018.
- *Colletotrichum acutatum* et *Xanthomonas arboricola* pv. *juglandis* : mise au point d'une méthode d'inoculation d'organes détachés de noyer en conditions contrôlées.

Nicolas Jammes
- DUT Mesures Physiques, à l'Université de Limoges.
- Stage d'avril à juin 2018.
- Création d'un protocole permettant la détermination de la dureté des noix.

Julie Crabier
- Master 2 Biologie Agrosciences, à l'Université de Bordeaux.
- Stage de janvier à juin 2020.
- Characterization and Genome-Wide Association Study of Nut Related Traits in Walnut (*Juglans regia* L.).

Résumé

Face aux conditions climatiques futures, la filière nucicole française, représentant le deuxième verger en surfaces cultivées, a besoin de nouveaux cultivars adaptés. L'arrêt du programme d'amélioration génétique en 2007, dont l'INRAE avait la charge, a entrainé une pause dans le renouvellement variétal en France. Or, l'institut possède une collection de ressources génétiques sur le site de l'Unité Expérimentale Arboricole (UEA) de Toulenne qui est une des plus riches d'Europe. En 2017, le CTIFL, centre opérationnel de Lanxade, en partenariat avec l'INRAE de Nouvelle-Aquitaine-Bordeaux, a pris la décision de relancer un travail sur le matériel végétal « noyer » et d'exploiter la collection : c'est ainsi que le projet financé par l'ANRT et la Région Nouvelle-Aquitaine « INNOV'noyer » a vu le jour. Dans le cadre de ce projet, cette présente thèse Cifre vise à étudier le déterminisme génétique des principaux caractères d'intérêt agronomique du noyer, de la fleur au fruit, en vue de mettre en place les outils nécessaires à une sélection assistée par marqueurs.

Une synthèse bibliographique résumant l'ensemble des études moléculaires, de diversité et « omics » réalisées sur le noyer à l'échelle du monde et donnant un aperçu des principaux programmes d'amélioration du noyer a tout d'abord été réalisée. Le travail de prospection, d'introduction et de caractérisation que l'on doit principalement à l'équipe d'Éric Germain a permis de rassembler sur l'UEA de Toulenne la majeure partie des espèces du genre *Juglans* et un nombre important d'accessions de noyer cultivé, *Juglans regia* L. L'exploitation des archives accumulées depuis près de 30 ans a permis de déposer en septembre 2018 d'importantes données chronologiques de phénotypage concernant la collection sur le système d'information GnpIS de l'INRAE de Versailles-Grignon. Grâce à ces données, l'avancée de la date de débourrement et des dates de floraisons mâle et femelle des deux variétés témoins 'Lara' et 'Franquette' au cours des 25 dernières années, en lien avec l'élévation des températures printanières, a été mise en évidence.

Sur la base des données de génotypage obtenues avec un ensemble de 13 microsatellites, des allèles spécifiques aux espèces *Juglans* ont été identifiés et la structure de 217 accessions de la collection a été étudiée. La structure s'est révélée être subdivisée en deux sous-groupes principaux, l'un comprenant des accessions de l'Europe de l'est à l'Asie et l'autre, comprenant des accessions de l'Europe de l'ouest aux Etat-Unis. Aussi, une core collection de 170 accessions a été définie et exploitée pour réaliser des études de Genome-

Wide Association Study (GWAS) sur les principaux caractères d'intérêt agronomique, grâce à l'utilisation d'une puce de 600 000 SNP mise au point par l'Université de Davis en Californie. Des associations entre des SNP et plusieurs caractères ont été mises en évidence, en particulier pour les caractères liés à la phénologie, grâce aux données des archives et à celles nouvellement acquises. Un SNP lié à la date de débourrement des feuilles et fleurs femelles a été identifié sur le chromosome 1 et co-localise avec un QTL détecté en parallèle sur une descendance F_1. Un marqueur de type Kompetitive Allele Specific PCR (KASP) a été validé grâce à un panel d'accessions provenant de l'Université de Davis. Il pourra à terme être utilisé en sélection assistée par marqueurs. Des associations ont également été identifiées pour le type de dichogamie et de fructification, ce dernier caractère intervenant directement sur le rendement.

L'utilisation de différents types de marqueurs durant cette thèse, SSR et SNP, a soulevé une question légitime sur le choix à réaliser dans le cadre du management des ressources génétiques. Un travail plutôt descriptif de comparaison de la puissance des deux marqueurs a été réalisé, sur les mêmes accessions, dans le but de donner des éléments de réponse dans ce choix à préconiser selon la tâche à entreprendre. Ces tâches sont par exemple la détermination de la structure, de l'apparentement et la création de core collections. D'une manière générale, il a été démontré que 13 marqueurs SSR donnent les mêmes résultats principaux que des milliers de SNP en ce qui concerne la structure et l'apparentement. De même, les core collections construites sont très similaires avec les deux types de marqueurs.

D'autres analyses GWAS ont également été réalisées sur les caractères liés au fruit. On peut citer la taille de la noix (hauteur, diamètre), son poids, le rendement au cassage et la force nécessaire pour rompre la noix. En parallèle, des méthodes utilisant des techniques de phénotypage robustes ont été développées, comme l'utilisation de la microtomographie à rayons X pour mesurer tous les caractères morphologiques, sans casser la noix. Suite à ces travaux, des outils seront maintenant disponibles et pourront être exploités pour initier un nouveau programme de création variétale qui sera mené par le CTIFL. Ces nouvelles variétés devront répondre aux critères recherchés dans les années à venir, prenant en compte le changement climatique. Ces outils appliqués sont notamment : 1) une base de données de phénotypage robuste du futur matériel de breeding, 2) des informations de parentés entre ces accessions, 3) des zones du génome contrôlant des caractères d'intérêt agronomique et des marqueurs moléculaires utilisables en sélection.

Abstract

Considering future climatic conditions, the French walnut sector, representing the second largest orchard in terms of cultivated area, needs new adapted cultivars. The end of the INRAE breeding program in 2007 led to a pause in varietal renewal in France. However, the Institute has a genetic resources collection in the Fruit Experimental Unit in Toulenne, which is one of the richest in Europe. In 2017, the CTIFL, operational centre of Lanxade, in partnership with the INRAE of Nouvelle-Aquitaine-Bordeaux, took the decision to restart working on walnut plant material and to exploit the collection: this is how the project "INNOV'noyer" financed by the ANRT and the Région Nouvelle-Aquitaine came into being. Within this project, this “Cifre” thesis aims to study the genetic determinants of the main agronomically interesting traits of walnut, from the flower to the fruit, in order to set up the necessary tools for marker-assisted selection.

Firstly, a literature review summarizing all the molecular, diversity and omics studies carried out on walnut worldwide and giving an overview of the main walnut breeding programs was carried out. The prospecting, introduction and characterization work conducted mainly by Eric Germain's team led to gather most of the species of the genus *Juglans* and a large number of accessions of cultivated walnut, *Juglans regia* L. in Toulenne. The use of archives accumulated over nearly 30 years has made it possible to deposit in September 2018 important chronological phenotyping data regarding the collection on the GnpIS information system of the INRAE of Versailles-Grignon. Thanks to these data, the advance of the budbreak date and the male/female flowering dates of the two control varieties 'Lara' and 'Franquette' over the last 25 years, has been highlighted in relation to climate change.

On the basis of genotyping data obtained with a set of 13 SSR markers, alleles specific to *Juglans* species were identified and the structure of 217 accessions of the collection was studied. The structure was found to be subdivided into two main subgroups, one comprising accessions from Eastern Europe and Asia and the other comprising accessions from Western Europe and America. Also, a core collection of 170 accessions has been defined and used to conduct Genome-Wide Association Studies (GWAS) on major traits of agronomic interest, using a 600,000 SNP array developed by the University of Davis in California. Associations between SNPs and several traits were found, particularly for phenological traits, using the legacy data and newly acquired data. A SNP linked to the budbreak date of leaves and female flowers was identified on chromosome 1 and co-localized with a QTL detected on a F_1

progeny. A Kompetitive Allele Specific PCR (KASP) marker was validated with plant material from the University of Davis. It could eventually be used in marker-assisted selection. Marker-trait associations have also been identified for dichogamy and bearing habit, the latter having a direct impact on yield.

The use of different types of markers during this thesis, SSR and SNP, raised a legitimate question about the choice to be made for genetic resources management. A descriptive work comparing the power of the two markers was carried out, on the same accessions, in order to provide elements of response in the choice to be recommended according to the germplasm management task. These tasks are, for example, the determination of the structure and the relatedness, and the construction of core collections. It has been shown that 13 SSR markers give globally the same main results as thousands of SNPs with regard to structure and relatedness. In addition, the core collections built are very similar with both types of markers.

Further genome-wide association studies were also carried out on fruit-related traits. These include nut size (height, diameter), weight, breaking yield and force required to break the nut. In parallel, robust phenotyping methods have been developed, such as the use of X-ray computed tomography to measure all morphological characteristics without breaking the nut. As a result, tools will now be available and can be exploited to initiate a new breeding program to be conducted by the CTIFL. These new varieties will have to meet the criteria sought in the coming years, taking into account climate change. These applied tools include: 1) a robust phenotyping database of future breeding material, 2) information on relatedness between accessions, 3) genomic regions controlling traits of agronomic interest and molecular markers that can be used in breeding processes.

Sommaire

Remerciements v

Productions scientifiques xi

1. Publications xi
2. Communications xii
3. Encadrement de stagiaires xiii

Résumé xv

Abstract xvii

Sommaire 1

Liste des tableaux 3

Liste des figures 5

Liste des annexes 7

Liste des abréviations 9

Introduction générale 15

Chapitre 1 – Revue bibliographique 21

1. Le noyer et son amélioration génétique 21
2. L'importance des ressources phytogénétiques 55
3. L'importance du phénotypage à l'ère du génotypage haut-débit 65
4. La méthodologie de la GWAS 69
5. Les objectifs de la thèse et la démarche scientifique 91
   - 5.1. Les objectifs 91
   - 5.2. La démarche scientifique 92

Chapitre 2 – Matériels et Méthodes 97

1. Matériel végétal 97
   - 1.1. La collection de ressources génétiques *Juglans* spp. de l'INRAE de Nouvelle-Aquitaine-Bordeaux 97
   - 1.2. La descendance F_1 de cartographie génétique 99
2. Caractérisation phénotypique 101
   - 2.1. Caractérisation de la phénologie et du type de fructification 101
   - 2.2. Caractérisation de la noix sèche 104
   - 2.3. Caractérisation de la coque et du cerneau 107
   - 2.4. Caractérisation de la qualité du cerneau 111

2.5. Visualisation et analyse des données de caractérisation phénotypique 112
2.6. Caractérisation de la sensibilité aux deux principaux pathogènes 112
3. Analyses moléculaires 115
3.1. Prélèvement du matériel végétal et extraction d'ADN 115
3.2. Amplification par PCR de marqueurs microsatellites 115
3.3. Génotypage à l'aide de la puce de génotypage Axiom™ *J. regia* 700K SNP 116
4. Analyses génétiques 117
4.1. Calcul des paramètres de diversité 117
4.2. Analyse de la structure de la population 117
4.3. Détermination des arbres de distance génétique 119
4.4. Cartographie génétique et détection de QTL 119
4.5. Analyse en génétique d'association 120

Chapitre 3 – Résultats et Discussion 125
1. Caractérisation phénotypique de la collection de ressources génétiques *J. regia* de l'INRAE 125
1.1. Données phénotypiques chronologiques issues des archives d'Eric Germain de l'INRAE 125
1.2. Données phénotypiques acquises pendant les 3 années de thèse, de la fleur au fruit 130
1.3. Caractérisation de la noix en coque via l'utilisation de la microtomographie à rayons X 137
2. Analyse de la diversité génétique et de la structure de la collection de ressources génétiques *Juglans* de l'INRAE via l'utilisation de marqueurs SSR 149
3. Etude du déterminisme génétique du type de fructification et des caractères liés à la phénologie 173
4. Etude du déterminisme génétique des caractères liés au fruit 203
5. Contribution au management des ressources génétiques du noyer 245

Conclusion et Perspectives 263
Références bibliographiques 275
Annexes 295

Liste des tableaux

Tableau 1 : Synthèse des campagnes de caractérisation phénotypique. 113

Liste des figures

Figure 1 : Cariatides de la Salle des Cariatides au Musée du Louvre. 22
Figure 2 : Nature morte de Georg Flegel, XVIIè. 22
Figure 3 : Modèle simple de la conservation des RPG et apport des outils de génomique. 59
Figure 4 : Changement annuel de la température de l'air de la surface terrestre et de la suface océanique. 61
Figure 5 : Date de débourrement moyenne pour 'Franquette' et 'Lara' en France entre 1989 et 2017. 62
Figure 6 : Approche de base pour la réalisation d'une GWAS. 74
Figure 7 : Représentation schématique du déséquilibre de liaison. 75
Figure 8 : Représentation fictive d'une association. 76
Figure 9 : Exemple fictif de diagramme quantile-quantile montrant la fiabilité des résultats de GWAS. 79
Figure 10 : Représentation fictive d'une association induite par l'hétérogénéité génétique. 81
Figure 11 : Concept de développement de FarmCPU. 83
Figure 12 : Description globale du matériel végétal étudié au sein de la collection de ressources génétiques. 99
Figure 13 : Stades phénologiques de l'inflorescence femelle du noyer *J. regia* proposés par Germain et al. 1999. 101
Figure 14 : Stades phénologiques de l'inflorescence mâle du noyer *J. regia* proposés par Germain et al. 1999. 102
Figure 15 : Mesures de la hauteur et des largeurs d'une noix. 105
Figure 16 : Dessins des différentes formes majeures de noix de face et de profil. 106
Figure 17 : Principe de la jauge de déformation. 108
Figure 18 : Photographies de cerneaux non-commercialisables. 109
Figure 19 : Principe de la tomographie 3D. 110
Figure 20 : Résultats de l'ACP réalisée sur l'ensemble des données de phénotypage acquises pendant la thèse. 131
Figure 21 : Graphique des individus de l'ACP sur les données de 2018. 134
Figure 22 : Utilisation des puits disponibles sur l'ensemble des 3 puces de génotypage. 170
Figure 23 : Nuage de points de la date de débourrement des 10 accessions phénotypées durant les 3 années de thèse. 200
Figure 24 : Données météorologiques des hivers 2017-2018 et 2018-2019. 201

Figure 25 : Schéma de synthèse des loci majeurs identifiés dans l'architecture génétique des caractères d'intérêt agronomiques du noyer. 267
Figure 26 : Idées de croisements pour apporter une tardiveté dans la phénologie..... 270
Figure 27 : Idées de croisements pour apporter un calibre plus important. 270
Figure 28 : Idées de rétrocroisements ou de croisements supplémentaires..... 271

Liste des annexes

Annexe 1 : Article « La collection de noyers de l'INRA, un réservoir de diversité essentiel pour l'avenir de la filière nucicole » dans la revue trimestrielle « Hommes et Plantes ». 295
Annexe 2 : Article « The INRA's Walnut Germplasm Collection, a Reservoir of Diversity Essential for the Future of French Walnut Industry » dans la revue « Nutfruit Magazine ». 304
Annexe 3 : Synthèse des travaux réalisés en pathologie. 306

Liste des abréviations

2DE – Two-dimensional gel Electrophoresis
A – Number of alleles
ACP – Analyse en Composantes Principales
ACoP – Analyse en Coordonnées Principales
A.D. – *Anno Domini*
A_e – Number of effective alleles
AFLP – Amplified Fragment Length Polymorphism
ANRT – Agence Nationale de la Recherche et de la Technologie
ANSES – Agence Nationale de Sécurité Sanitaire de l'alimentation, de l'environnement et du travail
AOP – Appelation d'Origine Protégée
BAC – Bacterial Artificial Chromosome
BAN – Brown Apical Necrosis
B.C. – Before Christ
BFP – Biologie du Fruit et Pathologie
BIC – Bayesian Information Criterion
BLUPs – Best Linear Unbiased Predictions
cDNA – complementary DNA
Cifre – Conventions Industrielles de Formation par la REcherche
CIMMYT – Centro Internacional de Mejoramiento de Maiz Y Trigo
CLRV – Cherry LeafRoll Virus
cM – centimorgan
CMR – Cancérogène, Mutagène et toxique pour la Reproduction
CNRS – Centre National de la Recherche Scientifique
CoMM – Collaborative Mixed Model
COST – COopération européenne en Science et Technologie
cpDNA – chloroplast DNA
CTIFL – Centre Technique Interprofessionnel des Fruits et Légumes
CTPS – Comité Technique Permanent de la Sélection
DAM – Dormancy-Associated MADS-Box
DO – Densité Optique
DOI – Digital Object Identifier
DL – Déséquilibre de Liaison
DNA – DeoxyriboNucleic Acid
DUT – Diplôme Universitaire de Technologie
EMMA – Efficient Mixed-Model Association
EST – Expressed Sequence Tag
FAIR – Findable, Accessible, Interoperable, Reusable
FAO – Food and Agriculture Organization of the United Nations
FDR – False Discovery Rate
F_{IS} – Inbreeding Coefficient
F_{ST} – Fixation Index
GC-MS – Gas Chromatography-Mass Spectroscopy
GEBV – Genomic Estimated Breeding Values
GEVES – Groupe d'Etude et de contrôle des Variétés Et des Semences
GFP – Green Fluorescent Protein
GIEC – Groupe d'experts Intergouvernemental sur l'Evolution du Climat
GIS – Groupement d'Intérêt Scientifique
GISS – Goddard Institute for Space Studies (NASA)
GISTEMP – GISS Surface Temperature Analysis
GLM – Generalized Linear Model
GPS – Global Positioning System
GWAS – Genome-Wide Association Study
H^2 – Broad-sense heritability
h^2 – Narrow-sense heritability
ha – hectare

HDL – High Density Lipoprotein
H_e – Expected Heterozygosity
H_o – Observed Heterozygosity
HPLC-q-ToF-MS – High-Performance Liquid Chromatography-Quadrupole-Time of Flight Mass Spectroscopy
HTS – High Throughput Sequencing
IBD – Identity By Descent
InDel – Insertion-Deletion
INRA – Institut National de la Recherche Agronomique
Devenu INRAE au 1er janvier 2020 : Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement
IPBES – Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services
IPGRI – International Plant Genetic Resources Institute
Devenu Bioversity International en 2006
ISSR – Inter-Simple Sequence Repeat
ITERG – Institut des Corps Gras
ITS – Internal Transcribed Spacers
IUCN – International Union for Conservation of Nature
KASP – Kompetitive Allele Specific PCR
kb – kilo base pairs
kcal – kilocalorie
KING – Kinship-based INference for Genome-wide association studies
kt – kiloton
LA – Linkage Analysis
LD – Linkage Disequilibrium
LDL – Low Density Liporpotein
LG – Linkage Group
LOD – Logarithm of Odds Ratio
Log_{10} – Logarithme decimal
LOWESS – Locally Weighted Scatterplot Smoothing
LPGA – Levure Peptone Glucose Agar
LUBEM – Laboratoire Universitaire de Biodiversité et Ecologie Microbienne
MAGIC – Multi-parent Advanced Generation Inter-Cross
MAS – Marker-Assisted Selection
Mb – Million base pairs
MCMC – Markov Chain Monte Carlo
MIM – Multiple Interval Mapping
MLM – Mixed Linear Model
mM – millimolaire
MLMM – Multi-Locus Mixed-Model
mRNA – messenger RNA
mya – million years ago
NAM – Nested Association Mapping
NASA – National Aeronautics and Space Administration
NBS-LRR – Nucleotide-Binding Site Leucine-Rich Repeat
NCBI – National Center for Biotechnology Information
ONU – Organisation des Nations Unies
PCD – Programmed Cell Death
PCR – Polymerase Chain Reaction
PCA – Principal Component Analysis
PCoA – Principal Coordinate Analysis
PDA – Potato Dextrose Agar
PIC – Polymorphism Information Content
PPO – PolyPhenol Oxidase
PUFA – PolyUnsaturated Fatty Acid
QTL – Quantitative Trait Locus
QTN – Quantitative Trait Nucleotides
RAPD – Ramdomly Amplified Polymorphic DNA
RARe – Ressources Agronomiques pour la Recherche
rDNA – ribosomal DNA
REAP – Relatedness Estimation in Admixed Populations

RFLP – Restriction Fragment Length Polymorphism
RIL – Recombinant Inbred Line
RNA – RiboNucleic Acid
RNAi – RNA interference
rRNA – ribosomal RNA
RPG – Ressources PhytoGénétiques
RRL – Reduced Representation Libraries
SAM – Sélection Assistée par Marqueurs
SCAR – Sequence-Characterized Amplified Region
SDS-PAGE – Sodium Dodecyl Sulfate-PolyAcrylamide Gel Electrophoresis
SENuRA – Station d'Expérimentation Nucicole en Rhône-Alpes
SLAF – Specific-Locus Amplified Fragment
SNP – Single Nucleotide Polymorphism
SSR – Single Sequence Repeat
SVT – Sciences de la Vie et de la Terre
SWEET – Sugar Will Eventually be Exported Transporter
T-LiDAR – Terrestrial Light Detection And Ranging
TDF – Transcript-Derived Fragment
tRNA – Transfer RNA
TWAS – Transcriptome-Wide Association Study
TZ – Transition Zone
UAV – Unmanned Aerial Vehicle
UEA – Unité Expérimentale Arboricole
UFC – Unité Formant Colonie
UMR – Unité Mixte de Recherche
UPGMA – Unweighted Pair Group Method with Arithmetic Mean
UPOV – Union for the Protection of New Varieties of Plants
URGI – Unité de Recherche Génomique Info
U.S.A. – United States of America
USDA – United States Department of Agriculture
VOC – Vertical Oozing Canker
WGIG – Walnut Genomics Implementation Group

Introduction générale

Introduction générale

Source de mythes et de légendes depuis l'Antiquité, le noyer commun *Juglans regia* L., originaire d'Asie Centrale, est un arbre emblématique de nos régions françaises, particulièrement en Isère et dans le Périgord. Actuellement, en termes de surfaces de plantation, les noyeraies représentent le deuxième verger en France derrière la pomme. La production mondiale de noix, dominée par la Chine et la Californie, ne cesse de croître pour atteindre presque 3,7 millions de tonnes en 2018.

Néanmoins, le changement climatique, un consensus dans notre communauté scientifique, met en péril cette production. L'augmentation des anomalies de température à la surface du globe en faveur d'un réchauffement global est alarmante depuis les années 1980. Ces conséquences, pour beaucoup dues à l'activité humaine, sont à l'origine de certaines difficultés d'adaptation chez les plantes, notamment les espèces pérennes comme les arbres fruitiers. Chez le noyer, les températures chaudes de plus en plus précoces au printemps entraînent une avancée des dates de débourrement et de floraison. Ainsi, des gelées printanières tardives peuvent anéantir une partie de la production. Ce changement climatique peut être aussi à l'origine de modifications de pressions des pathogènes, qu'ils soient connus ou nouveaux.

Dans ce contexte, il est crucial de disposer de collections de ressources génétiques végétales et de les valoriser. L'INRAE de Nouvelle-Aquitaine-Bordeaux possède une riche collection, fruit du travail d'Eric Germain, responsable du programme d'amélioration de 1977 à 2007. Grâce aux nombreuses collaborations et prospections qu'il a entreprises, Eric Germain a constitué un réservoir de diversité génétique et de variabilité phénotypique unique en Europe.

D'autre part, une importante dynamique dans la recherche académique sur le noyer est à souligner. L'Université de Davis en Californie est notamment à l'origine du séquençage du génome nucléaire et du développement d'une puce de génotypage à haut-débit de type Affymetrix® contenant plus de 600 000 SNP. Ainsi, les ressources génétiques du noyer peuvent aujourd'hui être valorisées non pas seulement par leur utilisation dans des processus de création variétale basés uniquement sur le phénotype, mais elles peuvent constituer le matériel de choix pour la mise en œuvre de travaux de génétique quantitative à l'échelle du génome.

En parallèle, il faut respecter notre devoir de réponse aux attentes des producteurs·trices qui sont nombreuses. Ils·elles souhaitent évidemment des cultivars adaptés aux conditions pédo-climatiques changeantes, tolérants aux deux principaux bio-agresseurs du noyer, tout en gardant une certaine typicité française concernant le goût et l'apparence générale de la noix. La plupart de ces caractères d'intérêt agronomiques sont quantitatifs et en théorie, la mise en œuvre d'une « analyse d'association tout génome » (GWAS en anglais pour Genome-Wide Association Study) permet de disséquer leur architecture génétique.

L'amélioration génétique des arbres fruitiers est un domaine passionnant extrêmement large. Il existe différents leviers mobilisables comme les leviers biologiques (travail du porte-greffe, variabilité épigénétique, édition de gène par exemple) et les leviers plutôt méthodologiques ou technologiques (approches prédictives en lien avec des environnements variés, sélection génomique). Mais étudier les ressources génétiques disponibles, une forme de biodiversité, est un des leviers biologiques incontournables pour le développement de variétés résilientes, en parallèle des conduites culturales, en accord avec les grands principes de l'agroécologie. Ainsi, le CTIFL, centre de Lanxade, en partenariat avec l'INRAE de Nouvelle-Aquitaine-Bordeaux, a souhaité construire un projet financé par l'ANRT et la Région Nouvelle-Aquitaine. Il vise à redynamiser le travail du matériel végétal « noyer » en France et à participer à l'acquisition de connaissances sur ce matériel pour le développement d'idéotypes.

La présente thèse expose les résultats d'une étude approfondie des ressources génétiques du noyer disponibles à l'INRAE de Nouvelle-Aquitaine-Bordeaux, un travail en vue de l'obtention d'outils, nécessaires en amont, pour la mise en place d'une sélection assistée par marqueurs. Principalement, cette étude a consisté en un phénotypage large des principaux caractères agronomiques, de la fleur au fruit en passant par la sensibilité aux principaux pathogènes, combiné à un génotypage dense du génome. Cette démarche a permis de répondre aux principaux objectifs : l'évaluation de la diversité génétique et de la variabilité phénotypique de la collection, l'élucidation du déterminisme génétique majeur de ces caractères et l'obtention d'outils pour le·la sélectionneur·euse, comme des marqueurs moléculaires.

L'originalité de cette thèse Cifre réside dans l'étude générale d'un matériel végétal riche et unique mais peu étudié jusqu'à présent. Les objectifs ainsi annoncés sont nombreux puisque la volonté est d'apporter les bases de connaissances les plus importantes possibles pour initier un programme de création variétale et non pas de répondre à une question

scientifique précise. Il est donc primordial d'adopter une vision plus globale, en accord avec le fonctionnement du dispositif Cifre. Ceci a conduit au choix d'annoncer un intitulé large de la thèse puisqu'elle présente un éventail important de résultats d'une étude exploratoire.

Le manuscrit proposé est décomposé de manière classique, en trois grands chapitres. Le chapitre 1 est une revue bibliographique présentant l'espèce étudiée, sa biologie et son amélioration génétique sous le format d'une « review », l'enjeu des ressources phytogénétiques face au contexte climatique, le rôle de la génétique quantitative dans les processus d'amélioration génétique et l'importance du phénotypage à l'ère de la génomique « haut-débit ». Le chapitre 2 détaille les « matériels et méthodes » utilisés pour mener à bien ces travaux. Le chapitre 3 de « résultats et discussion » regroupe principalement un ensemble d'articles publiés, ou en cours de publication, et articulés de sorte à constituer un ensemble cohérent. Des sous-chapitres permettent de faire des transitions entre ces différentes publications. L'histoire racontée par ce chapitre 3 permet d'apprécier la variabilité phénotypique de la collection ainsi que sa diversité génétique à l'aide de quelques marqueurs microsatellites. Il présente ensuite un travail de génétique d'association et de cartographie génétique lié à la phénologie du noyer, dont le panel pour la GWAS a été construit grâce aux précédents marqueurs microsatellites. Il présente également des travaux préliminaires de GWAS concernant cette fois les caractères liés au fruit. Enfin, il présente un travail plus descriptif concernant la contribution des microsatellites et des SNP dans le management optimal des ressources génétiques, comme la détermination de la structure de la collection et la création de core collections. Pour terminer, les perspectives à envisager pour la poursuite de ce travail sont proposées, ainsi que des idées pour un pre-breeding.

N. B. : Suite à la fusion de l'INRA et de l'Irstea pour former l'Institut national de recherche pour l'agriculture, l'alimentation et l'environnement, le choix a été fait d'utiliser le nouvel acronyme « INRAE » et le nom du centre « INRAE de Nouvelle-Aquitaine-Bordeaux » tout au long du manuscrit, devenus officiels respectivement au 1er janvier et au 7 février 2020. Toutefois, les informations mentionnant des travaux ou évènenements antérieurs à cette date font automatiquement références à « INRA » et au centre « INRA de Bordeaux », notamment lorsqu'il est question de l'ancien programme d'amélioration d'Eric Germain.

Chapitre 1 – Revue bibliographique

1. Le noyer et son amélioration génétique
2. L'importance des ressources phytogénétiques
3. L'importance du phénotypage à l'ère du génotypage haut-débit
4. La méthodologie de la GWAS
5. Les objectifs de la thèse et la démarche scientifique

Chapitre 1 – Revue bibliographique

1. Le noyer et son amélioration génétique

Introduction

Traversant les époques, le noyer est souvent un vieil arbre croisé au détour d'un chemin ou retrouvé au beau milieu d'un champ. C'est un arbre symbole du mois d'octobre lorsqu'il arbore ses teintes automnales. Son fruit, cassé depuis tous temps dans la paume de la main – *cela a donné son nom à la poignée de main vigoureuse dite « casse-noix » !* – offre un parfum particulier et une texture unique. Avant de décrire quelques mythes et légendes autour de cet arbre, il est nécessaire de s'intéresser à l'origine des mots « noix » et « *Juglans* ». La noix vient de *nux*, le noyau en latin, expliquant le *x* muet hérité de nos jours. Le noyer commun, en vieux français, est également baptisé *goguier*, *noguier*, *calottier*, *écalonnier* ou *gojeutier*. Cet arbre appartient au genre *Juglans*, dont le nom proviendrait de la contraction de *Jovis glans*, le « gland de Jupiter » en latin (Lamarck and Poiret 1783) car les Romains pensaient que Jupiter leur avait offert cet arbre si fécond.

Les noyers ont fait l'objet de nombreuses légendes antiques. A l'époque préhellénique, le culte des arbres était un fondement de la civilisation. L'arbre avait une âme qui se manifestait sous la forme d'une déesse (Rondeau 1997). Dans la mythologie, il est dit que Dionysos, en Laconie, s'éprit de la plus jeune fille du roi Dion, Carya. Ses deux sœurs furent transformées en rochers en raison de leur jalousie. Plus tard, Carya mourut de tristesse et fut immortalisée en noyer fécond par Dionysos. Avec l'essor de la civilisation hellénique, Carya – *qui donna son nom au genre du pacanier (noix de pécan)* – disparut peu à peu pour céder sa place à Artémis. Elle protège les lieux sauvages et ne descend que rarement dans la cité. Ainsi, lorsqu'elle n'est pas déesse de la chasse ou de la fécondité, elle devient maîtresse du noyer et des arbres portant des fruits, sous le nom d'Artémis *Karyatis* (Salles 2003). Un temple lui fut dédié et des colonnes en noyer, taillées aux formes de la déesse, furent érigées pour soutenir le chapiteau ; ce sont les caryatides ou cariatides. Aujourd'hui, il est possible d'admirer des cariatides en pierre datant de la fin du V^{è} siècle av. J.-C. à l'entrée de l'Erechthéion sur l'Acropole d'Athènes ou au Musée du Louvre, sculptées par Jean Goujon en 1550 (Figure 1).

Figure 1 : Cariatides de la Salle des Cariatides au Musée du Louvre.
Cette salle fut construite par Pierre Lescot pour le roi Henri II et les cariatides furent sculptées par Jean Goujon en 1550, pour soutenir la tribune des musiciens.

Arbre de vie ou arbre de mort ? Une certaine ambivalence s'observe puisque le noyer orne les cimetières en Poitou tandis que les noix sauvent les vertueux du Déluge selon une légende slave (Rondeau 1997). Au cours du IV^{è} siècle ap. J.-C., saint Augustin associe même la noix au symbole du Christ : le brou serait l'enveloppe charnelle de Jésus, la coque évoquerait le bois longitudinal de la croix et le noyau représenterait la divinité cachée du Christ. La noix est aussi un élément récurrent des natures mortes du peintre allemand Georg Flegel (Figure 2) et du peintre flamand Frans Snyders, au début du XVII^{è} siècle.

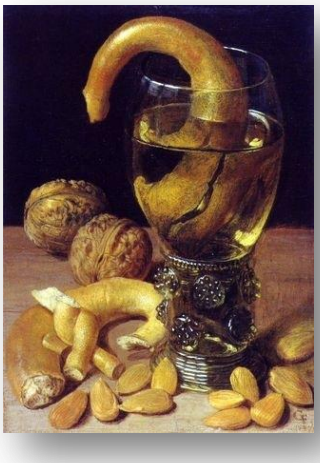


Figure 2 : Nature morte de Georg Flegel, XVIIè.
Au XVII^{è} siècle, la nature morte (genre artistique représentant des éléments inanimés) prend son essor en Flandre et en Hollande, notamment grâce aux Grandes Découvertes et à l'arrivée de nouveaux spécimens végétaux en Europe.

Anomalie source de superstitions, la noix à trois coutures préserverait celui ou celle qui la porte de toutes sortes de cataclysmes (maladies, ensorcellements, etc.) mais gare à celui·celle qui la perd ! De plus, un rêve de noyer était présage de prospérité… L'ambivalence est également présente au sujet du mariage. En effet, dans les Landes, suite à une demande en mariage, l'élue pouvait faire part de son refus par l'offrande d'un plat de noix en guise de dessert lors du repas suivant cette demande (Sébillot 1984). En parallèle, l'offrande de noix est un rite de la célébration nuptiale en Grèce et en Roumanie (Alexandri 1853). Mais le mythe le plus célèbre reste celui du Noyer de Bénévent.

Bénévent (*Benevento* en italien), ville de la région de Campanie en Italie, est réputée depuis l'époque de l'occupation lombarde au VII^{è} siècle, pour être la ville où les sorcières du monde entier se réunissaient pour célébrer le Sabbat sous un noyer séculaire. D'après la légende, les soldats Lombards accrochaient la peau des chèvres au noyer et les mangeaient après les avoir détruites avec des lances, pensant ainsi acquérir le pouvoir des chèvres. Il est dit également que, la nuit, des femmes – *les sorcières* – hurlaient et sautaient autour du noyer duquel pendaient des serpents. Pour arrêter ces rites, l'Evêque Barbato ordonna de couper le noyer vénéré. Au XVII^{è} siècle, l'apothicaire napolitain Pietro Piperno relate les racines de cette légende dans son essai de 1639 nommé *Della superstitiosa noce di Benevento*, traduit de l'italien en latin par *De Nuce Maga Beneventana* (Le Noyer Magique de Bénévent).

Aussi, parmi les 38 Elixirs floraux du Docteur Bach, médecin britannique du XX^{è} siècle, également appelés « fleurs de Bach », le 33ème élixir élaboré à base de fleurs de noyer est le remède pour se protéger des influences extérieures et de la sensibilité qui découle du changement (naissance, puberté, mariage, retraite, etc.). Aujourd'hui, les ancien·ne·s continuent d'affirmer que « rien ne pousse sous un noyer » ou que « dormir sous un noyer peut entraîner un mal de tête ». Il est donc clair que les croyances autour du noyer ne semblent pas prêtes de s'essouffler…

Pour aller plus loin, l'article en Annexe 1 publié en novembre 2018 dans la revue trimestrielle Hommes et Plantes, sous le nom de « La collection de noyers de l'INRA », apporte un complément d'informations sur l'histoire folklorique des noyers du genre *Juglans* spp.

Le genre botanique des noyers

Le noyer cultivé pour la production de noix, *Juglans regia* L., Persian walnut ou English walnut en anglais, constitue une des plus anciennes sources de nourriture pour l'être humain. Il est cultivé dans le monde entier dans les zones tempérées. En 2018, la production mondiale de noix en coque a atteint plus de 3,6 millions de tonnes (source FAOSTAT, 2020). Depuis 2008, cette production dépasse celles de l'amande et de la noisette. Globalement, cette production est croissante au cours des 20 dernières années. La Chine est le principal pays producteur, avec plus d'un million de tonnes en moyenne sur les 20 dernières années, suivie par la Californie et l'Iran. La France se situe à la 10ème position, mais représente le 3ème producteur européen derrière l'Ukraine et la Roumanie. En 2019, la France a produit plus de 35 000 tonnes de noix et presque 26 000 hectares sont alloués à sa production, ce qui en fait la deuxième culture arboricole fruitière en termes de surfaces, derrière la pomme (source Agreste, 2020). Avec ses deux AOP Noix de Grenoble et Noix du Périgord, la noix est un fruit à coque incontournable économiquement et culturellement.

Le genre *Juglans* appartient à l'ordre des Fagales, comme le noisetier et le châtaignier, puis à la famille des Juglandacées, comme le pacanier, qui donne la noix de pécan. Ce genre comprend plus de 20 espèces ; ce sont « les noyers ». Les fruits sont des drupes et toutes les espèces sont diploïdes avec 16 paires de chromosomes homologues (Woodworth 1930). L'espèce cultivée pour le fruit, le noyer commun *Juglans regia*, est originaire d'Asie centrale, sur les contreforts de l'Himalaya (Zeven and Zhukovsky 1975). C'est une espèce monoïque, donc qui possède les fleurs mâles et les fleurs femelles sur le même pied. Très souvent, une dichogamie est observée, soit un décalage dans les deux floraisons. Sa période juvénile dure en moyenne 6 ans, la pollinisation se fait grâce au vent et le génome compte environ 620 Mégabases (Marrano et al. 2020).

L'amélioration génétique du noyer

Malgré une culture très ancienne, l'amélioration génétique du noyer en tant que telle a réellement commencé au début du XXè siècle. L'historique des travaux en génétique menés sur le genre *Juglans* commence il y a presque un siècle, en 1930, avec la publication de la ploïdie (Woodworth 1930). A l'aide de différentes méthodologies, depuis les isozymes jusqu'aux récents progrès en matière de séquençage du génome, de nombreuses études génétiques du noyer ont été menées au cours des 30 dernières années. Ces études concernent notamment l'évaluation de la diversité génétique au sein de collections de ressources génétiques et de populations naturelles, la détermination de la phylogénie et de l'origine des

espèces et des accessions, la construction de cartes physiques et génétiques et les études concernant les stress biotiques et abiotiques.

Dates-clés de l'amélioration génétique

L'amélioration génétique du noyer a connu une grande évolution grâce aux pays du Moyen-Orient qui ont largement contribué à l'étude des caractéristiques morphologiques du noyer. Les premières cartes génétiques, quant à elles, ont été publiées entre 1994 et 2001 (Fjellstrom and Parfitt 1994a; Woeste et al. 1996b; Malvolti et al. 2001), mais la première carte décrivant les 16 groupes de liaison date de 2015 seulement (Luo et al. 2015). L'année charnière concernant la génomique du noyer est l'année 2016, durant laquelle les génomes nucléaire et chloroplastique ont été séquencés par l'Université de Davis en Californie (Martínez-García et al. 2016). Rapidement, un outil de génotypage haut-débit a été créé en 2018, une puce Axiom™ de 700 000 SNPs (Marrano et al. 2019a). Enfin, la version 2 du génome vient d'être publiée, avec une annotation plus précise des séquences codantes (Marrano et al. 2020).

Les besoins de la filière et le travail d'amélioration effectué en France

A l'échelle mondiale, la filière nucicole a besoin de variétés à haut rendement (meilleure production, meilleur calibre), donnant des noix facilement énoisables et ayant de bonnes qualités organoleptiques. De plus, la phénologie de ces variétés doit être adaptée aux conditions climatiques du pays producteur. Enfin, les variétés doivent être tolérantes aux principaux pathogènes (Cosmulescu and Botu 2012). Le travail de création variétale, que l'on doit aux principaux pays producteurs, montre un certain nombre de cultivars. Il est possible de citer la variété phare américaine 'Chandler', disponible depuis 1979, ou encore 'Franquette', une variété française locale ancienne, pourtant variété prédominante de nos vergers. En France, deux programmes d'amélioration du noyer se sont succédé entre 1977 et 2003, conduits par Éric Germain et son équipe à l'INRAE de Nouvelle-Aquitaine-Bordeaux, au sein de l'ex Unité de Recherche sur les Espèces Fruitières. Ces programmes étaient parmi les plus importants à l'échelle internationale, avec ceux de l'Université de Davis. Éric Germain a réalisé un premier programme dont l'objectif était d'apporter le caractère de fructification latérale, donc tout le long du rameau, ceci menant à une plus grande production. Il a donc réalisé 28 croisements entre les variétés françaises traditionnelles reconnues pour leur qualité gustative et leur adaptation locale, mais à fructification terminale (à l'extrémité des rameaux), et des cultivars californiens fructifiant latéralement. Ensuite, un second programme s'est orienté toujours sur cette fructification latérale, mais en y apportant de la diversité génétique.

En effet, pour les besoins de ses programmes, Éric Germain a été à l'initiative de nombreuses collaborations afin de réaliser des échanges de matériel et des prospections dans le monde entier. Il a ainsi constitué une riche collection de ressources génétiques *Juglans*, principalement entre 1988 et 2000. Éric Germain a aussi travaillé sur le caractère « cerneau rouge », une niche de diversification, recherchée notamment pour la pâtisserie. Seulement, un seul hybride est prometteur et en cours d'inscription au Catalogue. En 2007, l'INRAE décide de restreindre globalement le nombre d'espèces sur lesquelles l'institut développe des programmes de recherche en génétique. S'en suit la restructuration de l'Unité de Recherche sur les Espèces Fruitières, qui se concentre désormais sur les Rosacées

Les sept variétés modernes françaises sont issues du premier programme. Si ces variétés modernes sont à l'origine de meilleurs rendements, elles ne répondent pas pleinement aux futurs enjeux. En France, les variétés précoces sont sensibles aux gelées printanières et elles donnent globalement toutes des fruits d'un calibre moyen. Il est donc primordial de proposer une gamme plus large de variétés, via la sélection assistée par marqueurs, qui se justifie sur ce type d'espèce à longue période juvénile.

La revue suivante a été publiée en ligne le 3 décembre 2017 dans le journal Tree Genetics and Genomes, volume 14:1, sous le nom de « Walnut: past and future of genetic improvement » (DOI 10.1007/s11295-017-1214-0). Après une description de la taxonomie, de l'origine, de la biologie, de la phénologie, de la physiologie, de la production et de l'aspect nutritionnel, elle résume l'ensemble des études moléculaires et « omics » à l'échelle du monde et donne un aperçu des principaux programmes d'amélioration du noyer.

Tree Genetics & Genomes (2018) 14:1
https://doi.org/10.1007/s11295-017-1214-0

REVIEW

Walnut: past and future of genetic improvement

Anthony Bernard[1,2] · Fabrice Lheureux[2] · Elisabeth Dirlewanger[1]

Received: 20 April 2017 / Revised: 21 November 2017 / Accepted: 22 November 2017
© The Author(s) 2017. This article is an open access publication

Abstract
Persian or English walnut (*Juglans regia* L.), the walnut species cultivated for nut production, is one of the oldest food sources known. Persian walnuts, native to the mountain valleys of Central Asia, are grown worldwide in temperate areas. World production exceeds three million tons since 2012, mostly provided by China, the USA, and Iran. Despite very ancient culture of walnut species (*Juglans* spp.), breeding actually started in the twentieth century. Using a range of methodologies, from morphological markers to the most recent advances in genome analysis, many genetic studies of walnut have been conducted during the past 30 years, including examination of diversity, determination of relationships within or among germplasm collections and populations, phylogenetic and origin elucidation, genetic map construction, and biotic or abiotic stress investigations. The genetic improvement of walnut has undergone great evolution. The producing countries of the Middle East have widely studied morphological characteristics of walnut. The USA and France, for example, are behind important cultivar releases such as "Chandler" and "Franquette." Finally, genomics represents a major breakthrough in walnut improvement, in particular by recent sequencing of both chloroplast and nuclear genomes. This review summarizes worldwide molecular and "omics" studies and gives an overview of the main walnut breeding programs.

Keywords Genetic improvement · Breeding program · Molecular markers · Genomics · Walnut · *Juglans* spp.

Introduction

Nuts

Tree nuts are one of the oldest sources of food for man, birds, and wild animals (Woodroof 1967), and their history is very rich. Ancient fossils prove their presence on earth long before man arrived (Jaynes 1969). References to nut trees in ancient literature are numerous. Theophrastus, Greek philosopher born 370 B.C., referred to walnut and mentioned hazelnuts as growing in the mountains of Macedonia (Jaynes 1969). Furthermore, branches and flowers have been used on religious occasions since antiquity and several tree nuts, like almonds and pistachios, were mentioned in the Bible, Genesis 43:11 and Song of Solomon 6:11. It is mentioned that Persian walnuts (*Juglans regia* L.) were cultivated in King Solomon's garden, and almonds were depicted for ornamentation of the candlesticks in the Temple (Moldenke and Moldenke 1952). Nowadays, "nuts" are still a valuable food resource, widespread in the northern hemisphere due to their very high food and ornamental value (Jaynes 1969).

In botany, the word "nut" is very ambiguous. According to the *American Heritage Dictionary of the English Language, 5th Edition, 2011*, "a nut is an indehiscent fruit having a single seed enclosed in a hard shell, such as an acorn or hazelnut," or "any of various other usually edible seeds enclosed in a hard covering such as a seed coat or the stone of a drupe, as in a pine nut, peanut, almond or walnut." Finally, it seems that a consensus has been accepted by the scientific community in which a nut is a simple dry fruit with one seed in which the ovary wall becomes increasingly hard as it matures, and where the seed remains unattached within the ovary wall (Woodroof 1967; Jaynes 1969). All nuts are indehiscent and do not open at maturity. True nuts are produced for example by some species of the order *Fagales* (beech, chestnut, oak, hazelnut, etc.). However, in a general context, a wide variety of dried seeds are also called "nuts" by extension because of their shell (Jacquenet and Moneret-Vautrin 2007)—this is actually the

Communicated by S. C. González-Martínez

✉ Elisabeth Dirlewanger
elisabeth.dirlewanger@inra.fr

[1] INRA, UMR 1332 Biologie du Fruit et Pathologie, F-33140 Villenave d'Ornon, France

[2] Ctifl, centre de Lanxade, 24130 Prigonrieux, France

culinary definition. A botanical classification of species producing nuts was proposed by Jacquenet and Moneret-Vautrin (2007) (Fig. 1).

With the development of molecular biology in the 1990s, numerous works were undertaken in order to improve the characteristics of nuts, particularly of walnuts. These studies facilitated accumulation of knowledge regarding biochemistry of walnuts, their genetic diversity, and the genetic determinants of agronomic traits of interest such as yield or disease resistance. There are still efforts in progress, particularly thanks to the team at the University of California-Davis that sequenced the Persian walnut genome in April 2016 (Martínez-García et al. 2016).

Taxonomy and origin

Walnut is defined at the nut of any tree of the genus *Juglans* spp., part of the subclass *Rosidae* (Germain et al. 1999). Walnuts are among the most produced "nuts" in the world alongside almonds, hazelnuts, pistachios, and cashew nuts, according to the International Nut and Dried Fruit Council in 2015. The Persian walnut tree (*Juglans regia* L. or English walnut) is a wide spread monoecious tree species of the order *Fagales* (formerly, in the order *Juglandales*) and the family *Juglandaceae*. This family includes about 50 species of 11 genera of which *Carya* (hickory tree), *Pterocarya* (wingnut tree), and *Juglans* are the major members (Rehder 1947). The genus *Juglans* includes more than 20 species divided into four sections: *Trachycaryon* (butternut, a single species, *J. cinerea*), *Cardiocaryon* (heartnuts), *Rhysocaryon* (black walnuts, notably *J. nigra*), and *Dioscaryon* containing only *J. regia* (Manning 1978). This classification was confirmed with a study of the nucleotide sequences of the chloroplast gene *matK* and the internal transcribed spacers (ITS) and 5.8S gene of the nuclear ribosomal DNA (Stanford et al. 2000) (Table 1). *J. regia* is widely disseminated and grown in many temperate regions of Europe, North and South America, South Africa, Asia, Australia, and New Zealand (Aradhya et al. 2006). All species of the genus *Juglans* are diploid with $2n = 2x = 32$ chromosomes (Woodworth 1930).

Fossil evidence (Manchester 1987) traces the origin of genus *Juglans* to the Middle Eocene in North America (from ~48 to ~38 million years ago), well after the now extinct first tree-like plant, *Archaeopteris*, which likely first appeared during late Devonian (~382 mya). Subsequent expansion and diversification of *Juglans* from North America was made possible by the Bering land bridge during Middle Tertiary (from ~66 to ~3 mya) and the North Atlantic land bridge during late Eocene (~38 to ~33 mya). In addition, there was an adaptation to Neogene cool climate (~23 to ~3 mya) (Wolfe 1978) and other expansions towards Southeastern Europe and Central and South America during late Myocene (~12 to ~5 mya). From studying chloroplast DNA (cpDNA) intergenic spacer sequences, it seems that the division within genus *Juglans* (*Rhysocaryon*, *Cardiocaryon*, and *Dioscaryon* sections only) occurred during a period ranging from ~50 to ~40 mya, which matches with fossil analyses (Aradhya et al. 2006, 2007) (Fig. 2). The cradle of Persian walnut domestication is thought to be located in Central Asia (Zeven and Zhukovsky 1975), particularly the foothills of the Western Himalayas from the Kashmir region to Tajikistan and Kyrgyzstan. It first spread to the west, namely Uzbekistan, north Iran, Caucasus region, and

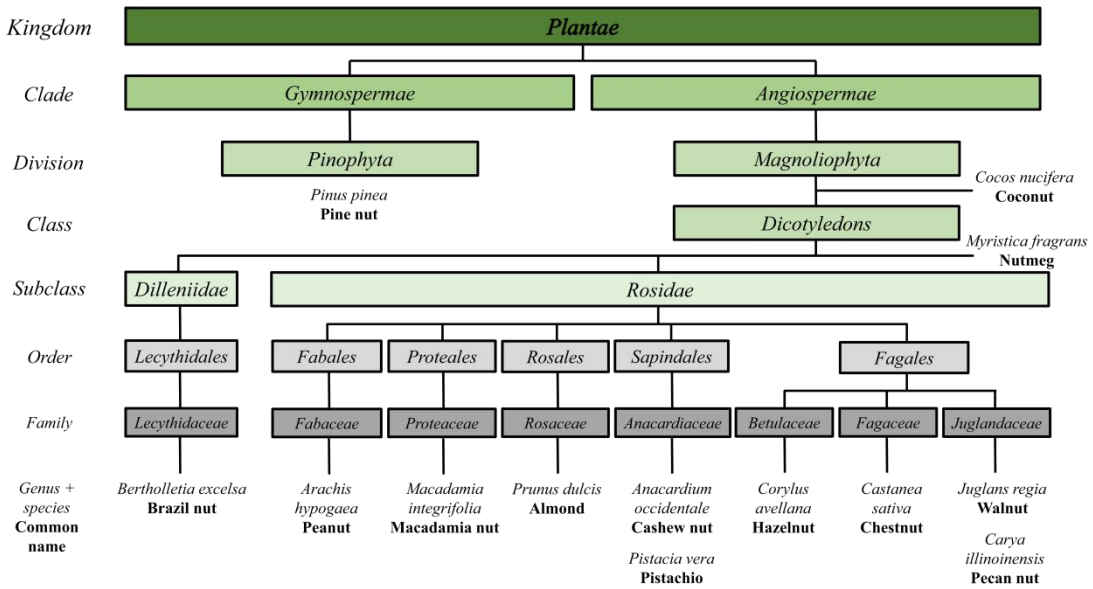


Fig. 1 Botanical classification of nuts adapted from Jacquenet and Moneret-Vautrin (2007) and according to the Angiosperm Phylogeny Group (APG III 2009)

Table 1 Short description of species of the genus *Juglans* L. adapted from Germain et al. 1999 and other sources

Section		Botanical characteristics	Species	Vernacular name	Main information
Common walnut	Dioscaryon	Leaves with small number of few or not dentate and hairless leaflets (7/9), absence of lacunae in the nut shell with thin primary and secondary walls, dehiscent husk when mature	*Juglans regia* L.	Persian or English walnut	Cultivated for nut and wood in the northern hemisphere (Europe, Asia, North America) and the southern hemisphere (Australia, New Zealand, South Africa, South America)
			Juglans sigillata Dode	Iron walnut	Not easily distinguished from *J. regia*, native to China, few information but nut is oval-shaped and tree used for its wood, commonly found in Yunnan province in China
Black walnut	Rhysocaryon	About 16 species from the Americas, leaves with dentate leaflets (9/25), nut with thick shell and presence of lacunae with well-developed and lignified primary and secondary walls, not dehiscent hairless husk or slightly pubescent	*Juglans nigra* L.	Eastern black walnut	Spontaneous in Eastern USA, introduced in Europe at the beginning of the XVIIth century for ornamental use and then for rootstock, bark and wood darker than *J. regia*, 25 to 40-m high tree, 13/23 leaflets lanceolate/narrow/elongated/finely dentate at their sides, nuts regularly larger than longer and slightly flattened, presence of black lacunae in the nut shell and difficult to break, pronounced taste appreciated in Northern America for ice creams and cakes, more resistant to extreme cold than *J. regia* but more susceptible to spring frosts (early budding), tolerant to *Armillaria mellea* rot, to ink disease *Phytophthora cinnamomi* and relatively not susceptible to crown gall *Agrobacterium tumefaciens.* Now used for wood and nut is correct
			Juglans hindsii Jeps.	Northern California black walnut or Hinds' black walnut	Native to Northern California, few in France, 20 to 30-m high tree, 15/21 leaflets lanceolate/narrow/finely dentate, green-blue, almost smooth nut, early budding, late flowering, appreciated for its wood, widely used in California as rootstock ("Paradox" rootstock is a hybrid from *J. hindsii* pollinated by an English walnut), tolerant to rot but susceptible to crown gall, also for ornamental use with small edible nut
			Juglans microcarpa Berl.	Texas or little black walnut	Native to Texas and Northern Mexico, semi-bushy, 10 m high at most, globulous nuts with thick shell and presence of lacunae, very furrowed, distinguished from other black walnuts by its small size (1.5 cm diameter), 15/23 leaflets dentate/small/narrow/lanceolate, lower side oftenly fluffy, can interbreed with *J. major* and *J. nigra*
			Juglans californica S. Wats.	Southern California black walnut	

Table 1 (continued)

Section	Botanical characteristics	Species	Vernacular name	Main information
				Native to Southern California, 11/15 leaflets small/lanceolate/narrow like *J. microcarpia* but hairless on lower side even juvenile, small globulous nuts, no more than 2.5 cm diameter, susceptible to roots diseases so not used as rootstock, ornamental use
		Juglans major Heller	Arizona black walnut	Native to Southwestern USA and Northern Mexico, up to 20 m high, stout trunk, distinguished from *J. microcarpa* by bigger nuts and larger leaflets fewer in number (9/15) more coarsely dentate, nuts with 2/3 cm diameter, less furrowed than *J. nigra*, less cracked bark, leaves slightly green-blue, used as rootstock as *J. nigra*, edible nuts and used as carpentry wood and firewood
		Juglans australis Griseb. (*J. brasiliensis* Dode)	Argentine or Brazilian walnut	Native to forests of Argentina and Bolivia, up to 20 m high, more resistant to frost than *J. regia*, edible nuts
		Juglans boliviana (C. DC.) Dode	Bolivian or Peruvian walnut	Native to Northern Bolivia and Central and Southern Peru, few information
		Juglans friburgensis		Native to Breisgau, area in Southwest Germany; few information
		Juglans hirsuta Mann.	Nuevo León walnut	Native to Mexico, 14/23 lanceolate leaflets, few information
		Juglans jamaicensis C. DC. (*J. insularis* Griseb.)	West Indies walnut	Found in Cuba, Dominican Republic, Haiti, and Puerto Rico but not native to Jamaica; can reach 25 m high; edible nuts of 2/3 cm diameter and wood used like *J. nigra*
		Juglans mollis Engelm.	Mexican walnut	Native to Mexico, up to 15 m high, edible nuts, few information
		Juglans neotropica Diels	Andean, Ecuadorian or Colombian walnut	Found in Colombia, Ecuador, and Peru; can reach 40 m high; reddish-brown bark; oval-shaped canopy; large leaves (40 cm); nuts eaten in local Ecuadorian market; more difficult to crack than *J. regia*; one of the most expensive wood because of its durability (floor and decoration)
		Juglans olanchana Standl. & L.O. Williams	Cedro Negro	Found in Costa Rica, Guatemala, El Salvador, Honduras, Mexico, and Nicaragua; can reach 40 m high; darker leaves on upper side; husk used to color leather; brown wood easy to work used for fine furniture and lute-making; edible nuts
		Juglans soratensis Mann.		Native to Bolivia, few information

Table 1 (continued)

Section		Botanical characteristics	Species	Vernacular name	Main information
			Juglans steyermarkii Mann.	Guatemalan walnut	Only found in Huehuetenango department of Guatemala, can reach 17 m high, few information
			Juglans venezuelensis Mann.	Venezuelan walnut	Native to Venezuela, small tree, 2 known wild populations of less than 100 individuals, few information
White walnut	Trachycaryon	Distinguished from black walnut by a more pubescent and viscous husk, nut shell does not have secondary wall, nuts are in clusters	*Juglans cinerea* L.	Butternut	Native to Northeastern USA, near to Canada, spontaneous until Minnesota and Arkansas, spreading shrub up to 20 m high, light gray bark and very cracked when mature, susceptible to sunburn, wood of lower grade than common and black walnut, very cropper, early budding, poor compatibility with *J. regia*, lengthened nut with thick shell, 11/17 lanceolate and dentate leaflets
–	Cardiocaryon	From the Far East, like *J. cinerea*: nuts in clusters without secondary wall, in pubescent and viscous husk but not dehiscent. Distinguished from *J. cinerea* by leaflets sparsely dentate and cordate leaf scar whose upper edge is concave	*Juglans sieboldiana* Maxim. (*J. ailantifolia* Carr.)	Japanese walnut	Native to Japan where it is abundant in forests, 15/20 m high, light gray bark, poor compatibility with *J. regia*, very cropper, early budding, nuts in cluster of 12/20, ovoid shape, pointed top, smooth shell, thick primary wall with lacunae, 11/17 oblong leaflets and few dentate, pubescent on lower side
			Juglans cathayensis Dode	Chinese walnut	Native to Central and Eastern China, spreading shrub up to 25 m high, used as rootstock in China, 9/17 acuminate leaflets, pubescent on lower side with glandulous petiole, nuts in cluster of 6/10, sherical to ovoid shape, thick smooth shell
			Juglans mandshurica Maxim.	Manchurian walnut	Native to Manchuria and Korea, found in Amur river valley, resistant to extreme cold, used as rootstock in cold areas of Northern China, spreading shrub, light gray bark, 30 m high, used for its wood, leaves until 90 cm, 9/17 oblong and acuminate leaflets, few dentate, pubescent on lower side, distinguished from *J. sieboldiana* by nuts in small cluster of 5/7, rough and thick shell

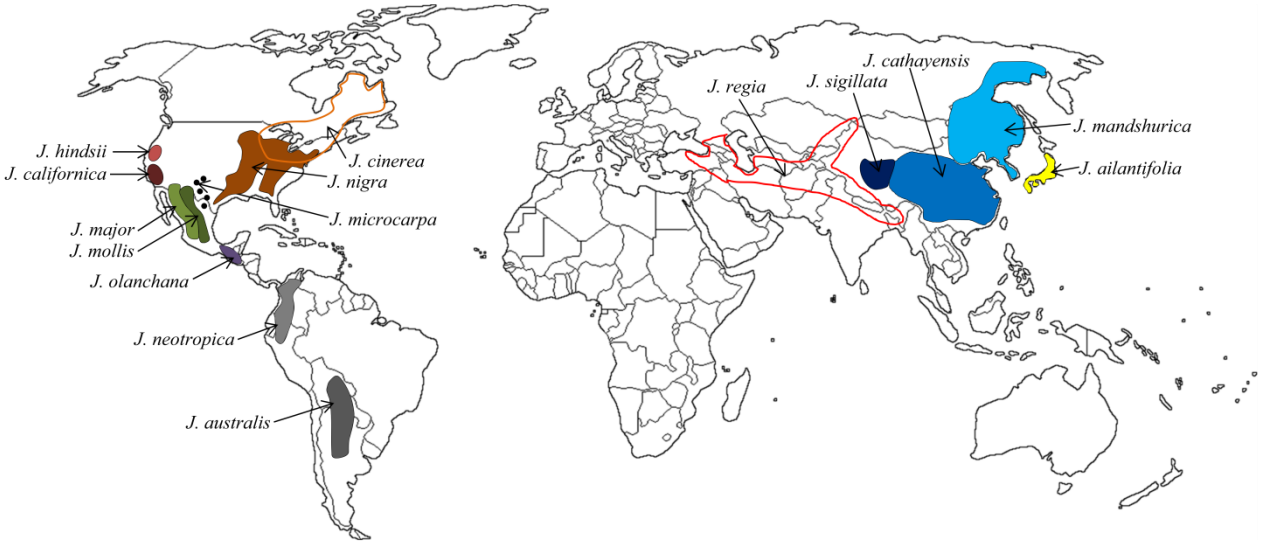


Fig. 2 Simplified distribution of *Juglans* taxa adapted from Aradhya et al. (2007) and Germain et al. (1999). Distribution of cultivated species *J. regia* extends beyond its natural range

eastern region of Turkey (Dode 1909; Forde 1975) and reached east into Northern India, Nepal, and Western China where it is still possible to find wild forests (Leslie and McGranahan 1988). Although Persian walnut was introduced into Europe and the Balkans before the last glacial period (~ 50.000 ya) (Carrion and Sanchez-Gomez 1992), it is still debated how trees have reached Europe. In relation to that, a very recent study focused on the history of its spread across Europe has integrated various data such as fossil pollen and microsatellite markers (Pollegioni et al. 2017) and came to the conclusion that natural populations existed before human intervention. This piece of work also confirmed the presence of walnut in glacial refugia in Western Europe and in the Balkans. This is why it would appear that human-mediated dispersal between Asia and the Balkans started in the Early Bronze Age. Another recent study confirmed this hypothesis and in addition showed that the highest genetic diversity of Persian walnut is found in South Asia, suggesting that this region is probably the origin of Persian walnut (Roor et al. 2017). Alternatively, Cro-Magnon people could have eaten ancestral forms of Persian walnuts (~ 17.000 ya) as archeological evidence found in the Perigord French region suggests. During the time of Ancient Greece, Persian walnuts became dispersed and more abundant thanks to Alexander the Great (~ 350 B.C.) and increased human dispersal of walnuts into Portugal; Spain and North Africa took place during the Roman Empire (~ 27 B.C. to ~ 395 A.D.). Evidence suggests human-mediated dispersal during the Roman Empire between Western Europe and the Balkans in Eastern Europe (Pollegioni et al. 2017). In the meantime, Chinese archives suggest a form of domestication of walnut in Southern Tibetan and the Yunnan region during the Han dynasty (~ 206 B.C. to ~ 220 A.D.) (De Candolle 1885). In the sixteenth century, the conquest of the New World enabled the spread of Persian walnut into South America, particularly Chile, before introduction in its present form to California in the nineteenth century (Germain et al. 1999).

Biology, phenology, and physiology

As described by Germain et al. (1999), walnut trees are wind-pollinated monoecious and dichogamous plants. Walnut trees are self-compatible but self-fertilization is limited due to the dichogamy, so it is often necessary to include a pollinizer genotype in orchards. Male inflorescences each consist of 100 to 160 flowers formed into a catkin that can produce around two million pollen grains. Female inflorescences of most walnut species, including Persian walnut, develop as individual flowers or most commonly as two or three per inflorescence. Female flowers of species in section *Cardiocaryon* are borne in racemes of approximately 7–19 flowers each. There are two main types of fruiting: fruiting only at the terminal position on new branches or fruiting at both terminal and lateral positions on shoots (fruiting all along branches). Rather rare, this last fruiting type was originally introduced into California by the cultivar "Payne." Walnut fruit consists of a fleshy green husk surrounding a nucleus of the walnut itself, composed of a shell containing the kernel.

There is evidence that tree phenology responds to climate changes: earlier spring budburst and later autumn senescence have extended the growing season by approximately 11 days since the 1960s at mid to high latitudes (Menzel and Fabian 1999). However, trees need to have completed a series of processes before the first frost arrives or they risk damage (Way 2011). In walnut, buds enter dormancy from mid-

August to early October and under temperate climate conditions; chilling requirements necessary for breaking dormancy are normally met by early January (Germain et al. 1999). Budburst dates vary according to genotype and climatic conditions, but under French climate conditions, leafing occurs between mid-March and early June. Fruit maturity typically occurs in September and October (Germain et al. 1999). A study of walnut phenology, conducted for two winters at both lowland and mountain locations, found no significant differences in chilling requirements among genotypes differing in budburst dates, but heat requirements were significantly different (Charrier et al. 2011). This means that from September until January, the acclimation was driven mainly by environmental factors, but from January until budburst, a prevalent genotype effect was found.

Production and nutritional aspects

Among the nut crops, almonds had the greatest production until 2008, then the walnut took over until 2013 (Fig. 3). During the 1973–2014 period, the five largest walnut producers were China, the USA, Iran, Turkey, and Ukraine with at the leading front the USA between 1973 and 1983, and China nowadays. In 2014, in-shell walnut production was as follows: (1) China, 1602 kt; (2) The USA, 518 kt; (3) Iran, 446 kt; (4) Turkey, 181 kt; (5) Mexico, 126 kt. France was in 10th position in 2014 with 34 kt. Regarding nutritional aspects, walnut is highly caloric with 654 kcal per 100 g and with high amount of proteins (15.2 g), lipids (65.2 g), carbohydrates (13.7 g), and micronutrients (US Department of Agriculture, Agricultural Research Service, Nutrient Data Laboratory 2015). Walnuts are particularly rich in unsaturated fatty acids, including oleic, linoleic, and linolenic acids (Maguire et al. 2004).

A study published by Martínez et al. (2010) focused on the biochemistry of walnut and highlights the nutritional importance of the oil made up of tocopherols, phospholipids, sphingolipids, sterols, and phenolic compounds. In addition, antioxidant and anticancer activities of walnut protein hydrolysates show possible benefits for prevention of breast and colon cancer (Jahanbani et al. 2016). Another recent review focused on the phytochemical profiles and biological activity of *J. regia* components and discusses the current high interest in isolation and identification of these active compounds and screening them for pharmacological activities (Panth et al. 2016). Concerning the effects on blood lipids, walnut consumption led to a decrease in low-density lipoprotein (LDL) cholesterol with no change in high-density lipoprotein cholesterol (Banel and Hu 2009). Taha and Al-Wadaan (2011) also reviewed the activities of walnut constituents, and similar observations have been published for tree nuts in general (Bolling et al. 2010).

Morphological studies as preliminary breeding approaches

Plant breeding has always impacted food production and played a vital role in improving human nutrition (Tester and Langridge 2010). However, this has also increased uniformity within the world's agricultural crops, contributing to increased genetic vulnerability to biotic and abiotic stresses (Keneni et al. 2012). For these reasons, it is important to understand better the impacts of modern plant breeding on genetic diversity. In the same way, intelligent management of this diversity could be of valuable assistance to breeders. Germplasm diversity is commonly evaluated with the help of morphological descriptors. This is usually the first step in classifying and describing germplasm and in studying heritability of traits for a new breeding program and selection of superior genotypes (Smith and Smith 1989). For this purpose, the International Union for the Protection of New Varieties of Plants (UPOV) offers guidelines in determining distinctness, homogeneity, and stability of new cultivars for most plants. Thus, morphological studies provide direction for the choice

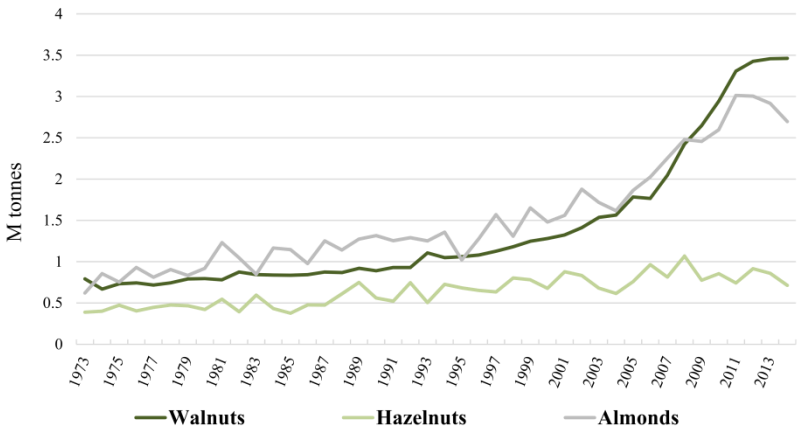


Fig. 3 World in-shell production of walnuts, hazelnuts, and almonds between 1973 and 2014 (FAOSTAT data 2016)

of cultivars suited to specific growing conditions (Solar and Stampar 2011).

For *J. regia*, many early morphological studies concerned the search of correlation between various horticultural traits (Sholokhov 1974; Komanich 1980; Sen 1985; Sharma 1996). While the relationship between yield level and type of dichogamy was found conflicting (Kornienko 1974; Majacka 1971), another study found no statistical difference between protogynous and protandrous cultivars on average yield and nut characteristics (Akça and Sen 1997). Then, a correlation was found between kernel weight and flowering type (Eskandari et al. 2005). Moreover, Amiri et al. (2010) found a strong correlation between lateral-bearing habit and yield in Iranian genotypes. However, lateral-bearing trees observed in Slovenia were reported to be more susceptible to walnut blight and cold (Solar et al. 2001) and susceptibility to walnut blight correlated with leaf abscission date (Abedi and Parvaneh 2016). Moreover, a negative correlation between yield and tree height was reported by Forde and McGranahan (1996), and a positive correlation was found between altitude and frost resistance, although altitude correlated negatively with yield (Atefi 1990).

All these traits are influenced by three components: genetic background, environment, and interaction between the two; it is of utmost importance to decipher the amount of contribution of each of these components. In Iran, an important study focused on the heritability of horticultural traits such as leafing and flowering times, fruit and kernel weights, kernel color, shell texture, etc. Broad-sense heritabilities (H^2) were higher than 60% among the genotypes tested, except for spring frost susceptibility and kernel percentage (Eskandari et al. 2005). A previous piece of work already focused on the narrow-sense heritabilities (h^2) for some traits and found many were high: leafing date 96%, female and male blooming dates > 80%, harvest date 85%, and fruit traits (shell thickness, nut and kernel weights) 85%. Conversely, yield and kernel color seemed to be non-heritable (Hansche et al. 1972).

Morphological studies can greatly assist programs in countries, such as Iran, using selection from existing natural seedling populations (Ghasemi et al. 2012). Promising genotypes there were selected by their phenological and pomological traits, including leafing date, lateral bearing, type of dichogamy (Ghasemi et al. 2012; Ebrahimi et al. 2015), bud break (Arzani et al. 2008), spring frost tolerance (Mahmoodi et al. 2016), nut yield, nut maturity, and nut characteristics (Haghjooyan et al. 2005; Arzani et al. 2008; Ahandani et al. 2014; Khadivi-Khub et al. 2015). Moreover, vigor of dwarf seedlings, which can be used to increase planting density, was also studied in Iran (Rezaee et al. 2009). A positive correlation was found between seedling height and other parameters, including seedling diameter, number of nodes, and length of internodes, suggesting that measurement of seedling height alone is sufficient for identification of dwarf genotypes.

Similar studies were performed in Turkey (Bayazit 2012). Superior genotypes were selected for yield, nut characteristics, and cold tolerance (Aslantaş 2006) and other promising trees were selected for late flowering and walnut blight and anthracnose tolerance (Yarilgac et al. 2001; Asma 2012; Fikret Balta et al. 2007; Karadag and Akça 2011). In Albania, evaluation of pomological and phenological traits of germplasm from the Dibra region found high variability in nut weight (Zeneli et al. 2005). In additional studies of this type, walnut cultivars were evaluated in Serbia (Miletic et al. 2003) as well as some foreign cultivars under the climatic conditions of south Bulgaria (Gandev and Dzhuvinov 2015). In India, differences between geographic and genetic diversity were identified using 15 nut traits (Sharma and Sharma 2001). In Italy, a study of a simulated seed orchard, planted with genotypes selected by morphology and phenology in order to preserve genetic resources, showed that it could have as much variability as a larger population (Ducci et al. 2010). In France, 2450 hybrids from 22 crosses between French and California varieties were studied. A correlation was found between leafing date of 1-year-old trees and the same parameter some years later on the same trees, showing that an early screening of this trait is possible (Germain 1990). Finally, the rate of apomixis was also investigated and variation among numerous European cultivars in Mediterranean climate was recorded (Asadian and Pieber 2005).

Historical overview of molecular marker development and use

Phenotypical studies are essential for a general description of an organism, and have been used since Darwin's time due to their low cost and wide application, for example for museum and fossil specimens. However, the arrival of molecular investigation methods in plant science research gives new insights in both selection processes and evolutionary studies. Indeed, when individuals are morphologically similar or indistinguishable, molecular studies can provide information based on their DNA and heritable variation (Hillis 1987). Molecular information seems less affected by environment than are morphological data, which often depict conflicting scenarios (Hillis 1987). On top of this, morphological markers can affect one another and make difficult to predict pleiotropic effects (Andersen and Lübberstedt 2003). As a consequence, nowadays, there is a debate on the importance and the future of morphological studies, especially in the context of phylogenetic systematics (Jenner 2004). Application of a wide range of molecular markers has been the focus of many walnut studies during the past 30 years. The following sections depict an overview of these molecular studies and their application (Table 2).

Table 2 Applications of molecular markers in walnut studies

Molecular markers	Applications	References
Isozymes	Inheritance evaluation	Arulsekar et al. 1986
	Intergeneric hybrids identification	McGranahan et al. 1986
	Genotype origin determination	Cheng and Yang 1987
	Genetic diversity/relationships determination	Malvolti et al. 1993, 1994; Fornari et al. 2001; Busov et al. 2002; Ninot and Aletà 2003
	Genotype characterization	Aletà et al. 1990; Solar et al. 1994; Vyas et al. 2003
isozymes + RFLPs	Somatic embryo origin determination	Aly et al. 1992
RFLPs	Genetic diversity/relationships determination	Fjellstrom et al. 1994; Fjellstrom and Parfitt 1994b
	Genetic mapping	Fjellstrom and Parfitt 1994a
RFLPs + RAPDs	Introgression evaluation	Woeste et al. 1998
	Genetic mapping	Woeste et al. 1996b
RAPDs	Interspecific hybrids identification	Ross-Davis et al. 2008; Emilia et al. 1997
	Biotic stress resistance	Woeste et al. 1996a
	Phenology/pomology evaluation	Keqiang et al. 2002; Li et al. 2007
	Genetic diversity/relationships determination	Nicese et al. 1998; Fatahi et al. 2010; Qianwen et al. 2010; Pop et al. 2010; Erturk and Dalkilic 2011
	Genetic mapping	Malvolti et al. 2001
ISSRs	Genotype origin determination	Malvolti et al. 2010
	Genetic diversity/relationships determination	Potter et al. 2002; Christopoulos et al. 2010; Ji et al. 2014; Pollegioni et al. 2003
AFLPs	Homogamous/dichogamous genetic diversity comparison	Sütyemez 2006
	Biotic stress investigation	Loreti et al. 2001; Mozaffarian et al. 2008
	Genetic diversity/relationships determination	Kafkas et al. 2005; Bayazit et al. 2007; Chen et al. 2008; Chen et al. 2009; Qing Guo et al. 2010; Wang et al. 2010; Ma et al. 2011; Ali et al. 2016
	Genotype characterization	He et al. 2010
AFLPs + RAPDs	Genetic diversity/relationships determination	Xu et al. 2012b
SSRs	Markers discovery	Woeste et al. 2002; Dangl et al. 2005; Hoban et al. 2008; Zhang et al. 2010; Qi et al. 2011; Yi et al. 2011; Wu et al. 2012; Zhang et al. 2013; Chen et al. 2013; Najafi et al. 2014; Topçu et al. 2015; Ikhsan et al. 2016; Dang et al. 2015; Hu et al. 2015; Dang et al. 2016
	Interspecific hybrids identification	Pollegioni et al. 2008, 2009
	Genetic diversity/relationships determination	Foroni et al. 2005; Dangl et al. 2005; Victory et al. 2006; Robichaud et al. 2006; Wang et al. 2008; Aradhya et al. 2009; Gunn et al. 2010; Karimi et al. 2010; Bai et al. 2010; Mohsenipoor et al. 2010; Ebrahimi et al. 2011; Ruiz-Garcia et al. 2011; Pollegioni et al. 2011; Qi et al. 2011; Grauke et al. 2012; Mahmoodi et al. 2013; Zhang et al. 2013; Pop et al. 2013; Najafi et al. 2014; Dang et al. 2015; Hu et al. 2015; Han et al. 2016; Noor Shah et al. 2016; Ćelepirović et al. 2016; Wang et al. 2016; Vischi et al. 2017; Ebrahimi et al. 2017
	Origin/history	Pollegioni et al. 2017; Roor et al. 2017
	Genotype characterization	Foroni et al. 2007
SNPs	Markers discovery	Liao et al. 2014
	Biotic stress resistance	Zhu et al. 2015
	Biotic stress investigation	Zerillo et al. 2014
	Synteny analysis	Luo et al. 2015
	Genotype characterization	Ciarmiello et al. 2011
	Genetic mapping	You et al. 2012

From isozyme markers in late 1980s to AFLPs markers in early 2000s

In the late 1980s, knowledge of fruit trees at the genetic level was poor but a major breakthrough was the development of enzyme-based markers using isozyme electrophoresis (Hunter and Markert 1957). Identification and characterization of isozymic genes in *J. regia*, such as that of phosphoglucomutase, was first reported by Arulsekar et al. (1986). Then, intergeneric hybrids between *Pterocarya* spp. and *J. regia* were identified based on isozyme variation (McGranahan et al. 1986) and the origin of different walnut types in China and Tibet was examined (Cheng and Yang 1987). Moreover, genetic diversity in Italian *J. regia* populations was reported (Malvolti et al. 1993) and in Slovenia, a study showed that pollen enzymes have more variability than those of leaves (Solar et al. 1994). Furthermore, a study of leaf and fruit morphology and isozyme variation in populations of *J. regia* from central Italy showed an exclusive correlation between genotype and leaf morphology and between leaf morphology and fruit morphology. This suggests a chain of causal relationships between genotype and the two studied aspects of phenotype (Malvolti et al. 1994). Other studies have used isozymes to analyze genetic diversity/relationships in *J. regia* (Fornari et al. 2001; Ninot and Aletà 2003), *J. nigra* (Busov et al. 2002), or for cultivar identification (Aletà et al. 1990; Vyas et al. 2003).

Molecular genotyping constitutes a valuable efficient method to screen germplasm collections and facilitate the use of these resources by breeders and researchers (Mason et al. 2015). The emergence of DNA manipulation techniques drove the shift from enzyme-based to DNA-based markers in the early 1990s (Schlötterer 2004). DNA markers are less limited in number, and there is no need to wait for a particular development stage of the plant, making early scoring possible (Andersen and Lübberstedt 2003). In *J. regia*, a study combining isozyme and restriction fragment length polymorphism markers (RFLPs) (Botstein et al. 1980) was performed in order to determine the origin (zygotic or maternal) of somatic embryos derived from ovule tissues (Aly et al. 1992). It led to a fast parentage determination of plants regenerated in vitro. Then, a California team conducted three studies in 1994 using RFLPs. The first used unweighted pair group method with arithmetic mean (UPGMA) cluster analysis (Nei 1972) to highlight that California germplasm was associated with that of France and Iran and had less similarity to that of Nepal, China, and Japan (Fjellstrom et al. 1994). A second study focused on analysis of an interspecific backcross [*J. hindsii* × *J. regia*] × *J. regia* created in order to construct the first linkage map in walnut, which included 12 linkage groups (LGs) covering 1660 cM (Fjellstrom and Parfitt 1994a). The third study concentrated on determination of taxonomic relationships between different *Juglans* species (Fjellstrom and Parfitt 1994b) and indicated that *J. cinerea* should be included in *Cardiocaryon* section and not as the unique section *Trachycaryon*, as suggested in existing literature (Manning 1978).

The main limitations of the RFLP technique are the low level of polymorphism, the complexity of its use (Wani et al. 2010), the lack of automation, and the large amount of pure DNA needed (Schlötterer 2004). Hence, new strategies based on polymerase chain reaction (PCR) were developed and among these were randomly amplified polymorphic DNAs (RAPDs) (Williams et al. 1990) and inter-simple sequence repeats (ISSRs) (Zietkiewicz et al. 1994). A California team identified RAPDs in a backcross population [*J. hindsii* × *J. regia*] × *J. regia* and created a revised genetic map of walnut that also included the set of RFLPs published earlier by Fjellstrom and Parfitt (1994a). This work expanded the number of identified LGs from 12 to 15, still short of the 16 total LGs present in walnut (Woeste et al. 1996b). Another study by this team identified and characterized an RAPD marker in walnut linked to the hypersensitivity response to cherry leafroll virus, which causes walnut blackline disease, a fatal necrosis of the graft union (Woeste et al. 1996a). Examination of another walnut backcross population {[*J. hindsii* × *J. regia*] × *J. regia*} found a low correlation between genetic and morphological distance during introgression (Woeste et al. 1998), suggesting that morphology-based selection during introgression is not an effective method for recovering the parent genome. RAPD markers also were used to study the genetic relatedness among *J. regia* cultivars in the collection of the breeding program at the University of California, including "Chandler," "Payne," "Meylan," and "Franquette" (Nicese et al. 1998). This led to the identification of two separate groups: (1) Asian genotypes and (2) California and French cultivars. Then, an additional linkage map was created using RAPDs and a *J. regia* intraspecific cross but the mapping efficiency was lower than with the interspecific cross (Malvolti et al. 2001). Phenological and pomological traits were also investigated using RAPD analysis. One marker was found to be related to early bearing trait (Keqiang et al. 2002), and one sequence-characterized amplified region (SCAR) marker derived from an RAPD was correlated with shell thickness (Li et al. 2007). In Italy, ISSRs were used successfully to differentiate varieties coming from the north or the south of the country (Pollegioni et al. 2003). An additional study was carried out using *Juglans* × *quadrangulata* (Carr.) Rehd. (pro sp.), a hybrid between *J. cinerea* (butternut) and *J. regia* (Ross-Davis et al. 2008). These hybrids are vigorous and more resistant to butternut canker disease, and RAPDs differentiating pure butternut from hybrids were found. Hybrids between *J. nigra* and *J. regia* were also differentiated using RAPDs (Emilia et al. 1997). RAPDs and ISSRs have been used since by Iranian (Fatahi et al. 2010), Romanian (Pop et al. 2010), Chinese (Qianwen et al. 2010;

Ji et al. 2014), Greek (Christopoulos et al. 2010), and Turkish (Erturk and Dalkilic 2011) teams to characterize their own local *J. regia* germplasm. In addition, the genetic relationships among *J. regia* cultivars from California, Europe, and Asia were examined using ISSRs but results showed that the ISSR method is limited for the study of germplasm coming from different geographic regions (Potter et al. 2002). Finally, a multidisciplinary approach was tried in Italy by combining the use of ISSRs, agronomic measures on fruit (diameter, form, dry weight), and chemical composition of cultivars from different areas, in order to determine genotype provenance. ISSRs showed a strong ability to discriminate walnut provenances, and the genotyping results were partly confirmed by the morphological and biochemical analysis of fruits (Malvolti et al. 2010).

Amplified fragment length polymorphism markers (AFLPs) (Vos et al. 1995) have been used in a number of studies. AFLPs were used to investigate the genetic diversity of the causal agent of walnut blight *Xanthomonas arboricola* pv. *juglandis*, one of the most important diseases of *J. regia* (Loreti et al. 2001). In this analysis, 66 isolates of *X. arboricola* pv. *juglandis* from different countries were characterized. AFLPs were also used to examine variation in populations of host-associated carob moth (*Ectomyelois ceratoniae*) on pomegranate, pistachio, and walnut in Iran (Mozaffarian et al. 2008). In addition, AFLPs were used in determining that the genetic diversity is higher in homogamous than in dichogamous genotypes, due to a higher level of heterozygosity (Sütyemez 2006). Then, Turkish *J. regia* trees were characterized using AFLPs in order to select superior genotypes (Kafkas et al. 2005) and AFLPs were used to study the genetic diversity among Turkish genotypes with a low chill requirement (Bayazit et al. 2007). In China, wild *J. regia* and *J. sigillata* populations from Sichuan province were studied using AFLPs to understand the genetic relationships between these two economically important species in the province (Chen et al. 2008, 2009). Other Chinese teams used AFLPs to fingerprint walnut cultivars (He et al. 2010; Ma et al. 2011; Xu et al. 2012b) and to study the genetic diversity of the early fruiting trait (Qing Guo et al. 2010). Eight *Juglans* species and one genotype of *Pterocarya stenoptera* were also studied in China with AFLPs to understand their genetic relationships (Wang et al. 2010). Finally, five walnut populations from the Kurdistan region of Iraq were analyzed using AFLPs and their genetic diversity was found to be low (Ali et al. 2016).

Peak of molecular studies with SSRs from the late 2000s

Simple sequence repeats (SSRs) (Morgante and Olivieri 1993), also known as microsatellites, are very powerful and informative markers for fingerprinting genotypes and determining genetic relationships, due to their abundance, co-dominant status, high reproducibility (Wani et al. 2010), and high polymorphism (Singh et al. 2008).

SSR markers have been employed for numerous walnut studies (Table 3). The first walnut SSR study was performed on *J. nigra* accessions to fingerprint clones with high-quality timber and edible nuts (Woeste et al. 2002). Two additional studies using SSRs on *J. nigra* showed a low genetic diversity among several populations from the central hardwood region of the USA (Victory et al. 2006) and identified parentage among progenies of several families (Robichaud et al. 2006). Another study characterized *J. nigra*, *J. regia*, and natural hybrids between the two species *Juglans* × *intermedia* (Carr.) with SSR markers (Pollegioni et al. 2008). Identification of parental species was found possible in a mixed Italian population, and this work confirmed the transferability of *J. nigra* SSRs to *J. regia*. A similar study focused on identification of parents that can easily hybridize for production of hybrid progeny, especially those characterized by increased vigor and a better wood quality (Pollegioni et al. 2009).

In China, genetic diversity of *J. regia* and *J. sigillata* populations, the two main cultivated species, was assessed using *J. nigra* SSR markers and a UPGMA dendrogram showed a geographic gradient rather than the conventional taxonomic classification (Wang et al. 2008). The results suggest that the genetic relatedness between the two species is strong, information relevant to determining if they are different species or merely different ecotypes. Moreover, a different study of the genetic diversity of *J. regia* and *J. sigillata* also using SSRs was directed at understanding the role of human impacts and biodiversity of walnut in Yunnan area in China (Gunn et al. 2010). The marker data showed that the two species were indistinguishable, and the authors concluded with two scenarios: *J. regia* and *J. sigillata* could be incipient species or they are distinct species that exchange genes. Next, expressed sequence tag (EST)-SSRs were developed and used to study trees of *Juglans* sections *Rhysocaryon*, *Cardiocaryon*, *Dioscaryon*, and *Carya illinoinensis* coming from the USA and Asia (Qi et al. 2011). The resulting UPGMA dendrogram clustered the samples according to conventional taxonomy, with *J. regia* and *J. sigillata* well separated, and proved the good transferability of EST-SSRs among *Juglans* species.

Other *Juglans* species, such as natural populations of *J. mandshurica* in northern and northeastern China, were also analyzed using SSRs (Bai et al. 2010). The results, which also included cpDNA fragments, showed that two independent refugia were maintained in northern China during the last glacial period, contrary to numerous other temperate forest trees which migrated to southern China. Then, newly developed ESTs from *J. regia* were used to develop SSR primers for *J. nigra* and *Carya* spp. investigation and many of these markers were found to be transferable with a high level of

Table 3 Microsatellite marker development in walnut

Authors and year	Generation source	*Juglans* species	Number of markers	Markers name
Woeste et al. 2002	Enriched (GA/CT)n SSR library from genomic DNA	*J. nigra*	30	WGA002, 004, 006, 007, 011, 017, 024, 025, 027, 032, 033, 042, 045, 047, 053, 054. 056, 058, 060, 065, 069, 070, 071, 072, 073, 074, 078, 079, 080, 082
Dangl et al. 2005	Enriched (GA/CT)n SSR library from genomic DNA	*J. nigra*	12	WGA001, 009, 089, 118, 202, 225, 276, 321, 331, 332, 349, 376
Hoban et al. 2008	Enriched (CA/GA/TAG)n SSR library from genomic DNA	*J cinerea*	13	jcin_A5, B110, B112, B114, B12, B121, B147, B157, B159, B212, B249, B262, B264
Zhang et al. 2010	5025 EST sequences from the NCBI database	*J. regia*	41	Contig_5, 40, 62, 104, 156, 216, 259, 268, 269, 352, 407, 541, 562-2, 566, 610, 642, 718, 719, 721, 795, 870, 1000, 1172, 1396, 1458, 1464, 1483, 1528, 1529, 1552, 1625, 1626, 1631, 1632, 1656, 1681, 1682, 1687, 1692, 1693, 1712
Qi et al. 2011	5025 EST sequences from the NCBI database	*J. regia*	18	ZMZ3, 5, 6, 11, 13, 14, 22, 26, 27, 30, 31, 34, 35, 44, 45, 46, and ZY1, 16
Yi et al. 2011	5213 EST sequences from the NCBI database	*J. regia*	30	/
Wu et al. 2012	48,218 BAC-end sequences from the NCBI database	*J. regia*	13,675	Available for downloading at http://walnutgenome.ucdavis.edu/
Zhang et al. 2013	13,559 EST sequences from the NCBI database	*J. hindsii* x *J. regia*	76	BFU-Jr10, 19, 29, 38, 40, 41, 42, 45, 46, 48, 49, 52, 62, 67, 68, 69, 70, 73, 82, 95, 97, 98, 99, 102, 104, 116, 121, 126, 130, 132, 144, 150, 152, 154, 161, 163, 167, 171, 179, 181, 182, 183, 184, 185, 195, 196, 201, 207, 217, 219, 228, 230, 233, 238, 240, 243, 245, 247, 249, 250, 254, 256, 263, 275, 276, 277, 284, 286, 288, 292, 298, 301, 306, 307, 319, 321
Chen et al. 2013	Fast isolation by AFLP of sequences containing repeats (FIASCO) approach	*J. mandshurica*	20	Pl, 3, 5, 8, 12, 14, 16, 18, 21, 23, 26, 28, 29, 30, 31, 35, 41, 44, 45, 47
Najafi et al. 2014	Fast isolation by AFLP of sequences containing repeats (FIASCO) approach	*J. regia*	13	ABRII/Jr1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13
Topçu et al. 2015	Enriched (CA/GA/AAC/AAG)n SSR library from genomic DNA	*J. regia*	185	46 SSRs from CA enriched library named CUJRAxxx, 80 SSRs from GA enriched library named CUJRBxxx, 14 SSRs from AAC enriched library named CUJRCxxx, and 45 SSRs from AAG enriched library named CUJRDxxx
Dang et al. 2015	Next-generation sequencing of transcriptome	*J. cathayensis*	12	JC6576, 9827, 4088, 4535, 3412, 2995, 7329, 5411, 8125, 1151, 3304, 4833
Hu et al. 2015	Next-generation sequencing of transcriptome	*J. hopeiensis*	25	JH6514, 2678, 9978, 2753, 2751, 8883, 1908, 6876, 6044, 0405, 8061, 1424, 9127, 9543, 2096, 4093, 4548, 1195, 0608, 8168, 0190, 2576, 9664, 4114, 5908
Ikhsan et al. 2016	48,218 BAC-end sequences from the NCBI database	*J. regia*	307	/
Dang et al. 2016	Next-generation sequencing of transcriptome	*J. regia*	39	JR0082, 0160, 1165, 1739, 1817, 2018, 2465, 2510, 2600, 2873, 3147, 3434, 3608, 3773, 4051, 4324, 4616, 4964, 4965, 4965.2, 5538, 5574, 6160, 6226, 6439, 6508, 6638, 6714, 6742, 6926, 7171, 7363, 7495, 7495.2, 7544, 8058, 8815, 9306, 9632

polymorphism (Zhang et al. 2013). Finally, *J. nigra*, *J. major*, and *J. microcarpa* individuals from natural population from Texas were investigated using plastid SSR markers suggesting that interspecific hybridization has occurred within these species. Indeed, many accessions considered to be *J. major* were hybrids with *J. microcarpa* (Grauke et al. 2012).

In *J. regia*, the most important Italian landrace "Sorrento" was fingerprinted with SSR markers (Foroni et al. 2005, 2007), as well as other Italian walnut resources, to discover and characterize long-term human impacts and frequent human-mediated domestication events controlling genetic diversity (Pollegioni et al. 2011). An analysis of the genetic structure in *J. regia* using SSRs revealed a high level of genetic diversity with five genetic groups linked to the centers of diversity of walnut in Eurasia (Aradhya et al. 2009). Later, SSR markers were also used in Iran to study the genetic diversity of genotypes (Karimi et al. 2010; Mohsenipoor et al. 2010; Ebrahimi et al. 2011; Mahmoodi et al. 2013; Najafi et al. 2014). Cultivars, mainly from Spain and the USA, were studied with SSR markers and data separated Spanish cultivars from those of California (Ruiz-Garcia et al. 2011). Additional studies using SSRs have focused on germplasm diversity and genetic relationships among Indian genotypes (Noor Shah et al. 2016), among trees from cold-temperate areas of the USA and Europe (Ebrahimi et al. 2017), and among germplasm in China (Han et al. 2016; Wang et al. 2016) and Romania (Pop et al. 2013). Finally, two recent papers have been published. The first focused on optimization of high-throughput multiplex PCR by combining 11 SSRs, particularly from the two first set of markers developed in *J. nigra*, in order to investigate the genetic diversity of Croatian germplasm (Ćelepirović et al. 2016). The other concerns natural populations of *J. regia* in the eastern Italian Alps where moderate genetic diversity was found (Vischi et al. 2017). These authors also claim that some SSR markers reported in the literature, especially from the two first set of *J. nigra* markers, were unable to provide reliable results with a lack of polymorphism, multilocus amplifications, and high frequency of null alleles.

Additional sets of SSR markers were developed in *J. regia* from bacterial artificial chromosome (BAC) end sequences (Wu et al. 2012; Ikhsan et al. 2016), from DNA genomic libraries with specific pattern repeats such as "AAC" or "AAG," improving the number of informative alleles (Topçu et al. 2015), and with EST sequences from the transcriptomes of many organs (Dang et al. 2016). These could potentially be useful in germplasm characterization and linkage mapping.

SNPs revolution from the early 2010s

Single nucleotide polymorphism (SNP) markers (Wang et al. 1998) present many advantages compared to the previous markers as they are highly abundant and are now the markers of choice for automated high-throughput analysis at moderate cost (Schlötterer 2004).

The initial use of SNP markers to characterize *J. regia* cultivars (Ciarmiello et al. 2011) employed the first and second internal transcribed spacer regions (ITS1 and ITS2) and 5.8S coding region of rDNA-repeat (ribosomal DNA) operon segment. Alignment of sequences revealed 244 SNPs and one short insertion-deletion (InDel). The phylogenetic analysis showed that these regions could be used for specific authentication of *J. regia* cultivars. Moreover, the development of a pipeline was completed in order to discover genome-wide SNPs in coding sequences. Twenty-two thousand seven hundred ninety-nine SNPs in "Chandler" were identified and 6000 SNPs were selected to construct a chip to be used to genotype a mapping population (You et al. 2012). Then, by partial genome pyrosequencing, 48,165 SNPs and 1037 InDels were found in *J. regia* (Liao et al. 2014). These sets of SNPs could provide great assistance for genetic diversity analysis, for understanding nucleotide variations in genome, for genetic map construction, and for other breeding applications such as marker-assisted selection. The causal agent of thousand cankers disease in the USA, *Geosmithia morbida*, was also investigated using SNP analysis of 209 isolates in a population genetic study (Zerillo et al. 2014).

Finally, analysis with SNPs assessed synteny between walnut and several other plant species. These included three long-lived perennials (*Vitis vinifera*, *Populus trichocarpa*, and *Malus domestica*) and three short-lived herbs (*Cucumis sativus*, *Medicago truncatula*, and *Fragaria vesca*). Results showed that the walnut genome was closer to the three genomes of woody perennials than to those of the herbs, and that degree of synteny was linked to nucleotide substitution rates (Luo et al. 2015).

Genetic/physical mapping and QTL detection

Genetic maps (or linkage maps) indicate the position and relative genetic distances between markers along chromosomes, based on the principle that markers segregate during meiosis as a result of chromosome recombination (Paterson 1996). Genetic mapping is the first step in localization of genes or quantative trait loci (QTLs) associated with agronomic traits of interest. Genetic maps are desirable for breeding and useful in marker-assisted selection (MAS) (Semagn et al. 2006). A number of intraspecific and interspecific populations have been developed in walnut for genetic mapping construction and QTL detection (Table 4).

First genetic maps with RFLPs and RAPDs

The first genetic map in walnut was created with the use of RFLP markers. A progeny of 63 individuals from an

Table 4 Summary of genetic linkage maps for walnut

Species/country	Mapping population type	No./types of offspring	No./types of markers	No. of LG/genome size (cM)	References
[*J. hindsii* × *J. regia*] × *J. regia* /USA	"Paradox Mom" × "Hartley"	63/BC1	42 RFLP	12/1660	Fjellstrom and Parfitt 1994a
[*J. hindsii* × *J. regia*] × *J. regia* /USA	"Paradox Mom" × "Hartley"	49/BC1	59 RAPD, 48 RFLP	15/NA	Woeste et al. 1996b
J. regia/Italy	"Lara 480" × "Chandler 1036"	81/F1	120 RAPD, 4 isozymes	11 (female), 10 (male)/NA	Malvolti et al. 2001
J. regia/China	"Chandler" × "Idaho"	425/F1	1525 SNP	16/1049.5	Luo et al. 2015
J. regia/China	"Yuan Lin" × "Qing Lin"	84/F1	2577 SLAF	16/2457.8	Zhu et al. 2015

interspecific backcross of [*J. hindsii* × *J. regia*] × *J. regia* was investigated. This backcross was from an open-pollinated male-sterile *J. hindsii* × *J. regia* tree ("Paradox Mom") growing in a *J. regia* "Hartley" orchard. Forty-two RFLPs were assigned into 12 LGs, and the estimation of genome size indicated that the walnut genetic map spanned 1660 cM (Fjellstrom and Parfitt 1994a).

Two years later, the same set of progeny was investigated using 66 RAPDs. A new expanded genetic map was created including the previously found RFLPs along with these new RAPD markers. One hundred seven markers were assigned into 15 LGs (Woeste et al. 1996b).

Five years later, an intraspecific cross of *J. regia* was analyzed with both RAPD markers and isozymes. An F_1 progeny of 81 individuals from a *J. regia* "Lara 480" and *J. regia* "Chandler 1036" cross was screened with 120 RAPDs and four isozymes (Malvolti et al. 2001). Only parental maps were created with 11 LGs for the female parent map and 10 for the male parent map.

Next-generation mapping

A 6K SNP iSelect Infinium BeadChip was generated (You et al. 2012) and used, 4 years later, to genotype an F_1 progeny of 425 individuals from an intraspecific cross between "Chandler" and "Idaho" (*J. regia*). A total of 1525 SNPs were assigned into 16 LGs. The LG lengths ranged from 37.7 cM (LG15) to 97.3 cM (LG7), and the length of the genetic map was 1049.5 cM (Luo et al. 2015).

This genetic map was used to construct a walnut physical map with 15,203 exonic BAC-end sequences. This physical map should represent a large part of the genome because its estimated total length, 736.1 Mb, is close to the estimated size of the genome, 606 Mb (Luo et al. 2015). With this mapping work, it has also been possible to find two markers on LG11 which surround the location of the *LB1* gene involved in the lateral bearing phenotype (Dvorak et al. 2015).

Finally, another mapping population, an F_1 progeny of 84 individuals, was created by an intraspecific cross between "Yuan Lin" and "Qing Lin" (*J. regia*). Two thousand five hundred seventy-seven specific-locus amplified fragment (SLAF) markers of three types were used: 2300 SNP only (89.25%), 87 InDel only, and 190 SNP and InDel. They were assigned into 16 LGs, and the integrated genetic map length obtained was 2457.82 cM (Zhu et al. 2015). Moreover, a QTL linked to anthracnose (*Colletotrichum gloeosporioides*) resistance, an important trait in Chinese cultivars, was found on LG14. The QTL interval detected was from 165.51 to 176.33 cM on LG14 with 10 markers and their LOD scores ranging from 3.22 to 4.04.

Genomics and transcriptomics studies

Genomics breakthrough

Genomics is an area within genetics that involves the study of genes and their function, including DNA sequencing, and aims to understand the structure of the genome of an organism. A primary approach to genome analysis is the cloning of exogenous DNA into BAC vectors which can hold genomic DNA fragments up to 350 kb (Choi and Wing 2000). This technique provides a further opportunity for marker development. Using two BAC libraries of 129,024 clones from shoots of "Chandler" grown in vitro and based on 48,218 BAC-end sequences (Table 5) covering 31.2 Mb (5.1% of the genome), 11.5% of the genome was reported to represent coding sequences and a physical map was constructed (Wu et al. 2012).

The same year, 54,912 BAC-end sequences from "Chandler" were generated by Sanger sequencing to identify genome-wide SNPs and to construct the first 6K SNP iSelect Infinium BeadChip (You et al. 2012).

Two years later in China, high-throughput pyrosequencing with Roche technology was used to produce 541,176 reads and 31,362 contigs were assembled covering 15.1 Mb, from which SNPs and InDels were discovered (Liao et al. 2014). In addition, the high-resolution SLAF sequencing (Sun et al. 2013), based on reduced representation libraries (RRLs) and high-throughput sequencing (HTS), was performed in *J. regia* and used to construct a genetic map (Zhu et al. 2015).

Table 5 Summary of walnut genomics and transcriptomics information on the NCBI database (https://www.ncbi.nlm.nih.gov)

Species	Total nucleotide sequences	Nucleotides	EST	GSS	Notes/references
J. regia	223,823	169,18	5253	49,405	40 ESTs unpublished/1–188 ESTs involved in cellular regulation and global metabolism from leaf under irradiance or in darkness/2–5025 ESTs from seed coat tissues/3
J. hindsii × *J. regia*	13,660	101	13,559	0	2715 ESTs from leaves and 10,844 from roots inoculated with *Pratylenchus vulnus* nematodes/4
J. nigra	211	129	78	4	78 ESTs involved in the regulation of heartwood production, related to programmed cell death, from stems/5
J. cathayensis	44	32	12	0	12 ESTs from fresh leaves, buds, female flowers, and male flowers to study genetic diversity/6
All other taxa	601	601	0	0	
Total	238,339	170,043	18,902	49,409	

In another approach, genomic DNA libraries enriched for CA, GA, AAC, and AAG repeats were created in Turkey from "Maras-18" and 624 clones were sequenced (Topçu et al. 2015). The greatest number of alleles and highest levels of polymorphism observed for these *J. regia* were found in the GA-repeats enriched library.

Chloroplast and nuclear genome sequencing

Very important results recently published in walnut genomics are chloroplast genome sequencing (Hu et al. 2016) and nuclear genome sequencing (Martínez-García et al. 2016). The chloroplast genome sequence provides information about phylogeography, genetic diversity, and evolution. The *J. regia* tissue was sequenced using Illumina technology and resulted in the first complete *Juglandaceae* chloroplast genome sequence published (Hu et al. 2016). With a length of 160,367 bp and 36.11% GC content, 137 genes were annotated. Among them, there were 86 genes coding for proteins, three pseudogenes (two *ycf15* and one *infA*), 40 transfer RNA genes (tRNA), and eight ribosomal RNA genes (rRNA). A phylogenetic analysis was performed showing this genome is closely related to others of the *Fagaceae* family.

In 2017, chloroplast genomes of four other Chinese *Juglans* species were investigated (*J. sigillata*, *J. cathayensis*, *J. hopeiensis*, and *J. mandshurica*) by a combination of de novo and reference-guided (Hu et al. 2016—NCBI Accession number KT963008) assembly of whole chloroplast genomes (Hu et al. 2017). Chloroplast genome lengths ranged between 159,714 (*J. hopeiensis*) and 160,367 (*J. regia*) bp with 36.1% GC content and contained 129 genes each, 88 protein-coding genes, 40 tRNA genes, and eight rRNA genes. Seventy-nine genes were chosen for analysis of selective pressure, and five genes were found to be under positive selection. Another recent study indicates some putative mis-annotated and unannotated genes from chloroplast genome sequence of *J. regia* (Chakraborty 2016).

Recently, the phylogenetic position of *J. cinerea*, the unique species of *Trachycaryon* section, was assessed based on chloroplast genomes and nuclear DNA sequences (Dong et al. 2017). Results showed that *J. cinerea* should be included in *Cardiocaryon* section due to its relatedness with *J. mandshurica*, as was also observed with AFLP markers by Fjellstrom and Parfitt (1994b).

The nuclear genome of *J. regia* was sequenced recently and represents the first reference genome sequence for a nut crop. The team of Martínez-García et al. (2016) obtained it from "Chandler" using Illumina sequencing of paired end reads (500 million reads) from short and large fragment libraries. The 667 Mbp were assembled using two methods (SOAPdenovo2 and MaSuRCA) resulting in an N50 scaffold size of 464,955 bp. A total of 221,640 contigs were obtained with 37% GC content. Annotation work led to 32,498 gene models. Among the most abundant molecular functions of annotated genes, were metabolism (25.4%), cellular (22.1%), and membrane components (18.6%) and a set of 10,130 genes had SNPs within their annotated coding sequences. This genome sequence has facilitated construction of a 600K SNP Affymetrix array which is about to become available (David Neale, personal communication, March 2017) and could provide a great impetus for walnut breeding. All of this has been made possible by the establishment of the Walnut Genomics Implementation Group (WGIG) (http://ucanr.edu/sites/wgig/) at Davis, California.

Progress on transcriptomics

Transcriptomics is the study of the complete set of RNAs encoded by the genome at specific times and under specific conditions. ESTs generate millions of short single-pass nucleotide sequence reads for transcriptome exploration (Nagaraj

et al. 2006). Availability of this information in walnut has increased the development of gene-based markers like EST-SSRs which are derived from cDNA libraries. These are more evolutionarily conserved and therefore have higher transferability to related species than non-coding sequences. The first construction of walnut ESTs was accomplished in *J. regia* (Table 5). Five thousand twenty-five ESTs from seed coat tissues covering 16.41 Mb (Muir et al. 2004) have been submitted to NCBI database (https://www.ncbi.nlm.nih.gov). Three years later, an additional 13,559 ESTs were developed (2715 from leaves and 10,844 from roots) from *J. hindsii* × *J. regia* plants infected by *Pratylenchus vulnus* nematodes (Britton et al. 2007).

Next, a *J. nigra* EST library was constructed in order to understand better the determinants of wood quality. Eighty ESTs assumed to be linked to programmed cell death (PCD) were found by analyzing the region of heartwood formation located in the transition zone (TZ) (Huang et al. 2009). A transcript encoding homeobox protein knotted-1-like 3 (KNAT3) was found to be expressed more strongly in the TZ than in other tissues and this transcript, called *JnKNAT3*-like, is 67% similar to *Knox1/Knox2* domains from *A. thaliana* KNAT3.

Three years later, 188 transcript-derived fragments (TDFs) from cDNA-AFLP in *J. regia* became available (Ben Bâaziz et al. 2012). These were obtained as part of a study of light-induced leaf hydraulic conductance (K_{leaf}), which plays a role in plant-regulation of water. A large number of these are linked to protein products involved in global metabolism and cellular regulation. More recently, a transcriptome study of *J. hopeiensis* using Illumina technology identified 39,708 genes (Hu et al. 2015). The same year, 40 brand new ESTs became available in *J. regia* (Zhao et al. 2015).

In 2016, the genetic diversity of Chinese germplasm was assessed using the sequence polymorphisms within the phenylalanine ammonia-lyase (*PAL*) gene (Han et al. 2016). This gene was previously characterized and designated as *JrPAL* (Xu et al. 2012a). Expression of the transcription factor gene *JrCBF* was also studied (Xu et al. 2014). This could have a key role in cold resistance mechanisms because increased expression is observed under cold stress conditions.

Finally, three studies were conducted using Illumina RNA-sequencing methods. The first focused on the transcriptome of the heartwood/sapwood transition zone in *J. nigra* (Chakraborty et al. 2015). A highly abundant transcript in that zone had an amino acid composition linked to a putative extensin, a cell wall protein. The second examined plant-microbe interactions and a large family of NBS-LRR resistance genes in *J. regia* tissues (Chakraborty et al. 2016). The last of these characterized the transcriptome of buds, leaves, female flowers, and male flowers in *J. regia* (Dang et al. 2016).

Metabolomics and proteomics studies

Metabolomics work

Metabolomics is the large-scale study of metabolites, at specific times and under specific conditions, within cells, tissues, or organisms, and provides knowledge about biochemical activity and mechanisms of responses at the molecular level of cells. Over the last few years, some studies have focused on the beneficial effects of walnut consumption on health, directly related to primary metabolites. Unlike several other nuts which contain high level of monounsaturated fatty acids, *J. regia* kernels are rich in polyunsaturated fatty acids (PUFA) such as linoleic (n-3) and linolenic (n-6) acids (Maguire et al. 2004). Several human clinical studies of coronary heart disease have demonstrated that walnut consumption can lead to lower blood LDL cholesterol concentrations (Feldman and Elaine 2002; Banel and Hu 2009). Moreover, a diet with increased PUFA from walnuts is also linked to long-term beneficial effects on type II diabetes (Tapsell et al. 2009). Dietary walnut fingerprinting using HPLC-q-ToF-MS analysis characterized some urinary biomarkers, such as fatty acids metabolites, (Garcia-Aloy et al. 2014) and positive effects of walnut consumption on inflammatory processes, vascular reactivity, glycemic control (Ros 2010), and oxidative stress (Zhao et al. 2014) were described. However, walnuts also can be responsible for food-induced allergies, an emergent public health problem, and techniques to detect and characterize *J. regia* and *J. nigra* allergens have been developed (Costa et al. 2014). Since August 2017, 115 bioactivity screening studies of *J. mandshurica* and *J. regia* have become available on NCBI database (https://www.ncbi.nlm.nih.gov).

Metabolite characterization can also be used for *J. regia* breeding. A study of dynamic metabolite changes during kernel maturation, using gas chromatography-mass spectrometry (GC-MS), investigated changes during kernel ripening (Rao et al. 2016). A total of 252 metabolites were detected and among these, 36.47% were organic acid metabolites, 20% were involved in carbohydrate metabolism, 17.64% in amino acid metabolism, 11.76% in amine metabolism, 7.05% in phosphate metabolism, and 5.88% in lipid metabolism. Higher levels of asparagine, proline, and leucine were found in the early stages of kernel development, involved respectively in nitrogen transport for protein synthesis, prevention of stress-induced damage, and synthesis of derived compounds such as sugars and organic acids. This work showed that a large part of carbohydrate and protein-derived carbon was transferred into other compounds like fatty acids during kernel ripening.

Biosynthesis of nonstructural polyphenols, which could play a role in pathogen resistance, also was studied in *J. regia* (Colaric et al. 2005; Solar et al. 2006; Farooqui et al. 2015). The shikimic acid and aromatic amino acid

biosynthesis pathways generate precursors of numerous secondary metabolites, such as phenolic compounds. In *J. regia*, a major group of polyphenols is the hydrolysable tannins. Polyphenol oxidase (PPO) is responsible for catalyzing oxidation of phenolic compounds including tannins into highly reactive quinones. Originally only a single PPO gene, *JrPPO1* was observed in walnut (Escobar 2013; Araji et al. 2014). Using the genome sequence information, a second gene *JrPPO2* has now been reported and is preferentially expressed in callus tissue (Martínez-García et al. 2016). PPO may be involved in pellicle darkening and silencing the responsible gene could lead to development of walnuts with lighter pellicle color (Butterfield et al. 2015), an important commercial trait.

Proteomics research

Proteomics is the high-throughput analysis of the structure, function, and interaction of proteins produced by the genes of a particular cell, tissue, or organism. In connection with the discovery of the *JrPPO* genes, a tyrosinase was isolated from *J. regia* leaves and identified as a PPO corresponding to the known *JrPPO1* sequence (Zekiri et al. 2014). Tyrosinases can catalyze both ortho-hydroxylation of monophenols to o-diphenols (monophenolase activity) and the subsequent oxidation of o-diphenols to the corresponding o-quinones (diphenolase activity). This *JrTYR* protein shows high monophenolase activity while PPOs of numerous plants lack it. In seeking to elucidate this monophenolase/diphenolase specificity, one hypothesis for the lack of monophenolase activity was structural restriction of active site. However, after the first crystal structure of the *JrTYR* was developed, it appeared that the distinction between the two activities does not depend on the degree of restriction of the active site but, instead, that amino acid residues located at the entrance and in the second shell of the active site play a major role (Bijelic et al. 2015).

Protein markers were developed to study the genetic structure of Pakistan *J. regia* germplasm (Khan et al. 2010). Based on total seed storage proteins analyses, a dendrogram was constructed and accessions were grouped into three clusters. Two accessions were quite different from others and could be useful for future breeding programs.

More recently, a study proposed to fractionate and more widely characterize walnut proteins in order to better understand the alleged health inducing contributions of walnut protein additives for use in food (Mao et al. 2014). Four extractable protein fractions were obtained according to their solubility (glutenin 72.06%, albumin 7.54%, globulin 15.67%, and prolamin 4.73%). These could constitute a good source of essential amino acids for adults. SDS-PAGE analysis showed that molecular weights ranged from 14.4 to 66.2 kDa and two-dimensional gel electrophoresis (2DE) profiles showed that isoelectric points were between 4.8 and 6.8.

Walnut allergen stability during roasting in *J. regia* was investigated by liquid chromatography-mass spectrometry (Downs et al. 2016). Results showed an increase in mature 7S and 11S globulins, known allergens, after roasting for 20 min at 180 °C, probably due to an increase in protein digestibility. Since August 2017, 56,050 protein sequences from *J. regia* are now available on NCBI database (https://www.ncbi.nlm.nih.gov) and will be a useful resource for additional proteomic studies.

Genetic engineering

Since identification of the Ti plasmid in *Agrobacterium tumefaciens* (Schell 1975) and development of the first genetically modified crop plant, an antibiotic-resistant tobacco plant (Fraley et al. 1983), the public acceptance of genetically engineered plants has become a source of debate. Genetic engineering has been used successfully in *J. regia* since methods for somatic embryogenesis from immature cotyledons and plant regeneration were first developed (Tulecke and McGranahan 1985; McGranahan et al. 1988, 1990). Walnut trees expressing a modified *cry1A(c)* gene from *Bacillus thuringiensis* for codling moth (*Cydia pomonella*) resistance were developed in field trials (Leslie et al. 2001). In addition, extensively used "Paradox" rootstock hybrids were modified with *rolABC* genes from *Agrobacterium rhizogenes*, genes of interest in increasing rooting potential (Vahdati et al. 2002).

RNA interference (RNAi) gene silencing also has been used in *J. regia*. The *iaaM* and *ipt* genes of *A. tumefaciens* responsible for crown gall disease were silenced by expressing homologous inducers of post-transcriptional gene silencing (Escobar et al. 2002). Co-transformation with *A. tumefaciens* (*iaaM* and *ipt* genes) and *A. rhizogenes* (*Pv010* gene) was successfully performed in an attempt to combine resistance to both crown gall and *Pratylenchus vulnus* nematode into the same rootstock (gene stacking) (Walawage et al. 2013). *A. tumefaciens* was also used to transfer the *fld* gene from cyanobacteria to *J. regia* in order to increase osmotic stress tolerance (Sheikh Beig Goharrizi et al. 2016).

Improvements in markers for somatic embryo transformed selection have continued in *J. regia*. β-glucuronidase (GUS) was used for most of the early work and has been the most widely employed scorable marker for walnut but a green fluorescent protein (GFP) marker has also been used (Escobar et al. 2000). Recently, a new red fluorescent protein from *Discosoma* sp. (DsRED) was shown to be more stable and reliable (Zhang et al. 2015). DsRED was used to examine gene translocation from transformed rootstock to wild-type scion and this gene was not translocated unlike mRNA was,

suggesting gene signal transport from rootstock to scion in *J. regia* (Liu et al. 2017).

Breeding

In the past, the best walnut genotypes were selected from seedling populations and propagated both for consumption and agricultural traits (McGranahan and Leslie 2012) but walnut breeding programs started in the twentieth century (Fig. 4). The past 50 years have seen the emergence of many workgroups worldwide focused on walnut breeding, genetics, and biotechnology, as well as propagation, nutrition, and production techniques. For access to exhaustive information, see Proceedings of Walnut Symposia: Acta Horticulturae numbers I, 284 (1990, Hungary), II, 311 (1993, Spain), III, 442 (1997, Portugal), IV, 544 (2001, France), V, 705 (2005, Italy), VI, 861 (2010, Australia), and VII, 1050 (2014, China). In general, common goals for walnut breeding are a high yield with ease of shell cracking, kernels of high visual and taste quality, well-adapted flowering and harvest dates, and resistance to major diseases like walnut blight and anthracnose (Cosmulescu and Botu 2012). However, specific characteristics will depend on the area of production and its climatic conditions. The following is a short summary of the main breeding programs that have been, or especially currently are led, by teams in California, France, China, and the Middle East, the most important walnut producing countries.

California breeding program

In the USA, germplasm collections are maintained by the Walnut Improvement Program at the University of California in Davis, CA, and the USDA National Clonal Germplasm Repository in Winters, CA. The UC Walnut Improvement Program objectives include breeding to improve scion cultivars and rootstocks for the California walnut industry, development of better knowledge and methods in order to make breeding more efficient, acquisition, development, and maintenance of diverse germplasm with traits useful for future breeding, and application of new molecular and genomic methods for walnut improvement. Detailed information about the Walnut Improvement Program is available from the Walnut Research Reports Database at the University of California, Fruit and Nut Research Information Center website (http://ucanr.edu/sites/cawalnut/). This breeding program originated from, and remains fundamentally based on, a set of controlled crosses initiated and evaluated by Eugene F. Serr and Harold I. Forde from 1948 to 1979 (Tulecke and McGranahan 1994). Their major goals were to increase yield and kernel quality, to select appropriate leafing and fruiting date, and to increase resistance to diseases. To do

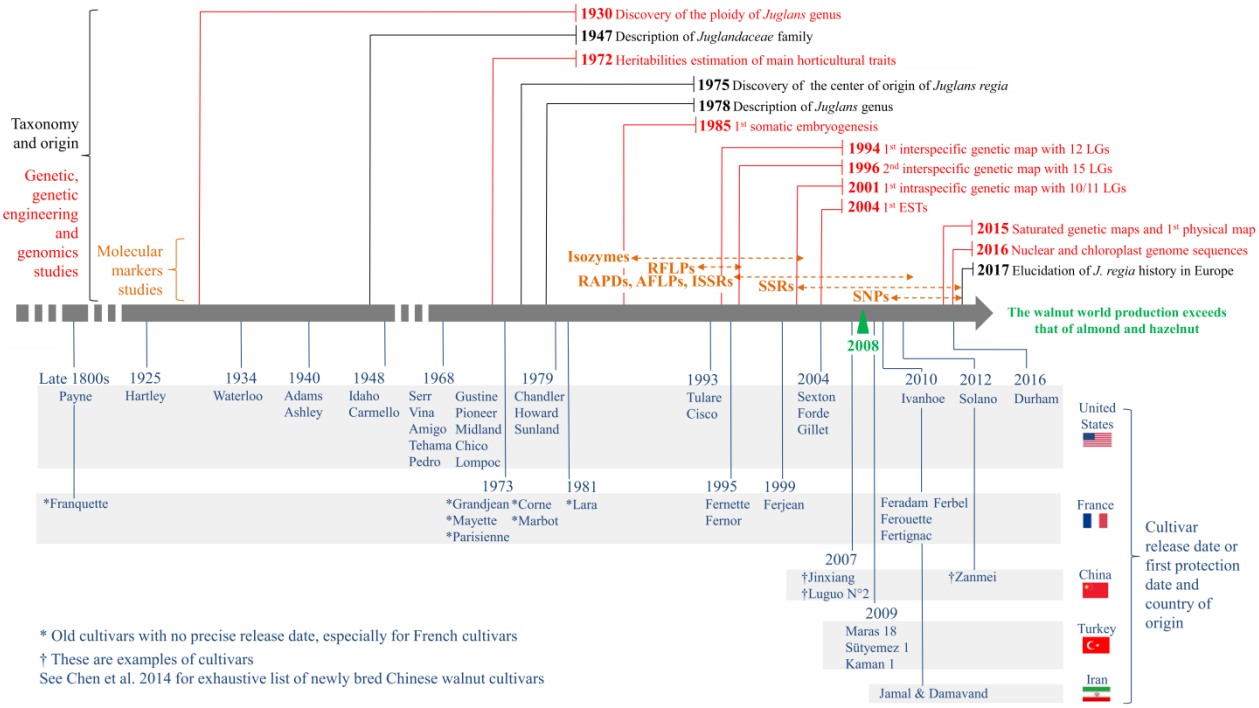


Fig. 4 Timeline of walnut breeding

this, they combined French cultivars for quality and late leafing with California cultivars, such as "Payne," chosen for lateral bearing and precocity. Late leafing helps avoid walnut blight (*X. arboricola* pv. *juglandis*) spread by spring rains. "Payne" is in the background of all current University of California cultivars. After evaluation of about 6000 progeny, ten cultivars were released in 1968. In 1979, an additional three cultivars, "Chandler," "Howard," and "Sunland," were released. The second phase of the breeding program was led by Gale H. McGranahan from 1982 to 2009. Initially directed at blackline resistance, the program objectives were soon expanded to include blight resistance, earlier harvest dates, yield and quality improvement, and rootstock improvement. "Tulare" was released in 1993, and current Californian walnut production is largely derived from "Chandler" (74.2%), "Tulare" (12.6%), and "Howard" (8.1%). Three additional cultivars, "Sexton," "Gillet," and "Forde," were released in 2004. The third phase of the breeding program has been led by Charles A. Leslie since 2009 with the continued goal of obtaining cultivars with early harvest dates, good yield, light kernel color, and resistance to walnut bacterial blight and cherry leafroll virus. Three new cultivars have been released recently: "Ivanhoe" (very early harvesting) in 2010, "Solano" in 2012, and "Durham" in 2016. All three of these cultivars have very light colored kernels and harvest earlier than "Chandler."

The program continues with projects directed at diverse aspects of walnut improvement (Leslie et al. 2015), including (1) Scion cultivar breeding is now emphasizing disease and abiotic stress resistance with 94 repropagated selections under advanced evaluation and more than 3900 seedlings under early stage observation. (2) Backcross breeding is directed at introgression of CLRV resistance into commercial cultivar. A DNA marker linked to CLRV resistance with an error rate of 10% is employed for preliminary screening and preselections are confirmed by direct virus patch testing. Individuals from 4th and 5th generation crosses population are under evaluation. (3) Rootstock breeding for resistance to soil-borne pathogens has emphasized controlled crosses of *J. regia* "Serr" pollen onto selected *J. cathayensis* and *J. microcarpa* female parents and testing for resistance to disease (Kluepfel et al. 2015). (4) An expanded emphasis on applied genomics has included generation and characterization of over 400 "Chandler" × "Idaho" seedlings for marker development, genetic mapping, and development of reference genomes. (5) New technology development includes evaluation of transgenic approaches to disease resistance and continuing improvement of micropropagation methods. (6) Collections of both field and in vitro germplasm resources are maintained for current and future use. Rootstock improvement is concentrated on developing and releasing disease-resistant "Paradox" (hybrids between *J. nigra* and *J. regia*) and improving the clonal propagation procedures needed for commercial production. In recent years, two new rootstocks have been released: RX1 (*J. microcarpa* × *J. regia*) with resistance to *Phytophthora* crown and root rots, and VX211 (*J. hindsii* × *J. regia*) which is very vigorous and shows tolerance to nematodes (McGranahan and Leslie 2012).

French breeding program

In France, germplasm collections are maintained by the Institut National de la Recherche Agronomique (INRA) (in English: National Institute of Agricultural Research) of Bordeaux, more precisely by the Centre for Genetic Resources *Prunus* which also manages *Castanea* and *Juglans* species. The first breeding program was led by E. Germain from INRA, in Unité de Recherche sur les Espèces Fruitières et la Vigne in Bordeaux (in English: Fruit and Vine Research Station) from 1977 to 1995. Twenty-eight crosses were made mainly with French cultivars as female parents ("Franquette," "Grosjean," "Marbot") because they have a late bud break and good fruit quality. Male parents were chosen for lateral bearing including California cultivars ("Pedro," "Chandler") (Germain 1997). Two cultivars were released in 1995: "Fernette" and "Fernor," both lateral bearing cultivars, from a cross between "Franquette" and "Lara." A third new cultivar, "Ferjean," a lateral bearing cultivar with quite small kernels, became available in 1999 from a cross between "Grosvert" and "Lara." In 2010, the last four cultivars from this program were released: "Feradam," "Ferbel," "Ferouette," and "Fertignac." The second French breeding program was led by E. Germain and F. Delort using a large and diverse collection of genetic material developed from 1996 to 2007. The best hybrids from the first program were crossed with genotypes from the Mediterranean area, Iran, Japan, and Central Asia to prevent problems with inbreeding, observed previously among French and California cultivar backcrosses. Although some promising hybrids were characterized and evaluated, the breeding program ended in 2007 due to a new orientation for INRA, following a decision to focus on a smaller number of model species, such as *Prunus* and *Malus*.

About 70% of French walnut production is provided by "Franquette," an ancestral cultivar with late leafing and terminal bearing habit, discovered in the early 1800s. There are two Appelations d'Origine Protégée (AOP) (in English: Protected Designations of Origin) for the two main production areas: AOP Noix du Périgord and AOP Noix de Grenoble. French walnut orchard areas increased by nearly 19% between 2000 and 2010, reaching approximately 20,000 ha. This development makes it the most important French fruit crop other than apples. France exports 80% of its production in-shell, mainly to Europe, thanks in particular to the quality of its product. The main goals of walnut breeding in France are similar to that of California for some traits. The French walnut industry needs new cultivars with lateral bearing, increased yield (~

4 t/ha versus ~ 2 t/ha for "Franquette"), and larger nut size. New cultivars also need high nut and kernel quality, including light color, thicker shell, and ease of cracking. They must also be well adapted to French climatic conditions, specifically winter hardiness so with late bud break to avoid late spring frosts but also with rapid fruit development. A team of INRA Clermont-Ferrand has been working on winter hardiness and found that the high heritability of autumn and winter hardinesses, and the lack of correlation with growth traits, may provide interesting selection pathways (Guàrdia et al. 2016). However, the development of dedicated models highlighted that early frost damages, although lower than at other times of the leafless period, would negatively influence fruit yield in the following year (Charrier et al. 2017). Last decade, a mechanistic modeling approach has been developed to forecast frost risks under higher climate variability and subsequent stress exposure (Charrier et al. 2015). This approach integrates the relevant physiological processes: dormancy, carbohydrate metabolism, and water status (Charrier and Améglio 2011; Charrier et al. 2013).

A new research project began mid-2017 in cooperation between the INRA Bordeaux-Aquitaine Centre and the Centre Technique Interprofessionnel des Fruits et Légumes (Ctifl) (in English: Fruit and Vegetable Interprofessional Technical Centre), supported financially by the Nouvelle-Aquitaine region. The main objectives are evaluation of the genetic diversity of germplasm resources (about 270 worldwide accessions), identification of the genetic determinants of traits of interest, and establishment of the necessary tools for MAS achievement. In the long term, the choice of genitors for the new breeding program led by the Ctifl will be possible. This project is carried out in several steps including (1) the collection of available data from E. Germain on the germplasm in the form of observation archives about phenology, architecture of the tree, and fruit quality. (2) The phenotyping of the germplasm regarding the three main pests and diseases of *J. regia* in France: walnut blight (*X. arboricola* pv. *juglandis*) and anthracnoses (*Gnomonia leptostyla* and *Colletotrichum acutatum*). (3) The genotyping of the germplasm firstly with SSR markers to eliminate genetically close accessions and secondly with the 600 K SNP Affymetrix array developed by the University of California in Davis to perform genome-wide genotyping on interesting accessions. (4) The completion of a genome-wide association study on traits of interest and the establishment of a germplasm collection.

Other breeding programs

As can be observed in the Proceedings of the Seventh International Walnut Symposium (Tian 2014), a number of other countries also have active walnut breeding programs, notably China and in the Middle East. Chinese breeding goals include early fruiting, high yield with high quality, and tolerance to anthracnose (*C. gloeosporioides*) and other diseases. Twenty-one Chinese provinces or autonomous regions grow walnuts. The major ones are the Xinjiang autonomous region, Yunnan, Shaanxi, Shanxi, Sichuan, Shandong, and Hebei provinces. There is a long cultivation history of walnut in China, and there are many germplasm repositories. Indeed, Chinese botanists have identified and collected more than 800 cultivars, including for example the 216 cultivars, 486 types, and 164 seedling landraces noted by Chen et al. (2014). In 1985, the National Chinese Walnut Germplasm Repository was established at the Shandong Pomology Institute, which currently maintains 415 accessions collected from 24 provinces and from the USA. Each province has developed many local cultivars, but the main ones of national interest include, for instance, "Bofeng," "Bokexiang," "Zha343," "Luguang," "Xiangling," "Hanfeng," "Wen185," the "Jinboxiang" series, the "Zhongling" series, the "Jinlong" series, and the "Luguo" series. Between 2006 and 2012, more than 20 bred cultivars were released, including "Yangza No.1" (tolerant to walnut blight) and "Yangza No.2 hao" (for warm and humid areas) from the Forestry Bureau of Dali in Yunnan in 2006, "Jinxiang" from the Shanxi Academy of Forestry Sciences in 2007 (high yield), "Zanmei" (early bearing), and "Zhenzhuxiang" (small fruits) from the Agricultural University of Hebei in 2012 (Chen et al. 2014). In addition, 17 newly bred cultivars were released by the Shandong Pomology Institute between 2003 and 2012, including "Luguo 2" in 2012 from an intraspecific hybridization and "Luwen 1" in 2009 from an interspecific hybridization (Zhang et al. 2014). Rootstock improvement work also has proceeded in China, including the "Jin RS-1" series released in 2011 by the Shanxi Academy of Forestry Sciences. "Jin RS-2" and "Jin RS-3" have been evaluated for cold, disease, and pest resistance, and seem to be ideal for northern Chinese areas subject to frost (Wang et al. 2014a). Finally, a program in the Xinjiang autonomous region is also assessing the merits of the second bearing habit found in many early ripening dwarf walnuts (Wang et al. 2014b).

In Turkey, walnut germplasm consists mainly of early leafing and terminal bearing cultivars. The production is made mostly in the Mid-eastern Anatolia and Western Black Sea regions. The first Turkish selection study dating to 1971 examined diverse areas of Turkey and 20 seedlings genotypes were selected, including "Yalova 1," "Yalova 3," "Bilecik," and "Şebin" which has become a reference in many production areas of Turkey (Akça 2014). Late leafing and lateral bearing were not yet considered in that work but in 1990, a new selection study was developed in order to select genotypes that would not be damaged by late spring frosts. Selection was conducted for new types combining lateral bearing, late leafing, late flowering, and tolerance to diseases (Akça and Ozongun 2004). At that time, all Turkish cultivars had been selected from native populations (Aslantaş 2006).

The National Walnut Cultivar Breeding Program, initiated in 1998, introduced new cultivars with late leafing and lateral bearing. Then, in 2000, a new breeding program was started to develop cultivars with improved nut quality traits including greater kernel weight and shell thickness (Akça and Polat 2007). In 2008, a variety breeding program directed at developing cultivars with higher yield, late leafing, improved nut quality, and resistance to blight made 13 crosses using Turkish cultivars ("Şebin," "Akça1," etc.) and foreign cultivars ("Franquette" and "Hartley"). From these, 1340 hybrids were obtained and are under evaluation (Akça et al. 2016). From the previous mass selection program, three cultivars were released in 2009: "Maras 18," "Sütyemez 1," and "Kaman 1" (Sütyemez 2016).

In Iran, most walnut trees also originated from seed (Ghasemi et al. 2012). A walnut improvement program started there in 1983 with selection of superior genotypes from native germplasm in Karaj, Shahrood, Mashhad, and Urumieh regions and two major goals: to create new cultivars and to develop efficient vegetative propagation methods (Hassani et al. 2014). Seven superior genotypes were used as parents, and progenies were evaluated along with eight foreign cultivars, including "Serr," "Hartley," etc. As a result, two cultivars were released in 2010: "Jamal" (protandrous) and "Damavand" (protogynous). In addition to these two cultivars, five foreign cultivars ("Chandler," "Pedro," "Hartley" for high yield, "Franquette," and "Ronde de Montignac" for pollination) are recommended for Iranian climatic conditions. Currently, other genotypes are under evaluation for release of new cultivars and development of micropropagation is proceeding (Hassani et al. 2014). To finish, in vitro propagation of walnut plays a very important role in multiplication of cultivars and Iranian teams have been particularly involved in the development of these techniques (Payghamzadeh and Kazemitabar 2011).

Conclusion and future prospects

There is already a long and productive history of walnut genetic improvement since the discovery of heritabilities of main agronomic traits in 1972. Molecular markers permitted the construction of the first genetic maps between 1994 and 2001 and genomics allowed to crack the walnut genome in 2016. The new cultivars, including "Chandler" released in 1979, and the new cropping practices have led to a continuous increase in walnut world production that exceeds since 2008 that of almond and hazelnut. The technological advances permitted to solve a number of questions about taxonomic classification and genetic diversity of worldwide *Juglans* species germplasm.

It is undeniable that "Chandler" genome sequencing now will facilitate and accelerate great advances around the world in walnut genomics, breeding, and identification of genetic determinants of traits of interest. In California, initial genomic selection model building is already planned for 2017 (Neale et al. 2015). The VIII International Walnut Symposium planned for Santiago, Chile in November 2017, will be an opportunity to discover the latest results of teams working on the field. In the near future, we are likely to see the application of CRISPR/Cas9-mediated techniques for walnut improvement, as has already observed in tomato (Pan et al. 2016) and other crops. In France, where the walnut breeding program stopped in 2007, there is now a willingness to initiate a new breeding program that will likely employ MAS, according to Ctifl.

In the context of climate change and global awareness of the ecological impact of chemical inputs, efforts will have to be made in walnut scion and rootstock breeding to release new materials adapted to local environmental conditions and tolerant to current and future pests and diseases, in order to maintain the production and the quality. In addition to genetic improvement, crop management must be continually improved and made more efficient, especially regarding the concepts of water, nutrition, soil, physiology, and defense. In ending words, in light of growing world production and consumption of the crop, rapidly increasing knowledge and use of genetics, new investments in various breeding programs, and widely expanding international research cooperation, there are good times ahead for walnuts.

Acknowledgments We gratefully thank Joseph A. Grant (Cooperative Extension Farm Advisor, San Joaquin County, California), Patrick J. Brown, and Charles A. Leslie (UC Davis, Department of Plant Sciences, California) for helping us to draft the "California breeding program" part. We acknowledge Philippe Jardel (Walnut producer in Martel, France), Eloïse Tranchand (Walnut Experimental Station of Creysse, France), and Guillaume Charrier (INRA Clermont-Ferrand, France) for providing us information about perspectives in French walnut. We thank again Charles A. Leslie and William Nicolas (CNRS Bordeaux, France) for English language revision. Finally, this work could be achieved thanks to a fixed-term contract funded by the INRA and the Ctifl.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

Abedi B, Parvaneh T (2016) Study of correlations between horticultural traits and variables affecting kernel percentage of walnut (*Juglans regia* L.) J Nuts 7:35–44

Ahandani EA, Ramandi HD, Sarmad J et al (2014) Evaluation of morphological diversity among some Persian walnut accessions

(*Juglans regia* L.) in Guilan, northern Iran. Int J Plant Biol Res 2: 1015–1022

Akça Y (2014) 'Şebin' Persian walnut. Acta Hortic 1050:139–142

Akça Y, Ozongun S (2004) Selection of late leafing, late flowering, laterally fruitful walnut (*Juglans regia*) types in Turkey. N Z J Crop Hortic Sci 32(4):337–342. https://doi.org/10.1080/01140671.2004.9514313

Akça Y, Polat AA (2007) Present status and future of walnut production in Turkey. Eur J Plant Sci Biotechnol 1:57–64

Akça Y, Sen SM (1997) The relationship between dichogamy and yield-nut characteristics in *Juglans regia* L. Acta Hortic 442:215–216

Akça Y, Sütyemez M, Yilmaz S, Karadag H (2016) The new walnut variety breeding program in Turkey. VIIth International Scientific Agricultural Symposium. Jahorina, Bosnia and Herzegovina, pp 461–466

Aletà N, Olarte C, Truco MJ, Arus P (1990) Identification of walnut cultivars by isozyme analysis. Acta Hortic 284:91–96

Ali AM, Zubair SJ, Abbas AM, Jubrael JMS (2016) Genetic diversity among walnuts (*Juglans regia*) population in Kurdistan region–Iraq using AFLP-PCR. ZANCO J Pure Appl Sci 28:50–55

Aly MA, Fjellstrom RG, McGranahan GH, Parfitt DE (1992) Origin of walnut somatic embryos determined by RFLP and isozyme analysis. Hortscience 27:61–63

Amiri R, Vahdati K, Mohsenipoor S et al (2010) Correlations between some horticultural traits in walnut. Hortscience 45:1690–1694

Andersen JR, Lübberstedt T (2003) Functional markers in plants. Trends Plant Sci 8(11):554–560. https://doi.org/10.1016/j.tplants.2003.09.010

APG III (2009) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants. Bot J Linn Soc 161:105–121

Aradhya MK, Potter D, Simon CJ (2006) Cladistic biogeography of *Juglans* (*Juglandaceae*) based on chloroplast DNA intergenic spacer sequences. In: Motley TJ, Zerega N, Cross H (eds) Darwin's harvest: new approaches to the origins, evolution, and conservation of crops. Columbia University Press, New York, pp 143–170. https://doi.org/10.7312/motl13316-008

Aradhya MK, Potter D, Gao F, Simon CJ (2007) Molecular phylogeny of *Juglans* (*Juglandaceae*) : a biogeographic perspective. Tree Genet Genomes 3(4):363–378. https://doi.org/10.1007/s11295-006-0078-5

Aradhya M, Woeste K, Velasco D (2009) Genetic diversity, structure and differentiation in cultivated walnut (*Juglans regia* L.) Acta Hortic 861:127–132

Araji S, Grammer TA, Gertzen R, Anderson SD, Mikulic-Petkovsek M, Veberic R, Phu ML, Solar A, Leslie CA, Dandekar AM, Escobar MA (2014) Novel roles for the polyphenol oxidase enzyme in secondary metabolism and the regulation of cell death in walnut. Plant Physiol 164(3):1191–1203. https://doi.org/10.1104/pp.113.228593

Arulsekar S, McGranahan GH, Parfitt DE (1986) Inheritance of phosphoglucomutase and esterase isozymes in Persian walnut. J Hered 77(3):220–221. https://doi.org/10.1093/oxfordjournals.jhered.a110225

Arzani K, Mansouri-Ardakan H, Vezvaei A, Roozban MR (2008) Morphological variation among Persian walnut (*Juglans regia*) genotypes from central Iran. N Z J Crop Hortic Sci 36(3):159–168. https://doi.org/10.1080/01140670809510232

Asadian G, Pieber K (2005) Morphological variations in walnut varieties of the Mediterranean regions. Int J Agric Biol 7:71–73

Aslantaş R (2006) Identification of superior walnut (*Juglans regia*) genotypes in north-eastern Anatolia, Turkey. N Z J Crop Hortic Sci 34(3): 231–237. https://doi.org/10.1080/01140671.2006.9514412

Asma BM (2012) Pomological and phenological characterization of promising walnut (*Juglans regia* L.) genotypes from Malatya, Turkey. Acta Sci Pol Horturum Cultus 11:169–178

Atefi J (1990) Preliminary research of Persian walnut and correlation between pair characters. Acta Hortic 284:97–104

Bai W-N, Liao W-J, Zhang D-Y (2010) Nuclear and chloroplast DNA phylogeography reveal two refuge areas with asymmetrical gene flow in a temperate walnut tree from East Asia. New Phytol 188(3):892–901. https://doi.org/10.1111/j.1469-8137.2010.03407.x

Banel DK, Hu FB (2009) Effects of walnut consumption on blood lipids and other cardiovascular risk factors: a meta-analysis and systematic review. Am J Clin Nutr 90(1):56–63. https://doi.org/10.3945/ajcn.2009.27457

Bayazit S (2012) Determination of relationships among kernel percentage and yield characteristics in some Turkish walnut genotypes by correlation and path analysis. J Anim Plant Sci 22:513–517

Bayazit S, Kazan K, Gülbitti S, Çevik V, Ayanoğlu H, Ergül A (2007) AFLP analysis of genetic diversity in low chill requiring walnut (*Juglans regia* L.) genotypes from Hatay, Turkey. Sci Hortic 111(4):394–398. https://doi.org/10.1016/j.scienta.2006.11.006

Ben Bâaziz K, Lopez D, Bouzid S, Cochard H, Venisse JS, Sakr S (2012) Early gene expression in the walnut tree occurring during stimulation of leaf hydraulic conductance by irradiance. Biol Plant 56(4): 657–666. https://doi.org/10.1007/s10535-012-0120-4

Bijelic A, Pretzler M, Molitor C, Zekiri F, Rompel A (2015) The structure of a plant tyrosinase from walnut leaves reveals the importance of "substrate-guiding residues" for enzymatic specificity. Angew Chem Int Ed Engl 54(49):14677–14680. https://doi.org/10.1002/anie.201506994

Bolling BW, McKay DL, Blumberg JB (2010) The phytochemical composition and antioxidant actions of tree nuts. Asia Pac J Clin Nutr 19(1):117–123

Botstein DR, White RL, Skolnick MH, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32(3):314–331

Britton M, Leslie C, McGranahan G, Dandekar A (2007) Analysis of genes expressed in nematode-infected walnut plants. Unpublished raw data

Busov VB, Rink G, Woeste K (2002) Allozyme variation and mating system of black walnut (*Juglans nigra* L.) in the central hardwood region of the United States. For Genet 9:315–322

Butterfield T, Phu M, Chakraborty S, et al (2015) Identifying the genetic determinant(s) of pellicle coloration in walnut. Walnut Res Rep

Carrion JS, Sanchez-Gomez P (1992) Palynological data in support of the survival of walnut (*Juglans regia* L.) in the western Mediterranean area during last glacial times. J Biogeogr 19(6):623–630. https://doi.org/10.2307/2845705

Ćelepirović N, Karija Vlahović M, Dounavi A, Ivanković M (2016) Optimizations of high throughput multiplex polymerase chain reaction with simple sequence repeat markers for genotyping of common walnut populations (Juglans regia L.). Silva Fenn 50 article id 1674

Chakraborty S (2016) YeATSAM analysis of the chloroplast genome of walnut reveals several putative un-annotated genes and mis-annotation of the trans-spliced rps12 gene in other organisms. Preprint bioRxiv 094441

Chakraborty S, Britton M, Wegrzyn J, Butterfield T, Martínez-García PJ, Reagan RL, Rao BJ, Leslie CA, Aradhaya M, Neale D, Woeste K, Dandekar AM (2015) YeATS-a tool suite for analyzing RNA-seq derived transcriptome identifies a highly transcribed putative extensin in heartwood/sapwood transition zone in black walnut [version 2; referees: 3 approved]. F1000Research 4:155, https://doi.org/10.12688/f1000research.6617.2

Chakraborty S, Britton M, Martínez-García PJ, Dandekar AM (2016) Deep RNA-Seq profile reveals biodiversity, plant–microbe interactions and a large family of NBS-LRR resistance genes in walnut (*Juglans regia*) tissues. AMB Express 6(1):12. https://doi.org/10.1186/s13568-016-0182-3

Charrier G, Améglio T (2011) The timing of leaf fall affects cold acclimation by interactions with air temperature through water and carbohydrate contents. Environ Exp Bot 72(3):351–357. https://doi.org/10.1016/j.envexpbot.2010.12.019

Charrier G, Bonhomme M, Lacointe A, Améglio T (2011) Are budburst dates, dormancy and cold acclimation in walnut trees (*Juglans regia* L.) under mainly genotypic or environmental control? Int J Biometeorol 55(6):763–774. https://doi.org/10.1007/s00484-011-0470-1

Charrier G, Poirier M, Bonhomme M, Lacointe A, Ameglio T (2013) Frost acclimation in different organs of walnut trees *Juglans regia* L.: how to link physiology and modelling? Tree Physiol 33(11): 1229–1241. https://doi.org/10.1093/treephys/tpt090

Charrier G, Ngao J, Saudreau M, Améglio T (2015) Effects of environmental factors and management practices on microclimate, winter physiology and frost resistance in trees. Front Plant Sci 6:259. https://doi.org/10.3389/fpls.2015.00259

Charrier G, Chuine I, Bonhomme M, Améglio T (2017) Assessing frost damages using dynamic models in walnut trees: exposure rather than vulnerability controls frost risks. Plant Cell Environ https://doi.org/10.1111/pce.12935

Chen L-H, Hu T-X, Zhang F, Li G-H (2008) Genetic diversities of four *Juglans* populations revealed by AFLP in Sichuan province, China. J Plant Ecol 32:1362–1372

Chen L-H, Hu T-X, Zhang F (2009) AFLP analysis on genetic diversity of *Juglans* populations in dry and dry-hot valleys of Sichuan province. J Fruit Sci 26:48–54

Chen C-M, Han S-J, Yuan S-S et al (2013) Isolation and characterization of 20 polymorphic microsatellite markers for *Juglans mandshurica* (*Juglandaceae*). Appl Plant Sci 1:1–4

Chen X, Xu L, Zhang SL, Liu ZQ (2014) Walnut Genebank in China National Clonal Plant Germplasm Repository. Acta Hortic 1050: 89–94

Cheng S, Yang W (1987) Taxonomic studies of ten species of the genus *Juglans* based on isozymic zymograms. Acta Hortic Sin 14:90–96

Choi S, Wing RA (2000) The construction of bacterial artificial chromosome (BAC) libraries. Plant Mol Biol Man H5:1–28

Christopoulos MV, Rouskas D, Tsantili E, Bebeli PJ (2010) Germplasm diversity and genetic relationships among walnut (*Juglans regia* L.) cultivars and Greek local selections revealed by inter-simple sequence repeat (ISSR) markers. Sci Hortic 125(4):584–592. https://doi.org/10.1016/j.scienta.2010.05.006

Ciarmiello LF, Piccirillo P, Pontecorvo G, de Luca A, Kafantaris I, Woodrow P (2011) A PCR based SNPs marker for specific characterization of English walnut (*Juglans regia* L.) cultivars. Mol Biol Rep 38(2):1237–1249. https://doi.org/10.1007/s11033-010-0223-y

Colaric M, Veberic R, Solar A, Hudina M, Stampar F (2005) Phenolic acids, syringaldehyde, and juglone in fruits of different cultivars of *Juglans regia* L. J Agric Food Chem 53(16):6390–6396. https://doi.org/10.1021/jf050721n

Cosmulescu S, Botu M (2012) Walnut biodiversity in south-western Romania-resource for perspective cultivars. Pak J Bot 44:307–311

Costa J, Carrapatoso I, Oliveira MBPP, Mafra I (2014) Walnut allergens: molecular characterization, detection and clinical relevance. Clin Exp Allergy 44(3):319–341. https://doi.org/10.1111/cea.12267

Dang M, Liu Z-X, Chen X, Zhang T, Zhou HJ, Hu YH, Zhao P (2015) Identification, development, and application of 12 polymorphic EST-SSR markers for an endemic Chinese walnut (*Juglans cathayensis* L.) using next-generation sequencing technology. Biochem Syst Ecol 60:74–80. https://doi.org/10.1016/j.bse.2015.04.004

Dang M, Zhang T, Hu Y, Zhou H, Woeste K, Zhao P (2016) *De Novo* assembly and characterization of bud, leaf and flowers Transcriptome from *Juglans Regia* L. for the identification and characterization of new EST-SSRs. Forests 7(10):247–263. https://doi.org/10.3390/f7100247

Dangl GS, Woeste K, Aradhya MK et al (2005) Characterization of 14 microsatellite markers for genetic analysis and cultivar identification of walnut. J Am Soc Hortic Sci 130:348–354

De Candolle A (1885) Origin of cultivated plants. University of Michigan. D. Appleton, New York, p 18. https://doi.org/10.5962/bhl.title.29067

Dode LA (1909) Contribution to the study of the genus *Juglans* (English translation by R. E. Cuendett). Bull Soc Dendrol Fr 12:165–215

Dong W, Xu C, Li W, Xie X, Lu Y, Liu Y, Jin X, Suo Z (2017) Phylogenetic resolution in Juglans based on complete chloroplast genomes and nuclear DNA sequences. Front Plant Sci 8:1148. https://doi.org/10.3389/fpls.2017.01148

Downs ML, Baumert JL, Taylor SL, Mills ENC (2016) Mass spectrometric analysis of allergens in roasted walnuts. J Proteome 142:62–69. https://doi.org/10.1016/j.jprot.2016.04.045

Ducci F, De Rogatis A, Proietti R (2010) *Juglans regia* L., phenotypic selection and assessment of genetic variation within a simulated seed orchard. Ann Silvic Res 36:139–150

Dvorak J, Aradhaya M, Leslie C, Luo M-C (2015) Discovery of the causative mutation of the lateral bearing phenotype in walnut. Walnut Res Rep

Ebrahimi A, Fatahi R, Zamani Z (2011) Analysis of genetic diversity among some Persian walnut genotypes (*Juglans regia* L.) using morphological traits and SSRs markers. Sci Hortic 130(1):146–151. https://doi.org/10.1016/j.scienta.2011.06.028

Ebrahimi A, Khadivi-Khub A, Nosrati Z, Karimi R (2015) Identification of superior walnut (*Juglans regia*) genotypes with late leafing and high kernel quality in Iran. Sci Hortic 193:195–201. https://doi.org/10.1016/j.scienta.2015.06.049

Ebrahimi A, Zarei A, McKenna JR, Bujdoso G, Woeste KE (2017) Genetic diversity of Persian walnut (*Juglans regia*) in the cold-temperate zone of the United States and Europe. Sci Hortic 220: 36–41. https://doi.org/10.1016/j.scienta.2017.03.030

Emilia MM, Spada M, Beritognolo I, Cannata F (1997) Differentiation of walnut hybrids (*Juglans Nigra* L. x *Juglans Regia* L.) through RAPD markers. Acta Hortic 442:43–52

Erturk U, Dalkilic Z (2011) Determination of genetic relationship among some walnut (*Juglans regia* L.) genotypes and their early-bearing progenies using RAPD markers. Romanian Biotechnol Lett 16: 5944–5952

Escobar MA (2013) RNA-seq and metabolite profiling reveal novel functions for the polyphenol oxidase enzyme in walnut (*Juglans regia*). Plant and Animal Genome XXI Conference, San Diego, CA, January 14

Escobar MA, Park J, Polito VS et al (2000) Using GFP as a scorable marker in walnut somatic embryo transformation. Ann Bot 85(6): 831–835. https://doi.org/10.1006/anbo.2000.1143

Escobar MA, Leslie CA, McGranahan GH, Dandekar AM (2002) Silencing crown gall disease in walnut (*Juglans regia* L.) Plant Sci 163(3):591–597. https://doi.org/10.1016/S0168-9452(02)00164-4

Eskandari S, Hassani D, Abdi A (2005) Investigation on genetic diversity of Persian walnut and evaluation of promising genotypes. Acta Hortic 705:159–166

Farooqui A, Khan A, Borghetto I, Kazmi SU, Rubino S, Paglietti B (2015) Synergistic antimicrobial activity of *Camellia sinensis* and *Juglans regia* against multidrug-resistant bacteria. PLoS One 10(2): e0118431. https://doi.org/10.1371/journal.pone.0118431

Fatahi R, Ebrahimi A, Zamani Z (2010) Characterization of some Iranians and foreign walnut genotypes using morphological traits and RAPDs markers. Hortic Environ Biotechnol 51:51–60

Feldman MD, Elaine B (2002) The scientific evidence for a beneficial health relationship between walnuts and coronary heart disease. J Nutr 132(5):1062S–1101S

Fikret Balta M, Dogan A, Kazankaya A et al (2007) Pomological definition of native walnuts (*Juglans regia* L.) grown in Central Bitlis. J Biol Sci 7:442–444

Fjellstrom RG, Parfitt DE (1994a) RFLP inheritance and linkage in walnut. Theor Appl Genet 89(6):665–670. https://doi.org/10.1007/BF00223702

Fjellstrom RG, Parfitt DE (1994b) Walnut (*Juglans* spp.) genetic diversity determined by restriction fragment length polymorphism. Genome 37(4):690–700. https://doi.org/10.1139/g94-097

Fjellstrom RG, Parfitt DE, McGranahan GH (1994) Genetic relationships and characterization of Persian walnut (*Juglans regia* L.) cultivars using restriction fragment length polymorphisms (RFLPs). J Am Soc Hortic Sci 119:833–839

Forde HI (1975) Walnuts. In: Janick J, Moore JN (eds) Advances in Fruit Breeding. Purdue University Press, West Lafayette, pp 439–455

Forde HI, McGranahan GH (1996) Walnuts. In: Janick J, Moore JN (eds) Fruit breeding, vol. 3, Nuts. Wiley, New York, pp 241–274

Fornari B, Malvolti ME, Taurchini D et al (2001) Isozyme and organellar DNA analysis of genetic diversity in natural/naturalised European and Asiatic walnut (*Juglans regia* L.) populations. Acta Hortic 544: 167–178

Foroni I, Rao R, Woeste K, Gallitelli M (2005) Characterisation of *Juglans regia* L. with SSR markers and evaluation of genetic relationships among cultivars and the 'Sorrento' landrace. J Hortic Sci Biotechnol 80(1):49–53. https://doi.org/10.1080/14620316.2005.11511890

Foroni I, Woeste K, Monti LM, Rao R (2007) Identification of "Sorrento" walnut using simple sequence repeats (SSRs). Genet Resour Crop Evol 54(5):1081–1094. https://doi.org/10.1007/s10722-006-9187-0

Fraley RT, Rogers SG, Horsch RB, Sanders PR, Flick JS, Adams SP, Bittner ML, Brand LA, Fink CL, Fry JS, Galluppi GR, Goldberg SB, Hoffmann NL, Woo SC (1983) Expression of bacterial genes in plant cells. Proc Natl Acad Sci U S A 80(15):4803–4807. https://doi.org/10.1073/pnas.80.15.4803

Gandev S, Dzhuvinov V (2015) Evaluation of some walnut cultivars under the climatic conditions of South Bulgaria. Agroznanje 15: 5–16

Garcia-Aloy M, Llorach R, Urpi-Sarda M, Tulipani S, Estruch R, Martínez-González MA, Corella D, Fitó M, Ros E, Salas-Salvadó J, Andres-Lacueva C (2014) Novel multimetabolite prediction of walnut consumption by a urinary biomarker model in a free-living population: the PREDIMED study. J Proteome Res 13(7):3476–3483. https://doi.org/10.1021/pr500425r

Germain E (1990) Inheritance of late leafing and lateral bud fruitfulness in walnut (*Juglans regia* L.). Phenotypic correlations among some traits of the trees. Acta Hortic 284:125–134

Germain E (1997) Genetic improvement of the Persian walnut (*Juglans regia* L.) Acta Hort 442:21–32

Germain E, Prunet J-P, Garcin A (1999) Le noyer, monographie, Ctifl

Ghasemi M, Arzani K, Hassani D (2012) Evaluation and identification of walnut (*Juglans regia* L.) genotypes in Markazi province of Iran. Crop Breed J 2:119–124

Grauke LJ, Thompson TE, Mendoza-Herrera MA (2012) Native walnuts of Texas. Acta Hortic 948:199–210

Guàrdia M, Charrier G, Vilanova A, Savé R, Ameglio T, Aletà N (2016) Genetics of frost hardiness in *Juglans regia* L. and relationship with growth and phenology. Tree Genet Genomes 12(5):83. https://doi.org/10.1007/s11295-016-1038-3

Gunn BF, Aradhya M, Salick JM, Miller AJ, Yongping Y, Lin L, Xian H (2010) Genetic variation in walnuts (*Juglans regia* and *J. sigillata*; *Juglandaceae*) : species distinctions, human impacts, and the conservation of agrobiodiversity in Yunnan, China. Am J Bot 97(4): 660–671. https://doi.org/10.3732/ajb.0900114

Haghjooyan R, Ghareyazi B, Sanei Shariat-Panahi M, Khalighi A (2005) Investigation of genetic variation of walnut in some region of Iran by using quantitative morphological characters. Pajouhesh Sazandegi 69:22–30

Han H, Woeste KE, Hu Y, Dang M, Zhang T, Gao XX, Zhou H, Feng X, Zhao G, Zhao P (2016) Genetic diversity and population structure of common walnut (*Juglans regia*) in China based on EST-SSRs and the nuclear gene phenylalanine ammonia-lyase (PAL). Tree Genet Genomes 12(6):111. https://doi.org/10.1007/s11295-016-1064-1

Hansche PE, Beres V, Forde HI (1972) Estimates of quantitative genetic properties of walnut and their implications for cultivar improvement. J Am Soc Hortic Sci 97:279–285

Hassani D, Dastjerdi R, Haghjooyan R et al (2014) Genetic improvement of Persian walnut (*Juglans regia* L.) in Iran. Acta Hortic 1050:95–102

He F, Wang H, Zhang Z et al (2010) Identification of walnut cultivars with AFLP fingerprinting. Acta Hortic 861:151–154

Hillis DM (1987) Molecular versus morphological approaches to systematics. Ann Rev Ecol Syst 18(1):23–42. https://doi.org/10.1146/annurev.es.18.110187.000323

Hoban S, Anderson R, McCleary T et al (2008) Thirteen nuclear microsatellite loci for butternut (*Juglans cinerea* L.) Mol Ecol Resour 8(3):643–646. https://doi.org/10.1111/j.1471-8286.2007.02030.x

Hu Y-H, Zhao P, Zhang Q, Wang Y, Gao XX, Zhang T, Zhou HJ, Dang M, Woeste KE (2015) *De novo* assembly and characterization of transcriptome using Illumina sequencing and development of twenty five microsatellite markers for an endemic tree *Juglans hopeiensis* Hu in China. Biochem Syst Ecol 63:201–211. https://doi.org/10.1016/j.bse.2015.10.011

Hu Y, Woeste KE, Dang M, Zhou T, Feng X, Zhao G, Liu Z, Li Z, Zhao P (2016) The complete chloroplast genome of common walnut (*Juglans regia*). Mitochondrial DNA Part B 1(1):189–190. https://doi.org/10.1080/23802359.2015.1137804

Hu Y, Woeste KE, Zhao P (2017) Completion of the chloroplast genomes of five Chinese *Juglans* and their contribution to chloroplast phylogeny. Front Plant Sci 7:1955. https://doi.org/10.3389/fpls.2016.01955

Huang Z, Meilan R, Woeste K (2009) A KNAT3-like homeobox gene from *Juglans nigra* L., JnKNAT3-like, highly expressed during heartwood formation. Plant Cell Rep 28(11):1717–1724. https://doi.org/10.1007/s00299-009-0771-6

Hunter RL, Markert CL (1957) Histochemical demonstration of enzymes separated by zone electrophoresis in starch gels. Science 125(3261): 1294–1295. https://doi.org/10.1126/science.125.3261.1294-a

Ikhsan AS, Topçu H, Sütyemez M, Kafkas S (2016) Novel 307 polymorphic SSR markers from BAC-end sequences in walnut (*Juglans regia* L.): effects of motif types and repeat lengths on polymorphism and genetic diversity. Sci Hortic 213:1–4. https://doi.org/10.1016/j.scienta.2016.10.006

Jacquenet S, Moneret-Vautrin DA (2007) Classification botanique des fruits à coque. Rev Fr Allergol 47:487–491

Jahanbani R, Ghaffari SM, Salami M, Vahdati K, Sepehri H, Sarvestani NN, Sheibani N, Moosavi-Movahedi AA (2016) Antioxidant and anticancer activities of walnut (*Juglans regia* L.) protein hydrolysates using different proteases. Plant Foods Hum Nutr 71(4):402–409. https://doi.org/10.1007/s11130-016-0576-z

Jaynes RA (1969) Handbook of North American nut trees. Northern Nut Growers Association

Jenner RA (2004) Accepting partnership by submission? Morphological phylogenetics in a molecular millennium. Syst Biol 53(2):333–342. https://doi.org/10.1080/10635150490423962

Ji A, Wang Y, Wu G, Wu W, Yang H, Wang Q (2014) Genetic diversity and population structure of North China mountain walnut revealed by ISSR. Am J Plant Sci 5(21):3194–3202. https://doi.org/10.4236/ajps.2014.521335

Kafkas S, Ozkan H, Sütyemez M (2005) DNA polymorphism and assessment of genetic relationships in walnut genotypes based on AFLP and SAMPL markers. J Am Soc Hortic Sci 130:585–590

Karadag H, Akça Y (2011) Phenological and pomological properties of promising walnut (*Juglans regia* L.) genotypes from selected native population in Amasya Province. Afr J Biotechnol 10:16763–16768

Karimi R, Ershadi A, Vahdati K, Woeste K (2010) Molecular characterization of Persian walnut populations in Iran with microsatellite markers. Hortscience 45:1403–1406

Keneni G, Bekele E, Imtiaz M, Dagne K (2012) Genetic vulnerability of modern crop cultivars: causes, mechanism and remedies. Int J Plant Res 2(3):69–79. https://doi.org/10.5923/j.plant.20120203.05

Keqiang Y, Yuejin W, Yindong Z, Xueqin Z (2002) RAPD analysis for the identification of the precocious trait in walnuts. Acta Hortic Sin 2002:06

Khadivi-Khub A, Ebrahimi A, Sheibani F, Esmaeili A (2015) Phenological and pomological characterization of Persian walnut to select promising trees. Euphytica 205(2):557–567. https://doi.org/10.1007/s10681-015-1429-9

Khan MW, Khan IA, Ahmad H et al (2010) Estimation of genetic diversity in walnut. Pak J Bot 42:1791–1796

Kluepfel D, Leslie C, Aradhya M, Browne G (2015) Development of disease-resistant walnut rootstocks: integration of conventional and genomic approaches (SCRI-match Year 3). Walnut Res Rep

Komanich IG (1980) Correlations between walnut fruit characters in a seed population. Bull Acad Sci RSS Mold 2:27–30

Kornienko NA (1974) Types of dichogamy in walnut. Plant Breed Abstr 44:452

Leslie CE, McGranahan GH (1988) Native populations of *Juglans regia*: a draft. In: International Conference on Walnuts. Hort. Research Inst., ed. 124, Yalova, Turkey, Ataturk Central

Leslie CA, McGranahan GH, Dandekar AM et al (2001) Development and field-testing of walnuts expressing the cry1A(c) gene for lepidopteran insect resistance. Acta Hortic 544:195–199

Leslie C, McGranahan G, Hackett W, et al (2015) Walnut improvement program 2015. Walnut Res Rep

Li Z, Lanying Z, Qianwen X (2007) Identification of RAPD markers linked to thickness gene of shuck in walnut. Adv Biol Res 5–6: 137–140

Liao Z, Feng K, Chen Y et al (2014) Genome-wide discovery and analysis of single nucleotide polymorphisms and insertions/deletions in *Juglans regia* L. by high-throughput pyrosequencing. Plant Omics J 7:445–449

Liu X, Walawage SL, Leslie CA, Dandekar AM, Tricoli DM, Hu H, Huang Y, Zhang J, Xv C, Huang J, Zhang Q (2017) In vitro gene expression and mRNA translocation from transformed walnut (*Juglans regia*) rootstocks expressing DsRED fluorescent protein to wild-type scions. Plant Cell Rep 36(6):877–885. https://doi.org/10.1007/s00299-017-2116-1

Loreti S, Gallelli A, Belisario A, Wajnberg E, Corazza L (2001) Investigation of genomic variability of *Xanthomonas arboricola* pv. *juglandis* by AFLP analysis. Eur J Plant Pathol 107(6):583–591. https://doi.org/10.1023/A:1017951406237

Luo M-C, You FM, Li P, Wang JR, Zhu T, Dandekar AM, Leslie CA, Aradhya M, McGuire PE, Dvorak J (2015) Synteny analysis in Rosids with a walnut physical map reveals slow genome evolution in long-lived woody perennials. BMC Genomics 16(1):707. https://doi.org/10.1186/s12864-015-1906-5

Ma Q, Zhang J, Pei D (2011) Genetic analysis of walnut cultivars in China using fluorescent amplified fragment length polymorphism. J Am Soc Hortic Sci 136:422–428

Maguire LS, O'Sullivan SM, Galvin K et al (2004) Fatty acid profile, tocopherol, squalene and phytosterol content of walnuts, almonds, peanuts, hazelnuts and the macadamia nut. Int J Food Sci Nutr 55(3):171–178. https://doi.org/10.1080/09637480410001725175

Mahmoodi R, Rahmani F, Rezaee R (2013) Genetic diversity among *Juglans regia* L. genotypes assessed by morphological traits and microsatellite markers. Span J Agric Res 11(2):431–437. https://doi.org/10.5424/sjar/2013112-3445

Mahmoodi R, Hassani D, Amiri ME, Jaffaraghaei M (2016) Phenological and pomological characteristics of five promised walnut genotypes in Karaj, Iran. J Nuts 7:1–8

Majacka AD (1971) Dichogamy and fertility in walnut. Plant Breed Abstr 41:186

Malvolti ME, Paciucci M, Cannata F, Fineschi S (1993) Genetic variation in Italian populations of *Juglans regia* L. Acta Hortic 311:86–94

Malvolti ME, Fineschi S, Pigliucci M (1994) Morphological integration and genetic variability in *Juglans regia* L. J Hered 85(5):389–394. https://doi.org/10.1093/oxfordjournals.jhered.a111484

Malvolti ME, Fornari B, Maccaglia E, Cannata F (2001) Genetic linkage mapping in an intraspecific cross of walnut (*Juglans regia* L.) using molecular markers. Acta Hortic 544:179–185

Malvolti ME, Pollegioni P, Bertani A et al (2010) *Juglans regia* provenance research by molecular, morphological and biochemical markers : a case study in Italy. Bioremediation Biodivers Bioavailab 4:84–92

Manchester SR (1987) The fossil history of *Juglandaceae*. Missouri Botanical Garden

Manning WE (1978) The classification within the *Juglandaceae*. Ann Mo Bot Gard 65(4):1058–1087. https://doi.org/10.2307/2398782

Mao X, Hua Y, Chen G (2014) Amino acid composition, molecular weight distribution and gel electrophoresis of walnut (*Juglans regia* L.) proteins and protein fractionations. Int J Mol Sci 15(2):2003–2014. https://doi.org/10.3390/ijms15022003

Martínez ML, Labuckas DO, Lamarque AL, Maestri DM (2010) Walnut (*Juglans regia* L.) : genetic resources, chemistry, by-products. J Sci Food Agric 90(12):1959–1967. https://doi.org/10.1002/jsfa.4059

Martínez-García PJ, Crepeau MW, Puiu D, Gonzalez-Ibeas D, Whalen J, Stevens KA, Paul R, Butterfield TS, Britton MT, Reagan RL, Chakraborty S, Walawage SL, Vasquez-Gross HA, Cardeno C, Famula RA, Pratt K, Kuruganti S, Aradhya MK, Leslie CA, Dandekar AM, Salzberg SL, Wegrzyn JL, Langley CH, Neale DB (2016) The walnut (*Juglans regia*) genome sequence reveals diversity in genes coding for the biosynthesis of nonstructural polyphenols. Plant J 87(5):507–532. https://doi.org/10.1111/tpj.13207

Mason AS, Zhang J, Tollenaere R, Vasquez Teuber P, Dalton-Morgan J, Hu L, Yan G, Edwards D, Redden R, Batley J (2015) High-throughput genotyping for species identification and diversity assessment in germplasm collections. Mol Ecol Resour 15(5):1091–1101. https://doi.org/10.1111/1755-0998.12379

McGranahan G, Leslie C (2012) Walnut. In: Badenes ML, Byrne DH (eds) Fruit Breeding. Springer, Boston, pp 827–846. https://doi.org/10.1007/978-1-4419-0763-9_22

McGranahan GH, Tulecke W, Arulsekar S, Hansen JJ (1986) Intergeneric hybridization in the *Juglandaceae*: *Pterocarya* sp. x *Juglans regia*. J Am Soc Hortic Sci 111:627–630

McGranahan GH, Leslie CA, Uratsu SL, Martin LA, Dandekar AM (1988) Agrobacterium-mediated transformation of walnut somatic embryos and regeneration of transgenic plants. Nat Biotechnol 6(7): 800–804. https://doi.org/10.1038/nbt0788-800

McGranahan GH, Leslie CA, Uratsu SL, Dandekar AM (1990) Improved efficiency of the walnut somatic embryo gene transfer system. Plant Cell Rep 8(9):512–516. https://doi.org/10.1007/BF00820198

Menzel A, Fabian P (1999) Growing season extended in Europe. Nature 397(6721):659. https://doi.org/10.1038/17709

Miletic R, Zikic M, Mitic N, Nikolic R (2003) Biological and pomological characteristics of superior walnut selections. Genetika 35(2): 123–130. https://doi.org/10.2298/GENSR0302123M

Mohsenipoor S, Vahdati K, Amiri R, Mozaffari MR (2010) Study of the genetic structure and gene flow in Persian walnut (*Juglans regia* L.) using SSR markers. Acta Hortic 861:133–142

Moldenke HN, Moldenke AL (1952) Plants of the Bible. Chronica Botanica Co, Waltham

Morgante M, Olivieri AM (1993) PCR-amplified microsatellites as markers in plant genetics. Plant J 3(1):175–182. https://doi.org/10.1111/j.1365-313X.1993.tb00020.x

Mozaffarian F, Mardi M, Sarafrazi A, Ganbalani GN (2008) Assessment of geographic and host-associated population variations of the carob

moth, *Ectomyelois ceratoniae*, on pomegranate, fig, pistachio and walnut, using AFLP markers. J Insect Sci 8:6. https://doi.org/10.1673/031.008.0601

Muir R, Baek J, Leslie A, et al (2004) Analysis of genes expressed in walnut seed coat tissue. Unpublished raw data

Nagaraj SH, Gasser RB, Ranganathan S (2006) A hitchhiker's guide to expressed sequence tag (EST) analysis. Brief Bioinform 8(1):6–21. https://doi.org/10.1093/bib/bbl015

Najafi F, Mardi M, Fakheri B, Pirseyedi SM, Mehdinejad N, Farsi M (2014) Isolation and characterization of novel microsatellite markers in walnut (*Juglans regia* L.) Am J Plant Sci 5(03):409–415. https://doi.org/10.4236/ajps.2014.53054

Neale D, Martínez-García PJ, Crepeau MW, et al (2015) Application of marker breeding in the walnut improvement program. Walnut Res Rep

Nei M (1972) Genetic distance between populations. Am Nat 106(949): 283–292. https://doi.org/10.1086/282771

Nicese FP, Hormaza JI, McGranahan GH (1998) Molecular characterization and genetic relatedness among walnut (*Juglans regia* L.) genotypes based on RAPD markers. Euphytica 101(2):199–206. https://doi.org/10.1023/A:1018390120142

Ninot A, Aletà N (2003) Identification and genetic relationship of Persian walnut genotypes using isozyme markers. J Am Pomol Soc 57:106–114

Noor Shah U, Mir JI, Ahmed N, Fazili KM (2016) Assessment of germplasm diversity and genetic relationships among walnut (*Juglans regia* L.) genotypes through microsatellite markers. J Saudi Soc Agric Sci. https://doi.org/10.1016/j.jssas.2016.07.005

Pan C, Ye L, Qin L, Liu X, He Y, Wang J, Chen L, Lu G (2016) CRISPR/Cas9-mediated efficient and heritable targeted mutagenesis in tomato plants in the first and later generations. Sci Rep 6(1):24765. https://doi.org/10.1038/srep24765

Panth N, Paudel KR, Karki R (2016) Phytochemical profile and biological activity of *Juglans regia*. J Integr Med 14(5):359–373. https://doi.org/10.1016/S2095-4964(16)60274-1

Paterson AH (1996) Making genetic maps. In: Paterson AH (ed) Genome mapping in plants. Academic Press, Austin

Payghamzadeh K, Kazemitabar SK (2011) *In vitro* propagation of walnut—a review. Afr J Biotechnol 10:290–311

Pollegioni P, Bartoli S, Cannata F, Malvolti ME (2003) Genetic differentiation of four Italian walnut (*Juglans regia* L.) varieties by inter simple sequence repeat (ISSR). J Genet Breed 57:231–240

Pollegioni P, Woeste K, Major A et al (2008) Characterization of *Juglans nigra* (L.), *Juglans regia* (L.) and *Juglans* x *intermedia* (Carr.) by SSR markers : a case study in Italy. Silvae Genet 57:68–78

Pollegioni P, Woeste K, Mugnozza GS, Malvolti ME (2009) Retrospective identification of hybridogenic walnut plants by SSR fingerprinting and parentage analysis. Mol Breed 24(4):321–335. https://doi.org/10.1007/s11032-009-9294-7

Pollegioni P, Woeste K, Olimpieri I, Marandola D, Cannata F, Emilia Malvolti M (2011) Long-term human impacts on genetic structure of Italian walnut inferred by SSR markers. Tree Genet Genomes 7(4):707–723. https://doi.org/10.1007/s11295-011-0368-4

Pollegioni P, Woeste K, Chiocchini F, Del Lungo S, Ciolfi M, Olimpieri I, Tortolano V, Clark J, Hemery GE, Mapelli S, Malvolti ME (2017) Rethinking the history of common walnut (*Juglans regia* L.) in Europe: its origins and human interactions. PLoS One 12(3): e0172541. https://doi.org/10.1371/journal.pone.0172541

Pop IF, Pamfil D, Raica P et al (2010) Assessment of the genetic variability among some *Juglans* cultivars from the Romanian National Collection at S.C.D.P. Vâlcea using RAPD markers. Romanian Biotechnol Lett 15:41–49

Pop IF, Vicol AC, Botu M, Raica PA, Vahdati K, Pamfil D (2013) Relationships of walnut cultivars in a germplasm collection: comparative analysis of phenotypic and molecular data. Sci Hortic 153: 124–135. https://doi.org/10.1016/j.scienta.2013.02.013

Potter D, Gao F, Aiello G et al (2002) Intersimple sequence repeat markers for fingerprinting and determining genetic relationships of walnut (*Juglans regia*) cultivars. J Am Soc Hortic Sci 127:75–81

Qi J, Hao Y, Zhu Y et al (2011) Studies on germplasm of *Juglans* by EST-SSR markers. Acta Hortic Sin 38:441–448

Qianwen X, Kaizhi W, Lanying Z et al (2010) RAPD markers and heterotic effect of walnut quality in Sichuan of China. Adv Biol Res 4: 81–85

Qing Guo M, Jing Q, Dong P (2010) FISH-AFLP analysis of genetic diversity of early-fruiting walnut cultivars. For Res 23:631–636

Rao G, Sui J, Zhang J (2016) Metabolomics reveals significant variations in metabolites and correlations regarding the maturation of walnuts (*Juglans regia* L.) Biol Open 5(6):829–836. https://doi.org/10.1242/bio.017863

Rehder A (1947) Manual of cultivated trees and shrubs. In: The MacMillan Co. ed pp 115–124

Rezaee R, Vahdati K, Valizadeh M (2009) Variability of seedling vigour in Persian walnut as influenced by the vigour and bearing habit of the mother tree. J Hortic Sci Biotechnol 84(2):228–232. https://doi.org/10.1080/14620316.2009.11512509

Robichaud RL, Glaubitz JC, Rhodes OE, Woeste KE (2006) A robust set of black walnut microsatellites for parentage and clonal identification. New For 32(2):179–196. https://doi.org/10.1007/s11056-005-5961-7

Roor W, Konrad H, Mamadjanov D, Geburek T (2017) Population differentiation in common walnut (*Juglans regia* L.) across major parts of its native range—insights from molecular and morphometric data. J Hered 108(4):391–404. https://doi.org/10.1093/jhered/esw122

Ros E (2010) Health benefits of nut consumption. Nutrients 2(7):652–682. https://doi.org/10.3390/nu2070652

Ross-Davis A, Huang Z, McKenna J, Ostry M, Woeste K (2008) Morphological and molecular methods to identify butternut (*Juglans cinerea*) and butternut hybrids: relevance to butternut conservation. Tree Physiol 28(7):1127–1133. https://doi.org/10.1093/treephys/28.7.1127

Ruiz-Garcia L, Lopez-Ortega G, Denia AF, Tomas DF (2011) Identification of a walnut (*Juglans regia* L.) germplasm collection and evaluation of their genetic variability by microsatellite markers. Span J Agric Res 9(1):179–192. https://doi.org/10.5424/sjar/20110901-227-10

Schell J (1975) The role of plasmids in crown gall formation by *A. tumefaciens*. In: Ledoux L (ed) Genetic manipulations with plant materials. Plenum Press, New York, pp 163–181. https://doi.org/10.1007/978-1-4684-2763-9_11

Schlötterer C (2004) The evolution of molecular markers—just a matter of fashion? Nat Rev Genet 5(1):63–69. https://doi.org/10.1038/nrg1249

Semagn K, Bjørnstad A, Ndjiondjop MN (2006) Principles, requirements and prospects of genetic mapping in plants. Afr J Biotechnol 5: 2569–2587

Sen SM (1985) Correlations between shell thickness, shell cracking resistance, shell seal and shell upright cracking resistance and some other fruit quality characters in walnut (*Juglans regia* L.) Doga Bilim Derg D2 Tarm Ve Orman 9:10–24

Sharma OC (1996) Variability in seedlings trees of Persian walnut in Solan area of Himachal Pardesh. Thesis, Solan University of Horticulture and Forestry, India

Sharma OC, Sharma SD (2001) Genetic divergence in seedling trees of Persian walnut (*Juglans regia* L.) for various metric nut and kernel characters in Himachal Pradesh. Sci Hortic 88(2):163–171. https://doi.org/10.1016/S0304-4238(00)00204-1

Sheikh Beig Goharrizi MA, Dejahang A, Tohidfar M et al (2016) Agrobacterium mediated transformation of somatic embryos of Persian walnut using *fld* gene for osmotic stress tolerance. J Agric Sci Technol 18:423–435

Sholokhov LV (1974) Quality characters of fruit in walnut and correlation between them. Tr Novocherkas Inzha Melior Inta 13:106–112

Singh RK, Mishra GP, Thakur AK, Singh SB (2008) Molecular markers in plants. In: Singh RK, Singh R, Ye G et al (eds) Molecular plant breeding: principle, method and application. Studium Press LLC, Texas, pp 35–78

Smith JSC, Smith OS (1989) The description and assessment of distance between inbred lines of maize. 2: The utility of morphological - biochemical - and genetic descriptors and a scheme for the testing of distinctiveness between inbred lines. Maydica 34:151–161

Solar A, Stampar F (2011) Characterisation of selected hazelnut cultivars: phenology, growing and yielding capacity, market quality and nutraceutical value. J Sci Food Agric 91(7):1205–1212. https://doi.org/10.1002/jsfa.4300

Solar A, Smole J, Stampar F, Virščekk-Marn M (1994) Characterization of isozyme variation in walnut (*Juglans regia* L.) Euphytica 77(1-2): 105–112. https://doi.org/10.1007/BF02551471

Solar A, Hudina M, Stampar F (2001) Relationship between tree architecture, phenological data and generative development in walnut (*Juglans regia* L.) Acta Hortic 544:275–285

Solar A, Colaric M, Usenik V, Stampar F (2006) Seasonal variations of selected flavonoids, phenolic acids and quinones in annual shoots of common walnut (*Juglans regia* L.) Plant Sci 170(3):453–461. https://doi.org/10.1016/j.plantsci.2005.09.012

Stanford AM, Harden R, Parks CR (2000) Phylogeny and biogeography of *Juglans* (*Juglandaceae*) based on matK and ITS sequence data. Am J Bot 87(6):872–882. https://doi.org/10.2307/2656895

Sun X, Liu D, Zhang X, Li W, Liu H, Hong W, Jiang C, Guan N, Ma C, Zeng H, Xu C, Song J, Huang L, Wang C, Shi J, Wang R, Zheng X, Lu C, Wang X, Zheng H (2013) SLAF-seq: an efficient method of large-scale *De Novo* SNP discovery and genotyping using high-throughput sequencing. PLoS One 8(3):e58700. https://doi.org/10.1371/journal.pone.0058700

Sütyemez M (2006) Comparison of AFLP polymorphism in progeny derived from dichogamous and homogamous walnut genotypes. Pak J Biol Sci 9:2303–2307

Sütyemez M (2016) New walnut cultivars: Maras 18, Sütyemez 1, and Kaman 1. Hortscience 51(10):1301–1303. https://doi.org/10.21273/HORTSCI10972-16

Taha NA, Al-Wadaan MA (2011) Utility and importance of walnut, *Juglans regia* Linn: a review. Afr J Microbiol Res 5:5796–5805

Tapsell LC, Batterham MJ, Teuss G, Tan SY, Dalton S, Quick CJ, Gillen LJ, Charlton KE (2009) Long-term effects of increased dietary polyunsaturated fat from walnuts on metabolic parameters in type II diabetes. Eur J Clin Nutr 63(8):1008–1015. https://doi.org/10.1038/ejcn.2009.19

Tester M, Langridge P (2010) Breeding technologies to increase crop production in a changing world. Science 327(5967):818–822. https://doi.org/10.1126/science.1183700

Tian J (2014) Proceedings of the VIIth International Walnut Symposium: Fenyang, China, July 20–23, 2013. Fenyang, China. Acta Hort 1050

Topçu H, Ikhsan AS, Sütyemez M, Çoban N, Güney M, Kafkas S (2015) Development of 185 polymorphic simple sequence repeat (SSR) markers from walnut (*Juglans regia* L.) Sci Hortic 194:160–167. https://doi.org/10.1016/j.scienta.2015.08.014

Tulecke W, McGranahan G (1985) Somatic embryogenesis and plant regeneration from cotyledons of walnut, *Juglans regia* L. Plant Sci 40(1):57–63. https://doi.org/10.1016/0168-9452(85)90163-3

Tulecke W, McGranahan G (1994) The walnut germplasm collection of UC Davis. DANR, UC Davis

US Department of Agriculture, Agricultural Research Service, Nutrient Data Laboratory (2015) USDA National Nutrient Database for Standard Reference, Release 28. https://ndb.nal.usda.gov/ndb/. Accessed 9 Dec 2016

Vahdati K, McKenna J, Dandekar AM et al (2002) Rooting and other characteristics of a transgenic walnut hybrid (*Juglans hindsii* x *J. regia*) rootstock expressing rolABC. J Am Soc Hortic Sci 127: 724–728

Victory ER, Glaubitz JC, Rhodes OE Jr, Woeste KE (2006) Genetic homogeneity in *Juglans nigra* (*Juglandaceae*) at nuclear microsatellites. Am J Bot 93(1):118–126. https://doi.org/10.3732/ajb.93.1.118

Vischi M, Chiabà C, Raranciuc S, Poggetti L, Messina R, Ermacora P, Cipriani G, Paffetti D, Vettori C, Testolin R (2017) Genetic diversity of walnut (*Juglans regia* L.) in the Eastern Italian Alps. Forests 8(3): 81–94. https://doi.org/10.3390/f8030081

Vos P, Hogers R, Bleeker M, Reijans M, Lee T, Hornes M, Friters A, Pot J, Paleman J, Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23(21):4407–4414. https://doi.org/10.1093/nar/23.21.4407

Vyas D, Sharma SK, Sharma DR (2003) Genetic structure of walnut genotype using leaf isozymes as variability measure. Sci Hortic 97(2):141–152. https://doi.org/10.1016/S0304-4238(02)00129-2

Walawage SL, Britton MT, Leslie CA, Uratsu SL, Li YY, Dandekar AM (2013) Stacking resistance to crown gall and nematodes in walnut rootstocks. BMC Genomics 14(1):668. https://doi.org/10.1186/1471-2164-14-668

Wang DG, Fan JB, Siao CJ, Berno A, Young P, Sapolsky R, Ghandour G, Perkins N, Winchester E, Spencer J, Kruglyak L, Stein L, Hsie L, Topaloglou T, Hubbell E, Robinson E, Mittmann M, Morris MS, Shen N, Kilburn D, Rioux J, Nusbaum C, Rozen S, Hudson TJ, Lipshutz R, Chee M, Lander ES (1998) Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science 280(5366):1077–1082. https://doi.org/10.1126/science.280.5366.1077

Wang H, Pei D, Gu R, Wang B (2008) Genetic diversity and structure of walnut populations in central and southwestern China revealed by microsatellite markers. J Am Soc Hortic Sci 133:197–203

Wang H, Zhao S, Zhang Z et al (2010) Genetic relationship and diversity of eight *Juglans* species in China estimated through AFLP analysis. Acta Hortic 861:143–150

Wang G, Wang JY, He Q et al (2014a) Agronomic evaluation and heritability of Jin RS-2 and Jin RS-3 walnut rootstocks. Acta Hortic 1050:113–121

Wang GA, Zhang Q, Huang MM, Yakup A (2014b) The breeding of six Xinjiang dwarf walnut cultivars. Acta Hortic 1050:151–159

Wang H, Wu W, Pan G, Pei D (2016) Analysis of genetic diversity and relationships among 86 Persian walnut (*Juglans regia* L.) genotypes in Tibet using morphological traits and SSR markers. J Hortic Sci Biotechnol 90:563–570

Wani N, Bhat MA, Ahmad MF et al (2010) Molecular markers and their application in walnut improvement. Int J Curr Res 3:6–11

Way DA (2011) Tree phenology responses to warming: spring forward, fall back? Tree Physiol 31(5):469–471. https://doi.org/10.1093/treephys/tpr044

Williams JG, Kubelik AR, Livak KJ et al (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res 18(22):6531–6535. https://doi.org/10.1093/nar/18.22.6531

Woeste K, McGranahan G, Bernatzky R (1996a) The identification and characterization of a genetic marker linked to hypersensitivity to the cherry leafroll virus in walnut. Mol Breed 2(3):261–266. https://doi.org/10.1007/BF00564203

Woeste K, McGranahan GH, Bernatzky R (1996b) Randomly amplified polymorphic DNA loci from a walnut backcross [(*Juglans hindsii* x *J. regia*) x *J. regia*]. J Am Soc Hortic Sci 121:358–361

Woeste K, McGranahan G, Bernatzky R (1998) Low correlation between genomic and morphological introgression estimates in a walnut backcross. J Am Soc Hortic Sci 123:258–263

Woeste K, Burns R, Rhodes O, Michler C (2002) Thirty polymorphic nuclear microsatellite loci from black walnut. J Hered 93(1):58–60. https://doi.org/10.1093/jhered/93.1.58

Wolfe JA (1978) A paleobotanical interpretation of tertiary climates in the northern hemisphere: data from fossil plants make it possible to reconstruct tertiary climatic changes, which may be correlated with changes in the inclination of the earth's rotational axis. Am Sci 66: 694–703

Woodroof JG (1967) Tree nuts: production, processing, products. AVI Pub. Co.

Woodworth RH (1930) Meiosis of micro-sporogenesis in the Juglandaceae. Am J Bot 17(9):863–869. https://doi.org/10.2307/2435868

Wu J, Gu YQ, Hu Y, You FM, Dandekar AM, Leslie CA, Aradhya M, Dvorak J, Luo MC (2012) Characterizing the walnut genome through analyses of BAC end sequences. Plant Mol Biol 78(1-2): 95–107. https://doi.org/10.1007/s11103-011-9849-y

Xu F, Deng G, Cheng S, Zhang W, Huang X, Li L, Cheng H, Rong X, Li J (2012a) Molecular cloning, characterization and expression of the phenylalanine ammonia-lyase gene from *Juglans regia*. Molecules 17(12):7810–7823. https://doi.org/10.3390/molecules17077810

Xu Z, Hu T, Zhang F (2012b) Genetic diversity of walnut revealed by AFLP and RAPD markers. J Agric Sci 4:271–276

Xu L, Chen X, Zhang LS, Liu QZ (2014) Molecular cloning and expression analysis of the transcription factor gene *JrCBF* from *Juglans regia* L. Acta Hortic 1050:41–47

Yarilgac T, Koyuncu F, Koyuncu MA et al (2001) Some promising walnut selections (*Juglans regia* L.) Acta Hortic 544:93–96

Yi F, Zhijun Z, Shelong Z (2011) Development of walnut EST-SSR markers and primer design. Agric Sci Technol 12:1810–1813

You FM, Deal KR, Wang J, Britton MT, Fass JN, Lin D, Dandekar AM, Leslie CA, Aradhya M, Luo MC, Dvorak J (2012) Genome-wide SNP discovery in walnut with an AGSNP pipeline updated for SNP discovery in allogamous organisms. BMC Genomics 13(1):354. https://doi.org/10.1186/1471-2164-13-354

Zekiri F, Molitor C, Mauracher SG, Michael C, Mayer RL, Gerner C, Rompel A (2014) Purification and characterization of tyrosinase from walnut leaves (*Juglans regia*). Phytochemistry 101:5–15. https://doi.org/10.1016/j.phytochem.2014.02.010

Zeneli G, Kola H, Dida M (2005) Phenotypic variation in native walnut populations of Northern Albania. Sci Hortic 105(1):91–100. https://doi.org/10.1016/j.scienta.2004.11.003

Zerillo MM, Ibarra Caballero J, Woeste K, Graves AD, Hartel C, Pscheidt JW, Tonos J, Broders K, Cranshaw W, Seybold SJ, Tisserat N (2014) Population structure of *Geosmithia morbida*, the causal agent of thousand cankers disease of walnut trees in the United States. PLoS One 9(11):e112847. https://doi.org/10.1371/journal.pone.0112847

Zeven AC, Zhukovsky PM (1975) Dictionary of cultivated plants and their centres of diversity excluding ornamentals, forest trees, and lower plants. Centre for Agricultural Publishing and Documentation, Wageningen

Zhang R, Zhu A, Wang X et al (2010) Development of *Juglans regia* SSR markers by data mining of the EST database. Plant Mol Biol Report 28(4):646–653. https://doi.org/10.1007/s11105-010-0192-2

Zhang ZY, Han JW, Jin Q, Wang Y, Pang XM, Li YY (2013) Development and characterization of new microsatellites for walnut (*Juglans regia*). Genet Mol Res 12(4):4723–4734. https://doi.org/10.4238/2013.October.18.10

Zhang MY, Xu Y, Xiang K, Wang XL (2014) Review of walnut breeding research at the Shandong Institute of Pomology. Acta Hortic 1050: 55–60

Zhang Q, Walawage SL, Tricoli DM, Dandekar AM, Leslie CA (2015) A red fluorescent protein (DsRED) from *Discosoma* sp. as a reporter for gene expression in walnut somatic embryos. Plant Cell Rep 34(5):861–869. https://doi.org/10.1007/s00299-015-1749-1

Zhao MH, Jiang ZT, Liu T, Li R (2014) Flavonoids in *Juglans regia* L. leaves and evaluation of in vitro antioxidant activity via intracellular and chemical methods. Sci World J 2014 article id 303878

Zhao P, Zhang T, Zhou HJ (2015) Identification, development, and application of 40 polymorphic EST-SSR markers for common walnut (Juglans regia L.). NCBI, EST unpublished raw data

Zhu Y, Yin Y, Yang K, Li J, Sang Y, Huang L, Fan S (2015) Construction of a high-density genetic map using specific length amplified fragment markers and identification of a quantitative trait locus for anthracnose resistance in walnut (*Juglans regia* L.) BMC Genomics 16(1):614. https://doi.org/10.1186/s12864-015-1822-8

Zietkiewicz E, Rafalski A, Labuda D (1994) Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. Genomics 20(2):176–183. https://doi.org/10.1006/geno.1994.1151

2. L'importance des ressources phytogénétiques

Introduction

La notion de « ressources génétiques » est contemporaine de la Révolution verte des années 1960 mais elle a profité d'une certaine reconnaissance au niveau mondial grâce à la Convention sur la Diversité Biologique adoptée lors du Sommet de la Terre de Rio de Janeiro en 1992. Elle stipule notamment le fait que l'Etat est souverain sur son patrimoine naturel. Une définition des ressources génétiques a été énoncée durant cette convention comme étant un « matériel d'origine végétale, animale, microbienne ou autre, contenant des unités fonctionnelles de l'hérédité » ou un « matériel génétique ayant une valeur effective ou potentielle ».

La biodiversité agricole – ou agrobiodiversité – a été largement augmentée par l'être humain à partir d'espèces sauvages qu'il·elle a su domestiquer depuis l'époque de la Préhistoire. En diversifiant le compartiment cultivé de l'espèce domestiquée, il·elle a ainsi créé différentes variétés par un processus de sélection. La sélection végétale, ou plus globalement, l'amélioration des plantes alors inconsciente, a débuté il y a environ 10 000 ans lorsque l'être humain était encore chasseur·euse-cueilleur·euse. Les premiers éleveurs·euses-agriculteurs·trices sont apparu·e·s et ont cohabité avec les premiers·ères en Europe durant 2 000 ans environ ; c'est le début de la maîtrise des ressources naturelles, ou domestication, et de la sédentarité. Aujourd'hui, il existerait dans le monde par exemple, plus de 12 000 variétés de tomates. Mais en parallèle, 75% de la diversité des cultures a été perdue entre 1900 et 2000, estime la FAO en 2010 et seules les céréales comme le riz, le blé, l'avoine, l'orge, le maïs, le mil et le sorgho fournissent 45% des bases alimentaires de l'Humanité toujours selon la FAO en 2003. Actuellement, l'ancêtre du maïs, la téosinte, se fait rare à l'état sauvage en milieu naturel, il est grandement menacé de disparition. D'où l'importance de conserver les ressources génétiques. Ainsi par exemple, le Centro Internacional de Mejoramiento de Maiz Y Trigo (CIMMYT) au Mexique conserve en semences environ 150 000 accessions de blé de plus de 100 pays et environ 28 000 variétés de maïs, dont la plus grande collection de variétés traditionnelles et d'espèces sauvages.

Les enjeux des ressources phytogénétiques

Les ressources phytogénétiques (RPG) font donc l'objet d'enjeux cruciaux ; elles permettent notamment de rechercher des traits désirés en agriculture dans un contexte d'augmentation de la population et de changement climatique (Ulukan 2011; Ogwu et al.

2014). Selon le dernier rapport de l'ONU en 2017, la population atteindra 9,1 milliards d'individus en 2050 et les RPG constituent alors les fondations de la pérennité de l'agriculture et de la sécurité alimentaire mondiale (Hammer and Teklu 2008). Au cours des trois dernières décennies, l'augmentation du nombre de travaux menés sur la découverte de nouvelles accessions de RPG et sur l'exploration de celles existantes a conduit à l'obtention de nombreuses « big data » phénotypiques et génotypiques qui servent à augmenter d'une part, l'efficacité de leur conservation et d'autre part, l'efficacité de leur utilisation pour créer de nouvelles variétés dans un objectif de développement durable. Cela a soulevé des interrogations sur les systèmes de gouvernance de ces ressources et les échanges de matériel (Halewood et al. 2018). Jusqu'à présent, ces questions restaient délicates à cause d'évidents enjeux financiers. Si jusqu'à la fin des années 1960, les RPG faisaient partie du « bien public », les pays développés ont commencé à faire pression en faveur de la reconnaissance de la propriété intellectuelle des nouvelles variétés créées par la mise en place de l'Union internationale pour la Protection des Obtentions Végétales (UPOV). De plus, les données digitales associées à ces ressources physiques demeuraient dispersées dans les différentes organisations et il n'existait pas de système globalement adopté, comme un identifiant unique pour chaque ressource, sauf dans les organismes du réseau de conservation des principales espèces cultivées tels que le CIMMYT. Mais l'inquiétude justifiée des pays en développement a conduit l'Organisation des Nations Unies (ONU) à créer des obligations internationales pour partager les bénéfices. Ainsi, ont été adoptés le Protocole de Nagoya en 2010 sur l'accès aux RPG et le partage juste et équitable des avantages découlant de leur utilisation et l'International Treaty on Plant Genetic Resources for Food and Agriculture (ITPGRFA) en 2001 qui prône notamment l'instauration d'un Digital Object Identifier (DOI) et des protocoles de standardisation pour chaque RPG.

L'état des ressources phytogénétiques dans le monde

La prise de conscience du besoin de partager et de caractériser les RPG ne résoud pas le problème de l'érosion génétique. Nikolaï Ivanovitch Vavilov, botaniste russe de la première moitié du XX^{è} siècle, est un des premiers chercheurs à être conscient de cette érosion génétique, définie comme étant la perte de la variabilité au sein des cultures et traduite notamment par une perte d'hétérogénéité allélique. Avec le développement rapide de nouvelles variétés de haute qualité et homogènes, les variétés locales ou « landraces » sont grandement menacées de disparition (Brush 1995). Mais de façon incontestable, le changement climatique est reconnu comme cause première de cette érosion des RPG (Myers 1994).

Selon le Deuxième Rapport de la FAO de 2010 sur l'état des RPG pour l'alimentation et l'agriculture dans le monde, il existerait 7,4 millions d'accessions (dont plus de 28% représentées par le blé, le riz et l'orge) conservées dans 1 625 banques. Néanmoins, la FAO met en exergue un constat en demi-teinte. Par exemple, le nombre et la couverture des aires protégées a augmenté de 30% durant la dernière décennie, ce qui a eu pour effet d'augmenter le niveau de protection des espèces sauvages apparentées aux plantes cultivées. Néanmoins, des progrès restent à faire en dehors de ces zones protégées. Pour ce qui est de la gestion *ex situ* des RPG, la majorité existe sous forme de semences et certaines collections sont en danger (Fu 2017). La vulnérabilité des RPG conservées *ex situ* s'explique notamment par le fait qu'elles sont généralement sous-financées, qu'il existe un retard dans la régénération dans le cas des banques de semences (il faudrait 13 ans pour renouveler les 150 000 accessions de blé du CIMMYT) et que l'évaluation et la caractérisation sont souvent imprécises ou inadéquates.

Globalement, l'utilisation des marqueurs moléculaires pour caractériser les RPG a augmenté durant les 15 dernières années mais d'un autre côté, la sélection végétale est de plus en plus délaissée du secteur public, là où le secteur privé a su prendre la relève depuis les années 1980 pour les grandes cultures. Ceci pourrait constituer un certain danger dans la mesure où ce secteur se concentre sur un nombre plus restreint de cultures. En ce qui concerne la mise en place de programmes nationaux dédiés aux RPG, parmi les 113 pays ayant fourni des informations pour l'élaboration de ce rapport, 71% dirigent un programme national, comme la France avec, par exemple, son programme national de conservation des ressources génétiques forestières focalisé notamment sur le chêne sessile, le hêtre, l'épicéa commun, le pin sylvestre, etc. Cependant, il est indiqué que les financements restent encore inadéquats et l'accès au public des bases de données nationales se fait encore rare. Enfin, l'ITPGRFA, qui a reconnu l'importance de l'agriculteur·trice pour la diversité génétique, a permis à 8 pays d'adopter des réglementations quant aux droits des agriculteurs·trices. Mais en parallèle, sur les 85 pays reconnaissant les droits de l'obtention, 60 les ont reconnus au cours de la dernière décennie, selon la FAO en 2018.

Le management des ressources phytogénétiques et le recours aux outils de génomique

Le bon management des RPG est crucial depuis leur conservation jusqu'à leur utilisation. La figure 3 présente le modèle global de la conservation des RPG proposé par Maxted et al. en 1997. Dans le cas des espèces fruitières pérennes en général, la conservation des RPG se fait généralement *ex situ*, en vergers, ce qui présente certains avantages et

inconvénients. Les principaux avantages sont que les RPG peuvent être conservées dans les conditions climatiques de leur utilisation escomptée et elles peuvent être évaluées durant leur conservation. Mais en contrepartie, elles demandent beaucoup d'espace, le coût de conservation est conséquent et elles sont exposées aux ravageurs et aux maladies (Hammer and Teklu 2008). De nos jours, les outils de génomique apportent leur contribution à chaque étape du management des RPG (Wambugu et al. 2018). En amont tout d'abord, l'utilisation des séquences des génomes chloroplastiques peut résoudre des questions phylogénétiques, comme cela a été le cas chez le noyer. Ces outils viennent aussi décrypter les bases génétiques et moléculaires de caractères d'intérets agronomiques grâce à la découverte de marqueurs SNP, suite au reséquençage de plusieurs accessions comparées au génome de référence. Chez le noyer, une puce de génotypage à haute densité de 700K SNP est disponible depuis 2017 (Marrano et al. 2019a). Ce type de technologies peut donc servir à réaliser des études de génétique d'association et être utilisés dans les processus de sélection. Aussi, ces outils permettent de trouver les accessions génétiquement proches ou redondantes. Il est donc possible de créer des core collections, qui abriteront le maximum de diversité génétique au sein du plus petit nombre d'accessions, entraînant notamment des réductions de coûts de conservation. Enfin, ces outils peuvent également intervenir dans le cadre d'échanges de RPG. L'une des principales responsabilités d'un organisme impliqué dans ce type d'échange est de fournir un matériel de qualité. Cependant, les tendances protectionnistes, parfois par crainte de biopiraterie, entraînent la réduction des échanges. De plus, des cas de détournement de matériel et de manquements au droit de la propriété intellectuelle ont été signalés (McGrown 2006). Ces phénomènes entravent la bonne application du protocole de Nagoya et de l'ITPGRFA et a instauré un climat de méfiance vis-à-vis de ce type d'échange. Aujourd'hui, la preuve par l'ADN est admissible devant un tribunal (Henry 2012) et offre donc un moyen de protection légal en cas de violation de la propriété intellectuelle.

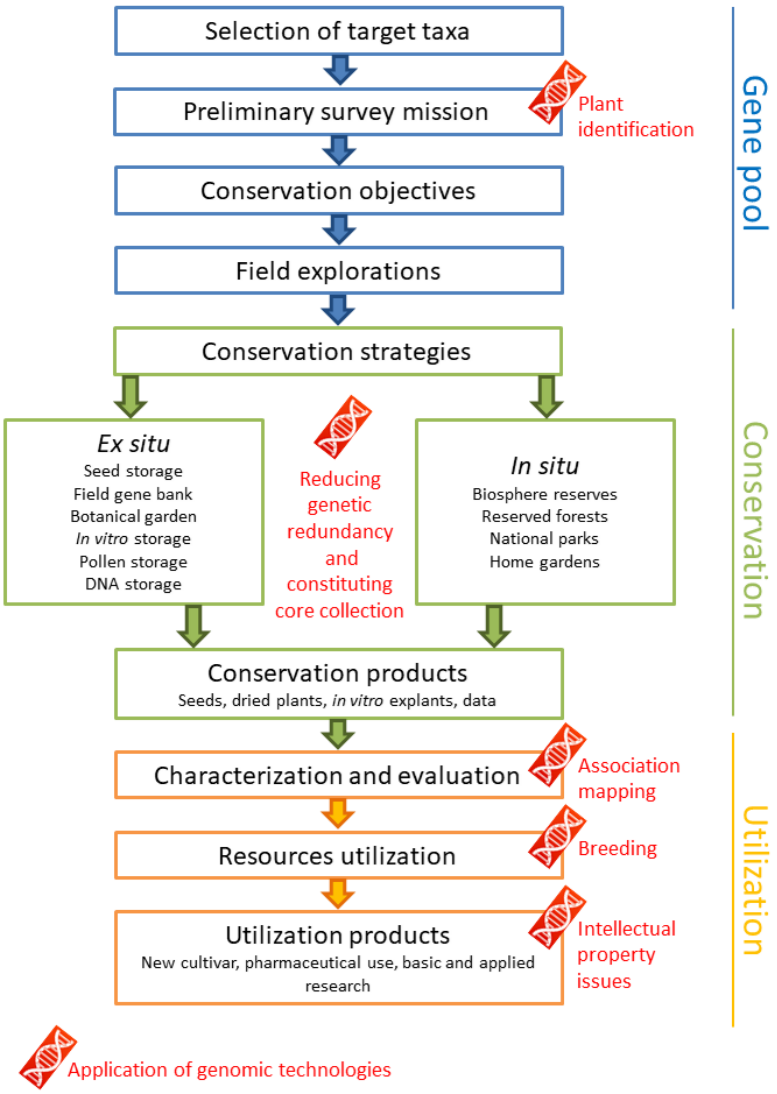


Figure 3 : Modèle simple de la conservation des RPG et apport des outils de génomique.
Le processus de conservation des RPG se décompose en trois principales étapes : la collecte d'un patrimoine génétique, sa conservation et son utilisation, d'après Maxted et al. 1997 et modifié par Hammer and Teklu 2008. Wambugu et al. 2018 montre l'importance de l'apport des outils de génomique à tous les niveaux de ce processus.

En France, le bureau des ressources génétiques, structure nationale du Ministère de la Recherche datant de 1983, avait pour but de collecter, conserver et étudier les RPG. Il a finalement été remodelé pour donner, avec l'institut français de la biodiversité, la fondation pour la recherche sur la biodiversité en 2008. En parallèle en France depuis 2017, en écho aux exigences du protocole de Nagoya et de l'ITPGRFA, des actions sont menées grâce au Groupe d'Etude et de contrôle des Variétés et des Semences (GEVES) et à la section « ressources génétiques » du Comité Technique Permanent de la Sélection (CTPS). Ces actions ont notamment pour but d'identifier les acteurs·trices impliqué·e·s dans la gestion des RPG, de mettre en place des stratégies pour sauvegarder celles en danger et de mettre en

réseau tous ces acteurs·trices. Enfin, au sein de l'INRAE, l'infrastructure Ressources Agronomiques pour la Recherche (RARe) permet de mettre en réseau différents centres de ressources biologiques sur l'animal domestique, la plante cultivée et sauvage, la forêt et le micro-organisme. Cela montre la volonté nationale de conserver nos ressources les plus précieuses.

En ce qui concerne plus particulièrement la collection de noyers de l'INRAE, des difficultés sont également rencontrées dans son management. D'abord, le taux de réussite lors de la multiplication par greffage est particulièrement bas et est lié à l'espèce, aux alentours de 50% environ, contre 90% pour le pommier par exemple. Aussi, le financement de l'entretien de la collection est assuré principalement par l'INRAE en auto-financement, via les ressources de l'Unité Expérimentale, les projets de recherche comme INNOV'noyer, ou la vente des fruits. Concrètement, hors coûts des salaires publics, le coût de l'entretien d'une parcelle de collection de noyers s'élève à 3 200€ par hectare et par an.

L'importance des ressources phytogénétiques du noyer pour la filière nucicole

La biodiversité s'est créée durant des milliards d'années par le jeu des évènements géologiques et climatiques, des interactions entre les espèces et des forces évolutives. Elle fournit à l'être humain des biens indispensables au quotidien comme l'oxygène, la nourriture, les médicaments, le bois et les fibres mais elle offre également des services de la plus haute importance comme la pollinisation des végétaux, la fertilisation des sols et l'épuration de l'air et de l'eau pour ne citer que ces quelques exemples. Mais de nos jours, le constat d'une sixième extinction de masse est fortement appuyé par un travail de synthèse publié en juillet 2017, concernant toutes les populations de vertébrés (Ceballos et al. 2017) et par le dernier rapport d'évaluation de mai 2019 de l'Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES). Ce dernier, fruit de 3 ans de travail avec non moins de 15 000 articles scientifiques synthétisés par 145 experts de 50 pays, affirme le déclin global sans précédent de la biodiversité. Une autre étude récente prédit d'ici à 2080 une perte allant de 48% des animaux et 69% des végétaux si le réchauffement climatique entraîne une augmentation de plus de 2°C des températures moyennes (Warren et al. 2018). Les facteurs de cette actuelle disparition des espèces – ou érosion de la biodiversité – sont principalement attribués à l'activité humaine. Ils ont été identifiés comme étant la disparition progressive des milieux naturels à cause de l'urbanisation, la surexploitation due à la surpêche et à la déforestation, la pollution de l'eau, de l'air et du sol, l'introduction d'espèces invasives liée à la mondialisation et le changement climatique.

L'effet du rechauffement climatique sur la phénologie

Le changement climatique est une réalité. Selon la National Aeronautics and Space Administration (NASA), plusieurs preuves appuient ce constat : l'augmentation de la température mondiale, le réchauffement des océans, la diminution des calottes glaciaires et de la couverture neigeuse, l'augmentation du niveau de la mer, l'augmentation des évènements météorologiques extrêmes et l'acidification des océans due au dioxyde de carbone. Le réchauffement climatique induit par l'être humain est la preuve la plus probante de ce changement climatique, reconnu aujourd'hui comme consensus scientifique (Cook et al. 2016). Le Groupe d'experts Intergouvernemental sur l'Evolution du Climat (GIEC), dans son cinquième rapport de synthèse de 2014 concernant le changement climatique, arrive aux mêmes conclusions : « Human influence on the climate system is clear, and recent anthropogenic emissions of greenhouse gases are the highest in history ». La figure 4 illustre parfaitement le réchauffement climatique par l'augmentation progressive des températures anormales au cours du dernier siècle, avec une progression très inquiétante depuis 2010.

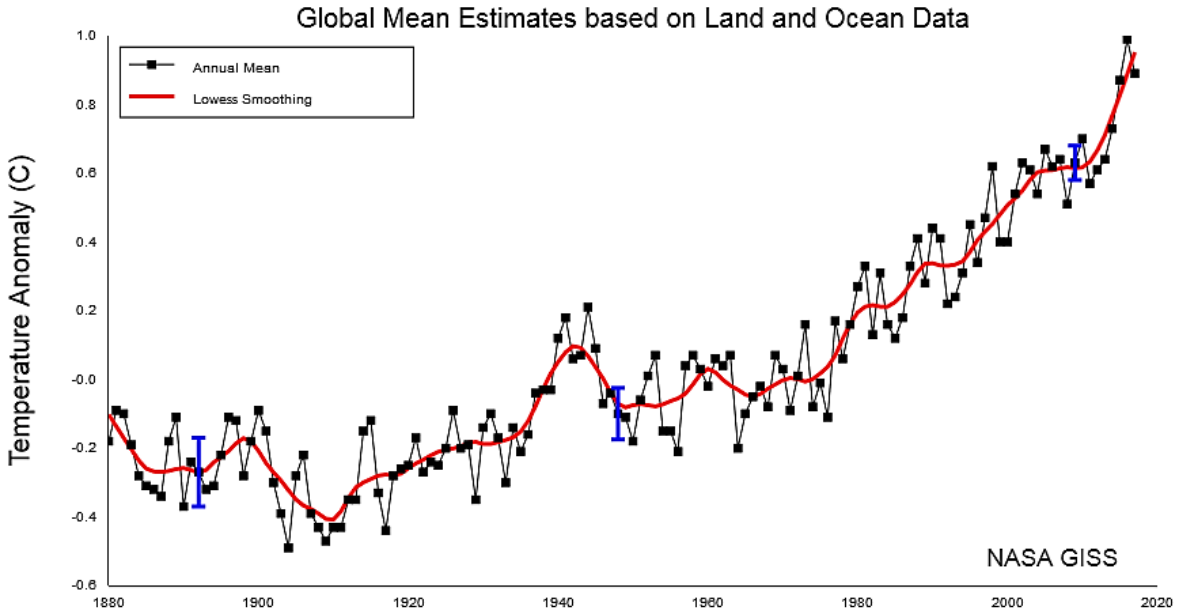


Figure 4 : Changement annuel de la température de l'air de la surface terrestre et de la suface océanique.
Ce graphique représente l'indice d'anomalies de température (°C) pour la période de 1880 à nos jours, estimé grâce aux données moyennes des températures de l'air aux niveaux terrestre et océanique. La ligne noire est la moyenne annuelle globale, la ligne rouge est la courbe de régression « LOWESS » pour Locally Weighted Scatterplot Smoothing. En bleu, les barres d'erreur pour un intervalle de confiance de 95%. Données GISTEMP et Hansen et al. 2010 mises à jour.

L'impact du changement climatique sur les plantes est observé depuis des décennies. Il n'affecte pas seulement le niveau de biodiversité, mais altère également la façon dont la

plante se comporte face à cet environnement changeant, notamment sur la phénologie. Chez le noyer, il existe peu d'études concernant l'effet du changement climatique sur la phénologie. Néanmoins, en Slovénie, la date de débourrement de deux variétés à été étudiée : une variété à date de débourrement plutôt tardive, 'Franquette' et une variété à date de débourrement intermédiaire, 'Geisenheim 139'. Durant la période de 1984 à 1990, la température moyenne entre janvier et avril était de 3,7°C et les dates de débourrement moyennes étaient le 23 avril et le 5 mai pour 'Franquette' et 'Geisenheim 139' respectivement. Durant la période de 2000 à 2006, la température moyenne entre janvier et avril était supérieure de 0,9°C, soit environ 4,6°C et les dates de débourrement moyennes ont avancé de 7 jours pour 'Franquette' et de 3 jours pour 'Geisenheim 139' (Črepinšek et al. 2009). De plus, grâce à des données phénologiques disponibles à l'INRAE de Nouvelle-Aquitaine-Bordeaux, suite à l'important travail de l'équipe d'E. Germain, il est possible de constater le même phénomène en France. Sur la période de 1989 à 2017, une avancée de la date de débourrement de 5 jours en moyenne est observée pour 'Franquette' et 'Lara' en figure 5.

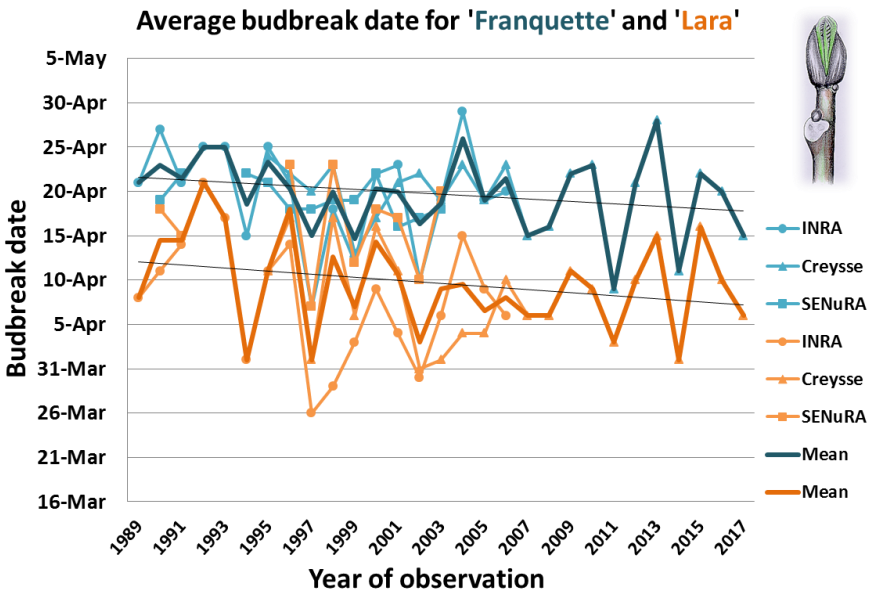


Figure 5 : Date de débourrement moyenne pour 'Franquette' et 'Lara' en France entre 1989 et 2017.
Ce graphique représente la date de débourrement moyenne de 'Franquette' en bleu et 'Lara' en orange entre 1989 et 2017, sur trois sites : UEA de Toulenne, Station Expérimentale de la Noix de Creysse et Station d'Expérimention Nucicole en Rhône-Alpes. La courbe de régression linéaire montre une tendance à l'avancée de la date de débourrement. Données INRAE, valorisées par Bernard et al. 2018 dans le poster « The French walnut improvement program: preliminary investigations » lors du XXX International Horticultural Congress (IHC 2018) à Istanbul, Turquie, du 12 au 16 août 2018.

Le noyer n'est pas la seule espèce touchée. Une étude publiée en 2001 a analysé des données phénologiques européennes du réseau « International Phenological Gardens » concernant le bouleau pubescent, le cerisier, le sorbier et le groseillier pour la période entre

1969 et 1998. En 30 ans, une avancée de 8 jours de la période de croissance moyenne au printemps a été constatée (Chmielewski and Rötzer 2001). Une autre étude s'est penchée sur un ensemble important de données grâce au réseau « International Phenological Gardens » et au réseau « COopération européenne en Science et Technologie (COST) ». Au total, plus de 125 000 séries d'observations de 542 plantes au sein de 21 pays européens entre 1971 et 2000 ont conduit à démontrer une avancée des évènements phénologiques printaniers de 2,5 jours par décennie (Menzel et al. 2006), allant dans le sens de l'étude précédente. Dans le bassin méditerranéen, il a été observé durant les trois dernières décennies une avancée plus rapide de la phénologie des insectes, en comparaison à celle des plantes. Ainsi, certaines interactions plantes-insectes pourraient se retrouver découplées, telles que celles entre fleurs et pollinisateurs (Gordo and Sanz 2005).

La phénologie est un phénomène complexe dépendant du génotype et de l'environnement. En ce qui concerne le débourrement, les bourgeons en dormance deviennent capables de croître après satisfaction de besoins en froid en hiver, ou endodormance, et de besoins en chaud au début du printemps, ou écodormance. Cette dernière période est critique en cas de gelées plus ou moins sévères. Ainsi, il est essentiel de comprendre quelle est la part du génotype ou de l'environnement dans ce phénomène. Une étude réalisée en France sur quatre cultivars de noyer a démontré que la période d'écodormance, outre les effets de l'environnement, est déterminée par un effet important du génotype (Charrier et al. 2011b). Ces résultats démontrent l'importance du choix du cultivar pour la date de débourrement et la possibilité de sélectionner ce caractère en amélioration.

Les autres effets du changement climatique

L'impact du changement climatique ne se fait pas uniquement remarquer sur la phénologie, mais également sur d'autres aspects des plantes comme la qualité de la croissance. Une étude de grande ampleur a été réalisée sur des peuplements forestiers d'Europe Centrale depuis 1900 jusqu'à nos jours. Si la dynamique de croissance s'est accélérée, et donc le volume de bois, de par la hausse des températures et de la prolongation des saisons, une diminution significative de la densité du bois a été notée. L'étude menée sur 41 forêts d'épicéa, de pin sylvestre, de hêtre ou de chêne sessile montre que la densité du bois a diminué de 8 à 12% en une centaine d'années (Pretzsch et al. 2018).

Chez le noyer, le changement climatique pourrait également venir modifier la pression exercée par les principaux pathogènes. Il a été constaté en France, à la Station d'Expérimentation Nucicole en Rhône-Alpes (SENuRA), que les pluies printanières sont

responsables de la dissémination des conidies de *Colletotrichum acutatum*, ascomycète responsable d'une nouvelle forme d'anthracnose du noyer. Ce pathogène, présent dans les vergers français depuis le début des années 2000 (Belisario et al. 2002), a entraîné l'explosion du nombre de noix noires en 2011 avec une perte de récolte estimée à près de 70% par la SENuRA. Le changement climatique pourrait donc modifier les équilibres au sein des différents écosystèmes. En Californie, second producteur mondial de noix, un modèle de prédiction met en relation l'augmentation des degrés-jour dans des scénarios futurs de réchauffement climatique et l'augmentation du nombre de génération des ravageurs retrouvés dans les vergers californiens, comme le carpocapse, *Cydia pomonella*. L'augmentation de la pression générale des ravageurs dans les années à venir est donc attendue (Luedeling et al. 2010).

D'autre part, les épisodes de sécheresse devraient être plus nombreux et plus longs dans les prochaines décennies. Si les arbres possèdent le moyen de réguler leur transpiration par la fermeture des stomates des feuilles, une longue période de sécheresse peut entraîner un phénomène d'embolie gazeuse, c'est-à-dire la pénétration d'air par cavitation dans les vaisseaux du xylème et bloquant la bonne circulation de la sève. Au sein du genre *Juglans*, certaines espèces apparentées au noyer sont largement utilisées en tant que porte-greffe de par leur plus grande tolérance à certaines maladies telluriques. En parallèle, un porte-greffe devra, à l'avenir, apporter une tolérance à la sécheresse. Ainsi, des travaux ont montré, via l'utilisation de la microtomographie à rayons X, des différences de sensibilité à l'embolie gazeuse au sein des espèces du genre *Juglans*. Ainsi, *J. microcarpa*, originaire des zones arides d'Amérique du Nord, est plus résistant que *J. hindsii*, notamment grâce à des fosses plus petites dans son xylème (Knipfer et al. 2018). Ces travaux ajoutent un nouvel élément à prendre en considération dans la sélection de porte-greffe du noyer.

Plutôt que de se demander si les cultures pourront s'adapter dans le futur à cet environnement changeant, il est aussi pertinent de se poser la question inverse : à savoir, est-ce qu'un environnement aura suffisamment d'aptitudes pour nos cultures. Chez le noyer, un travail récent a été réalisé en Iran, troisième producteur mondial de noix, sur l'adaptabilité des différentes régions de production face au changement climatique à l'horizon 2049. Il a été démontré, après analyse de données de 375 stations concernant les températures, l'humidité et les besoins en froid, que les surfaces adaptées au noyer reculeront de 6,19% pour passer de 582 844 à 546 710 km² (Vahdati et al. 2019).

3. L'importance du phénotypage à l'ère du génotypage haut-débit

Introduction

La détermination de l'architecture génétique d'un caractère quantitatif demande conjointement des données de génotypage et des données de phénotypage. Il est d'une grande importance de bien réfléchir à la méthode de phénotypage puisque l'héritabilité d'un caractère donné peut dépendre de la précision et de la qualité des données phénotypiques (Burghardt et al. 2017). Aujourd'hui, les progrès technologiques liés au phénotypage ne sont pas en reste et le développement des capteurs par exemple constituent une révolution (Gallais 2018).

Les progrès des méthodes de phénotypage et leurs liens avec le génotypage haut-débit

Penser autrement le phénotypage

Afin d'identifier les bases génétiques des caractères multigéniques très complexes, il est peut-être plus bénéfique de s'intéresser aux différents caractères individuels en amont. Imaginons le cas d'un caractère contrôlé par plus de 100 gènes, chacun ayant un degré d'effet identique. Il faudrait alors que le dispositif de GWAS utilisé soit capable d'avoir la puissance statistique nécessaire pour identifier un degré d'effet de moins de 1%. Néanmoins, si un autre caractère héritable expliquant par exemple 20% de la variation du rendement peut se phénotyper, il semble plus judicieux de travailler sur cet autre caractère. Si ce nouveau caractère est contrôlé seulement par une dizaine de gènes, la GWAS aura plus de pouvoir de détection des loci qui finalement contribuent eux aussi au caractère multigénique de départ (Burghardt et al. 2017; Korte and Farlow 2013).

Vers le phénotypage haut-débit

La GWAS a pu voir le jour grâce aux progrès technologiques en matière de génotypage, notamment avec la disponibilité des puces de génotypage SNP comme celles d'Illumina® et Affymetrix® (Bush and Moore 2012). Néanmoins, les progrès de la génomique pour l'amélioration des plantes sont parfois limités par la qualité du phénotypage du caractère étudié ; phénomène d'ailleurs appelé « phenotyping bottleneck » (Furbank and Tester 2011). Ce phénomène, avec le manque de transfert des fruits de la génomique au·à la sélectionneur·euse, constitue un frein supplémentaire aux progrès de l'amélioration des plantes (Stamp and Visser 2012). Un phénotypage performant est synonyme de meilleure précision, non destructif, de haut-débit, tout en réduisant les coûts et le temps via l'automatisation et l'utilisation de capteurs.

L'intérêt pour le phénotypage haut-débit est, sans surprise, grandissant au cours de la dernière décénnie. En 2009, moins de 300 articles publiés traitaient de « high-throughput phenomics » mais en 2016, ce chiffre atteignait les 700 (Mir et al. 2019). Ainsi, des plateformes de phénotypage haut-débit se sont mises en place utilisant des hautes technologies comme la thermographie infrarouge pour enregistrer des températures ou la microscopie à fluorescence pour enregistrer des activités photosynthétiques. En France, il en existe plusieurs : la plateforme PhénoField d'Arvalis, dédiée à l'étude de la tolérance à la sécheresse espèces des grandes cultures, la plateforme de l'UMR « Agroécologie » de l'INRAE Bourgogne-Franche-Comté, possédant une spécificité concernant les intéractions plantes-microorganismes du sol, celle de l'UMR « Laboratoire d'Ecophysiologie des Plantes sous Stress Environnementaux » de l'INRAE Occitanie-Montpellier, centrée sur la modélisation de la réponse adaptative aux stress abiotiques (notamment hydriques), pour ne citer qu'elles.

Cependant, bon nombre des techniques utilisées sont spécifiques des grandes cultures et des particularités propres à l'arboriculture rendent difficile la mise en place de telles plateformes. Chez les espèces fruitières et forestières, les caractères communément mesurés en recherche vont être par exemple liés à la hauteur de l'arbre et le volume du tronc et du houppier, car ils servent de moniteurs globaux pour certains travaux agronomiques tels que l'étude de l'utilisation de l'eau, pour la nutrition ou l'élagage (Torres-Sánchez et al. 2015). Ainsi, les drones ont fait leur apparition (Unmanned Aerial Vehicle en anglais ou UAV) ces dernières années dans le domaine de la recherche agronomique et ils ont su prouver leur efficacité en matière de phénotypage (Luo et al. 2014; Marris 2013).

Le pommier est une espèce modèle pour le phénotypage à haut-débit des arbres fruitiers. Une core collection de 241 accessions en verger a récemment été phénotypée en haut-débit (Coupel-Ledru et al. 2019). Différentes variables originales ont pu être obtenues. Par exemple, le degré de chevauchement des feuilles, déterminant l'efficacité de pénétration de la lumière dans la canopée, a été enregistré via l'utilisation d'un T-LiDAR (Terrestrial Light Detection and Ranging). Cette étude montre d'abord la faisabilité mais surtout la pertinence d'utiliser différentes méthodes de phénotypage haut-débit pour explorer les bases génétiques des caractères quantitatifs des plantes pérennes en verger. Un autre travail montre l'intérêt de ces technologies utilisant les capteurs concernant le feu bactérien du pommier causé par *Erwinia amylovora* (Jarolmasjed et al. 2019). Les auteurs·trices ont testé indépendemment plusieurs caméras multispectrales (caméra RGB pour red, green, blue ;

caméra en proche infrarouge) et un système portatif de mesure de photosynthèse afin d'estimer la progression de l'infection chez de jeunes pommiers. Aussi bien la conductance stomatale mesurée par le système portatif que certaines bandes de spectres ont montré une corrélation avec la sévérité des symptômes.

Des caméras multispectrales, cette fois installées sur drônes, ont également été utilisées pour l'acquisition de données chez le citronnier. En faisant appel à l'intelligence artificielle et au « machine learning », des citronniers en verger ont pu être détectés, comptés, géolocalisés et groupés selon l'envergure de leur canopée (Ampatzidis and Partel 2019). Sur noyer, l'application de telles méthodes n'a pas encore été publiée mais leur potentiel, dans cette ère de la génomique haut-débit, semble évident.

4. La méthodologie de la GWAS

Introduction

L'évolution des méthodes et des outils de l'amélioration des plantes est dépendante de l'évolution des connaissances en biologie et plus particulièrement en génétique. L'amélioration des plantes s'est donc déroulée de manière empirique durant des millénaires en favorisant la récolte des semences des plantes aisées à cultiver et des individus les plus productifs ou les plus résistants (Gallais 2011). Ce n'est qu'en 1694 qu'un certain Rudolf Jakob Camerarius, botaniste et médecin allemand, prouve l'existence d'une sexualité chez les plantes, en réalisant des expériences de castration des étamines et des pistils sur le ricin. Le célèbre naturaliste suédois Carl von Linné en fait le fondement de sa classification, qui donna les bases du système moderne de la nomenclature binomiale. Enfin, la découverte de l'hérédité grâce aux travaux de Gregor Mendel, botaniste tchèque au XIXè siècle et la découverte de l'ADN par le duo James Watson et Francis Crick en 1953 posèrent les bases de la génétique moderne. Aujourd'hui, le métier de sélectionneur·euse fait appel aux biotechnologies pour répondre aux besoins de l'agriculture moderne en créant de nouvelles variétés adaptées aux conditions du sol et du climat, pour répondre aux besoins alimentaires de la population, pour assurer un revenu aux agriculteurs·trices ; et ce, tout en préservant l'environnement.

Quelques définitions

La génétique formelle de Mendel s'est peu à peu diversifiée pour donner naissance à la génétique des populations et à la génétique quantitative. Pendant longtemps, à la redécouverte des lois de Mendel, les caractères quantitatifs (dont la variation est continue et dont la correspondance statistique avec le phénotype se mesure par la notion d'héritabilité), par opposition aux caractères qualitatifs (dont la variation est discontinue et dont la correspondance avec le phénotype est sans équivoque), étaient supposés ne pas suivre ces lois (Gallais 2018). Leur étude plus complexe a permis le développement d'une approche statistique indissociable aujourd'hui de la génétique quantitative.

La génétique des populations vise à expliquer les structures génétiques des populations et les modifications du patrimoine héréditaire de ces populations, par l'étude de la distribution des gènes, sous l'influence des pressions ou forces évolutives, telles que la sélection naturelle, la dérive génétique et les mutations. La génétique quantitative, quant à elle, est la branche de la génétique des caractères dont l'observation passe par une mesure et qui étudie leur

transmission à l'aide de modèles statistiques. Les caractères d'intérêt agronomique sont très souvent quantitatifs (comme par exemple la taille d'un plant de blé, le nombre de grains de son épi ou bien la date de sa floraison) et ces modèles statistiques, à partir de l'estimation des variances dues au génotype et à l'environnement, ont fait émerger le concept d'héritabilité. Sir Ronald Aylmer Fisher, statisticien et biologiste britannique, pose les bases de la génétique quantitative avec son modèle polygénique infinitésimal (Fisher 1918). Selon lui, la variation du caractère est due à l'action combinée d'un grand nombre de gènes à hérédité mendélienne et à l'effet de l'environnement, puis la ressemblance entre individus apparentés résulte du partage d'allèles en commun. Puisque cette variation du caractère résulte d'un grand nombre de facteurs à effet additif, elle suit logiquement une loi Normale.

Découlant de ce modèle, l'héritabilité d'un caractère est spécifique d'une population donnée et étudiée dans un environnement donné, ce n'est donc pas une mesure absolue. Cette héritabilité est susceptible d'évoluer au cours du temps en fonction des changements de l'environnement et des pressions évolutives (sélection, mutation, migration, etc.). Quoi qu'il en soit, une tendance est constatée : plus un caractère est complexe, plus il fait intervenir d'étapes dans le développement de la plante et moins il est héritable (Verrier et al. 2001).

Il est donc légitime de vouloir déterminer un facteur génétique causal ou prédictif d'un caractère quantitatif donné, ou bien d'avoir une idée du déterminisme génétique, de l'architecture génétique de ce caractère, comme le nombre de loci contribuant à la variation de ce caractère et leur contribution (Korte and Farlow 2013). Ainsi, l'apport des marqueurs moléculaires est incontestable dans l'étude de l'hérédité de ces caractères quantitatifs puisqu'ils donnent l'accès à une multitude d'informations concernant le génome. La détection de QTL ou la GWAS, faisant globalement appel aux marqueurs SNP de part leur abondance dans les génomes (Schlötterer 2004), sont encore aujourd'hui les deux principales approches utilisées à cet effet en génétique quantitative.

La naissance de la génétique d'association et son application aux ressources phytogénétiques

Les bases de la génétique d'association ont vu le jour grâce à un article fondateur (Risch and Merinkagas 1996) dans lequel les auteurs·trices expliquent que l'approche par analyse de liaison directe mutation/gène causal est sans succès pour certaines maladies humaines complexes comme le diabète et la schizophrénie. Ces auteurs·trices montrent qu'une approche par analyse d'association est plus puissante d'un point de vue statistique. Le design expérimental original de l'étude d'association pangénomique, ou analyse d'association

tout génome (Genome-Wide Association Study (GWAS) en anglais) a ensuite été créé dans le milieu de la recherche médicale au Japon. En effet, en avril 2000 est née une collaboration entre le SNP Research Center de Tokyo, le Human Genome Center de Tokyo, le Institute of Medical Science de Tokyo, l'Université de Tokyo et le Japan Science and Technology Agency. La mission de cette collaboration était d'identifier plus de 150 000 SNP au sein du génome humain en moins de deux ans (Ikegawa 2012). Peu de temps après, la première séquence du génome humain est publiée en 2001 (Lander et al. 2001) grâce à un travail collaboratif de l'International Human Genome Project (regroupant des laboratoires des Etats-Unis, du Japon, de Chine, de France, du Royaume-Uni et d'Allemagne) et de la société Celera Genomics.

La naissance de la génétique d'association

Avec l'aide de cette séquence malgré tout incomplète, la première base de données de SNP « population-spécifiques » a été établie au Japon (Haga et al. 2002) et la première GWAS fut alors publiée sur l'étude d'un gène de susceptibilité à l'infarctus du myocarde (Ozaki et al. 2002). L'équipe japonaise a mis en évidence un SNP fonctionnel dans le gène de la lymphotoxine-alpha, associé à cette pathologie. Très rapidement, l'achèvement du séquençage du génome humain (99% des zones contenant des gènes avec une précision de 99,99%) est annoncé en 2003 (International Human Genome Sequencing Consortium, 2004) et un nouveau projet voit le jour – le International HapMap Project – qui vise à rechercher les polymorphismes de séquence au sein de nombreux individus de 4 populations représentatives de la diversité génétique humaine (africaine, caucasienne, chinoise et japonaise).

Après l'ère HapMap, une véritable ruée vers l'or de la GWAS s'est mise en place, notamment grâce à la disponibilité de dispositifs de génotypage commerciaux, comme ceux de la société Illumina. Ces dispositifs sont de manière générale créés après identification d'un panel de SNP de référence au sein de quelques individus et ces mêmes SNP sont ensuite disposés sur une puce afin de génotyper un plus grand nombre d'individus pour lesquels on possède des données phénotypiques. De très nombreux articles ont été publiés dans des journaux prestigieux concernant des gènes de susceptibilité aux principales pathologies comme l'obésité, le diabète et l'hypertension. Par exemple, Nature Genetics avec un facteur d'impact de 36,38 en 2010, publiait cette même année 181 articles dont 56% étaient en rapport à la GWAS (Ikegawa 2012).

GWAS versus détection de QTL

La cartographie génétique au sens large utilise le phénomène de recombinaison génétique pour identifier un locus de caractère quantitatif lié à une variation phénotypique. En effet, la carte génétique va alors représenter l'agencement linéaire de loci le long d'un génome, dont les positions relatives ont été déterminées à partir de ces taux de recombinaison. Ces recombinaisons surviennent durant la méiose et plus une région chromosomique est grande, plus la probabilité qu'une recombinaison survienne est importante.

La différence principale entre les deux méthodes réside dans le contrôle que possède le·la chercheur·euse sur le phénomène de recombinaison génétique. D'un côté, le « family or linkage mapping » exploite des individus dont le lien de parenté est connu, via la mise en place d'un croisement biparental contrôlé. La cartographie permet donc un contrôle fin de l'histoire des ségrégations puisque le nombre de générations est connu. De l'autre, le « population mapping » exploite des individus non liés, ou dont le lien de parenté n'est peu ou pas contrôlé et donc, peu ou pas connu (Myles et al. 2009). Dans ce dernier cas, il est plus difficile de savoir où et quand le phénomène de recombinaison génétique s'est produit.

La première carte génétique utilisant les marqueurs moléculaires a été établie chez l'être humain avec l'essor des marqueurs RFLP (Botstein et al. 1980). La détection de QTL via l'utilisation d'une famille en ségrégation a été et reste une méthode puissante pour identifier les zones du génome impliquées dans la variation d'un caractère. Par exemple, les composantes clés de la période de floraison chez *Arabidopsis thaliana* ont été découvertes à partir de détections de QTL chez des familles Recombinant Inbred Line (RIL) et des F_2 (Alonso-Blanco et al. 1998; Clarke et al. 1995; Kowalski et al. 1994). Néanmoins, lors de la mise en œuvre de la cartographie génétique classique issue de l'étude d'un croisement biparental, seule une quantitée limitée de la diversité génétique peut être exploitée (Burghardt et al. 2017). Au contraire, la GWAS fait appel à des individus plus représentatifs de l'échelle d'une population naturelle et tire parti d'un historique d'évènements de recombinaison génétique plus important, tenant compte de l'histoire évolutive de l'espèce (Zhu et al. 2009).

Le bénéfice de la GWAS appliqué aux ressources phytogénétiques a été d'abord observé au sein des espèces autogames. La première étude d'association a notamment été réalisée sur le maïs en 2008, à l'aide de 8590 SNP et 553 lignées élites, dans le but de déterminer les gènes impliqués dans le contenu en acides gras du grain (Beló et al. 2008). Mais depuis le séquençage du génome de référence du maïs B73 en 2009 (Schnable et al. 2009), de nombreuses GWAS ont été réalisées, passant de 5 articles publiés en 2009 à 119 en 2015 (Xiao et al. 2017).

L'application de la GWAS aux espèces fruitières et forestières

Un des avantages de l'application de la GWAS aux espèces fruitières et forestières, donc pérennes avec des temps de génération très longs, réside également dans sa plus grande facilité de mise en place et son coût réduit, en comparaison au développement d'une famille en ségrégation (Zhu et al. 2009). Chez l'abricot, 2 SNP impliqués dans la résistance partielle au chancre bactérien ont été trouvés à l'aide d'un panel de 73 accessions d'une core collection, phénotypée durant 4 ans et génotypée à l'aide d'un peu plus de 63 000 SNP (Omrani et al. 2019). Chez la pêche, ce sont 3 SNP couvrant une région de 43 kb environ en lien avec la maturité du fruit qui ont été trouvés et qui pourront servir à mettre en place une sélection assistée par marqueurs en faveur d'une maturité précoce (Elsadr et al. 2019). Chez la pomme, l'utilisation de la puce de génotypage à haute densité de SNP Axiom® Apple 480K sur un panel de plus de 1000 accessions a mis notamment en évidence 2 SNP associés à la date de floraison et le fait que la GWAS peut permettre de réduire l'intervalle de confiance des régions associées identifiées au préalable par la détection de QTL (Urrestarazu et al. 2017).

Finalement, chez le noyer, la première publication de GWAS utilisant la puce Axiom® *J. regia* 700K est très récente et fait apparaître 55 SNP associés à des caractères concernant la noix en coque et le cerneau, comme l'épaisseur de la coque, le poids de la noix en coque et le taux de remplissage (Arab et al. 2019). Néanmoins, ces premiers résultats méritent d'être approfondis puisque le niveau de diversité des accessions sélectionnées est limité à quelques populations naturelles uniquement iraniennes et le phénotypage de ces accessions n'a été réalisé qu'une seule année, donc l'effet de l'environnement n'a pas pû être apprécié.

De la génétique d'association à la sélection assistée par marqueurs

L'approche pour conduire une GWAS peut s'envisager comme selon la figure 6 d'après la synthèse de Burghardt et al. 2017. L'étape (1) consiste en la sélection d'un panel de génotypes divers et celui-ci peut représenter une collection de génotypes à l'échelle de l'aire de répartition représentant le maximum de diversité présente à l'état naturel, ou bien une collection de génotypes relatifs à un programme d'amélioration. L'étape (2) se découpe en deux parties : l'obtention des données génotypiques et l'obtention des données phénotypiques pour le ou les caractère(s) héritable(s) souhaité(s). Plus le phénotypage est précis et plus le génotypage est dense, plus les chances de réussite de la GWAS sont élevées. L'étape (3) réside dans les analyses statistiques visant à relier, associer le polymorphisme nucléotidique à

la variabilité phénotypique, via l'utilisation du modèle linéaire. L'étape (4) donne la priorité aux associations candidates pour la poursuite des travaux. Il est donc conseillé, entre autres, d'étudier la force de l'association statistique, de rechercher des données cartographiques complémentaires et d'explorer l'éventuelle présence de gènes dans la région. Enfin, l'étape (5) consiste en la validation des associations candidates en utilisant, par exemple, un panel indépendant ou une méthodologie de génétique inverse.

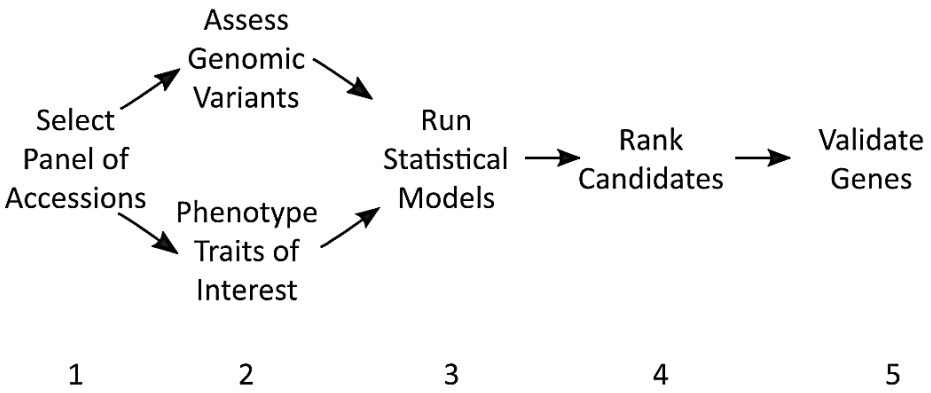


Figure 6 : Approche de base pour la réalisation d'une GWAS.
D'après Burghardt et al. 2017.

Un protocole unique ?

Evidemment, il n'existe pas de protocole unique en matière de conduite de GWAS. Néanmoins, certaines étapes de l'approche décrite précédemment ont fait l'objet de « procédures » globalement utilisées et admises dans la communauté scientifique. Par exemple, l'étape (2) qui consiste à obtenir les données de génotypage est suivie par une étape incontournable, le contrôle qualité des SNP. Des procédures ont été décrites et peuvent constituer une aide pour le·la chercheur·euse (Anderson et al. 2010; Turner et al. 2011; Marees et al. 2017), en utilisant des programmes comme PLINK (Purcell et al. 2007). Ce contrôle qualité se traduit entre autres par l'identification des données de génotypage manquantes et l'identification des individus dupliqués.

L'importance du déséquilibre de liaison

Si la densité du génotypage est naturellement primordiale, le nombre de marqueurs disponibles n'est pas le seul critère à considérer. En effet, il n'est pas assuré que les SNP génotypés soient toujours ceux qui sont responsables de la variation phénotypique. Un paramètre plus important est donc le degré du déséquilibre de liaison (DL), ou association allélique (linkage disequilibrium en anglais), entre un polymorphisme nucléotidique quelconque et un autre dit « causal » du phénotype, comme expliqué en figure 7.

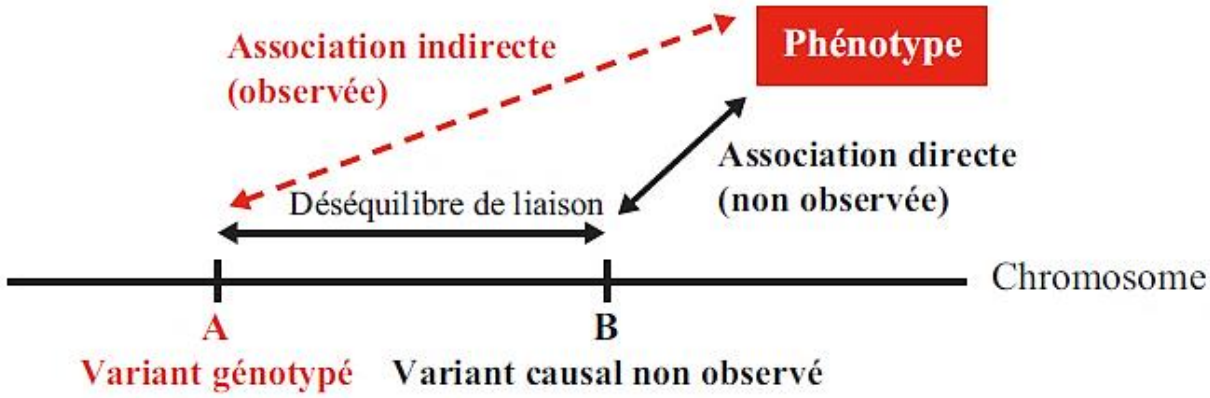


Figure 7 : Représentation schématique du déséquilibre de liaison.
D'après Astle and Balding 2009, traduit par Debette 2012. Ici, A et B représentent la distinction entre association génétique et relation causale. Le polymorphisme variant génotypé peut parfois être le variant causal.

Le DL se définit comme l'association non-aléatoire de polymorphismes nucléotidiques à différents loci (Clarke et al. 2011), ce qui correspond à un écart à l'équilibre d'Hardy-Weinberg provoqué notamment par les forces évolutives (Flint-Garcia et al. 2003). Ce déséquilibre de liaison est spécifique d'un panel et vient donc définir la densité de génotypage nécessaire. L'espoir d'un·e chercheur·euse réside alors dans le fait que les SNP génotypés soient en fort DL avec le SNP causal pour que l'association significative soit elle-même détectable, comme représenté en figure 9. En plus du panel, le DL dépend également du nombre de chromosomes et du nombre de générations impliquées dans la création du panel. Par exemple chez l'être humain, la population africaine est la plus ancestrale et possède les régions de DL les plus petites, puisqu'elle a accumulé plus d'évènements de recombinaison, en comparaison avec les populations européennes ou asiatiques qui possèdent donc un plus grand niveau de DL (Bush and Moore 2012).

Sur cette figure 8, le SNP causal en gris n'est pas génotypé mais deux loci situés de part et d'autre sont eux génotypés. Le SNP à sa droite est en fort DL avec le causal, contrairement à celui à sa gauche. Finalement, une association est trouvée entre l'allèle C à ce SNP en fort DL et le nombre plus important de baies, le caractère étudié, puisque généralement, la force de correlation entre deux marqueurs est fonction de la distance qui les sépare : le DL est plus fort si les marqueurs sont proches (Myles et al. 2009). La densité du génotypage nécessaire et la résolution de la zone du génome détectée vont donc dépendre de la taille du génome et du degré du déséquilibre de liaison, tout deux étant dépendants de l'espèce. Ce degré de déséquilibre de liaison s'apprécie en fonction de comment diminue le DL selon la distance. Ainsi, l'autogamie a tendance à réduire les évènements de

recombinaison génétique, ce qui augmente le DL, tandis que l'allogamie entraîne généralement un DL qui diminue plus rapidement (Garris et al. 2005; Heuertz et al. 2006). Le niveau de consanguinité des accessions choisies vient aussi faire varier le DL au sein d'une même espèce et parfois, si le DL est très étendu, la génétique d'association n'apportera pas plus d'avantages que la détection de QTL, puisqu'il sera difficile de déterminer un gène candidat. Il est donc primordial de sélectionner un panel d'individus diversifiés et exploitant le maximum d'évènements de recombinaison génétique qui se sont produits dans l'histoire de l'espèce étudiée.

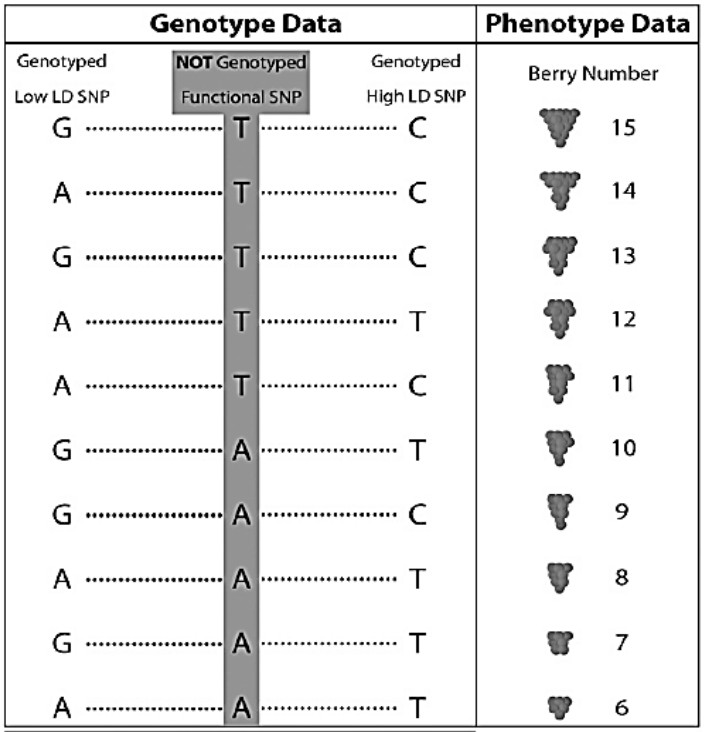


Figure 8 : Représentation fictive d'une association.
D'après Myles et al. 2009.

Concernant la significativité de l'association, la p-value, elle est liée à la probabilité de faire une erreur de type 1 et donc de rejeter l'hypothèse nulle H0 si elle est vraie (soit d'obtenir une association par chance). Elle est généralement présentée sous la forme d'un diagramme de Manhattan (Manhattan plot en anglais), une forme particulière de nuage de points. Plus précisément, ce Manhattan plot représente le logarithme décimal négatif de la p-

value sur l'axe des ordonnées ($-\log_{10}$) et la position génomique du SNP sur l'axe des abscisses. Puisque les associations les plus « fortes » possèdent les plus petites p-values, leur $-\log_{10}$ sont donc les plus grands. En biologie, il est classique d'utiliser une valeur de p-value de 0,05 comme valeur critique pour rejeter l'hypothèse nulle mais dans le cas d'un scan de génome avec 1 million de SNP, 5% donne 50 000 SNP susceptibles d'être trouvés comme significatifs. Ainsi en GWAS, la correction de Bonferroni (0,05 divisé par le nombre de SNP) est largement utilisée.

Les limites attribuées à la GWAS

La GWAS surmonte les principales limitations de la détection de QTL mais présente également plusieurs inconvénients (Korte and Farlow 2013). La puissance qu'aura la GWAS à déterminer les zones du génome liées au caractère donné est fonction de 6 facteurs communs avec la détection de QTL : la complexité de l'architecture génétique de ce caractère, son héritabilité, le nombre d'accessions analysées, la densité du génotypage, la précision du phénotypage et le niveau de proximité génétique. En GWAS, un facteur vient s'ajouter : la structure de la population (Burghardt et al. 2017).

Ainsi, la probabilité de réussite d'une conduite de GWAS dépend du nombre et du degré d'effet, ou taille d'effet des gènes impliqués dans la variation du caractère (Korte and Farlow 2013). Un caractère multigénique très complexe ne donnera pas, d'une manière générale, de signal fort en GWAS comme dans le cas de l'étude de l'architecture génétique de la hauteur du plant de maïs (Peiffer et al. 2014), un caractère pourtant déterminé comme étant hautement héritable. D'autres facteurs augmentent les risques d'erreur et deux d'entre eux, souvent reliés, sont la structure de la population et le niveau de proximité génétique entre les accessions, qui peuvent mener à l'obtention de faux-positifs (erreurs de type 1) (Aranzana et al. 2005; Clarke et al. 2011). Lorsqu'un·e chercheur·euse travaille avec des RPG, il·elle n'est pas dans les conditions classiques de génétique des populations et sa « population artificielle » en question a certainement subi des pressions de sélection artificielle. Ces pressions, qui peuvent être la conséquence de croisements contrôlés pour l'obtention de variétés modernes, ont donc entraîné des phénomènes de structure et de proximité génétique (aussi appelée apparentement) dans les populations étudiées (Flint-Garcia et al. 2005). Cela se rencontre également dans le cas d'utilisation de populations « naturelles » qui peuvent aussi être soumises à d'autres pressions de type migration, mutation ou sélection naturelle, entrainant une structuration génétique.

Si le caractère étudié possède une héritabilité non nulle, il existe donc des zones du génome qui sous-tendent sa variation. Néanmoins, il est relativement courant en GWAS de ne pas obtenir d'association avec une p-value significative. Ceci est notamment courant en appliquant la correction de Bonferroni qui fait l'hypothèse d'indépendance des tests de chaque SNP ; or il existe un certain niveau de DL entre ces SNP. Ainsi, se pose la question de l'interprétation stricte de cette p-value comme seuil de significativité et celle-ci pourrait plutôt servir de « guide » des associations à étudier plus précisément (Burghardt et al. 2017). En parallèle des faux-positifs, la GWAS peut donc mener à l'obtention de faux-négatifs. L'héritabilité manquante constitue justement une autre limite de la GWAS associée à l'obtention de faux-négatifs. Ce phénomène est observé lorsque, pour un caractère pourtant connu pour être hautement héritable, les SNP associés ne comptent que pour un très faible pourcentage de l'héritabilité. Une explication se trouve du côté des allèles « causals » de faible fréquence. Si un allèle rare possède un fort effet sur le caractère étudié, la génétique d'association aura une faible puissance pour le détecter (Myles et al. 2009). Aussi, cette héritabilité manquante peut être due aux phénomènes d'hétérogénéité génétique et d'épistasie (Brachi et al. 2011). Enfin, des causes sont peut-être à chercher du côté de l'épigénétique.

Les différentes approches et de nouvelles méthodes

Classiquement, les analyses statistiques en GWAS utilisent le modèle linéaire (Burghardt et al. 2017) où le phénotype est la variable à expliquer, le polymorphisme nucléotidique est la variable explicative et dans lesquels plusieurs covariables peuvent être incluses pour réduire certains effets :

$$\mathbf{Y} = \mu + \mathbf{X}\beta + \mathbf{G}\gamma + \mathbf{R}$$

- **Y** : phénotype
- μ : moyenne générale
- **X**β : covariables
- **G**γ : effet du SNP testé
- **R** : résidus

Afin de se faire une idée de la fiabilité des résultats, il est possible de regarder le diagramme quantile-quantile (Q-Q plot en anglais) des $-\log_{10}$ des p-values observées par rapport aux $-\log_{10}$ des p-values attendues, comme représenté en figure 9. Cette méthode de lecture graphique est importante à réaliser car les éventuels écarts peuvent venir modifier les p-values.

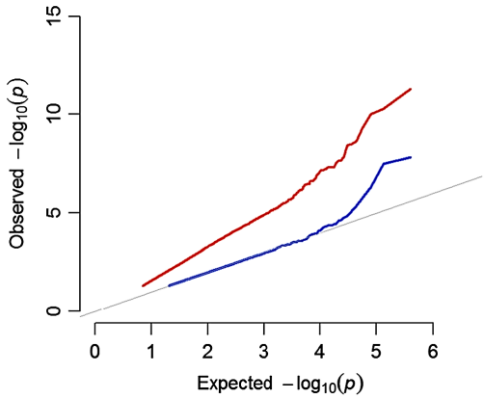


Figure 9 : Exemple fictif de diagramme quantile-quantile montrant la fiabilité des résultats de GWAS.
D'après Korte and Farlow 2013. Le Q-Q plot illustre la relation entre les tests statistiques observés (en ordonnée) et attendus (en abscisse), selon une loi uniforme. Ici, en rouge, les p-values observées sont déviées de façon positive et suggèrent un modèle non adéquat, par exemple ne prenant pas en compte les « confounding effects ». A l'inverse, en bleu, le modèle suit globalement la ligne y = x et seuls les points extrêmes suggèrent une association.

Si tel est le cas, il est alors possible de transformer les données de phénotypage afin que les résidus (différences entre les valeurs observées et les valeurs estimées, ils représentent la partie non expliquée par le modèle) possèdent les propriétés d'une distribution normale (Goh and Yap 2009). Une autre solution réside dans l'utilisation d'un modèle linéaire « amélioré », comme le modèle linéaire généralisé (generalized linear model en anglais ou GLM) permettant, entre autres, d'utiliser des données de phénotypage qui ne respectent pas scrupuleusement la loi normale.

La structure de la population, constituant une des sources principales de facteur d'erreur, a été rapidement intégrée dans ce GLM. Il est donc possible de corriger le modèle en considérant l'effet du génotype comme aléatoire et en intégrant la structure de la population comme covariable. Le paramètre K, qui estime la proportion d'appartenance de chaque accession à une population « ancestrale », peut être déterminé à l'aide de programmes dédiés comme STRUCTURE (Pritchard et al. 2000) ou fastSTRUCTURE (Raj et al. 2014) pour des données avec de nombreux SNP. A l'issue, une « Q-matrix » est obtenue et intégrée en covariable à effet fixe. Cette Q-matrix peut s'obtenir également en réalisant une analyse en composantes principales (ACP) (Price et al. 2006; Tian et al. 2008). Lorsque la proximité génétique est due à des relations complexes faisant intervenir différents niveaux de pedigree, elle n'est pas forcément prise en compte dans les modèles utilisés par STRUCTURE ou en ACP. Il existe alors une approche complémentaire visant à estimer directement la parenté entre tous les individus deux à deux, constituant une « K-matrix » (pour kinship, parenté en anglais). La structure de la population et la proximité génétique entre les accessions ont donc

été intégrées simultanément dans un nouveau type de modèle linéaire, le modèle linéaire mixte (mixed linear model en anglais ou MLM) (Yu et al. 2006; Zhang et al. 2010). Cette K-matrix est alors intégrée en covariable à effet aléatoire et cette nouvelle méthode combinant Q+K s'avère intéressante pour diminuer encore un peu plus le risque d'obtention de faux-positifs (Myles et al. 2009). Un nouveau modèle mixte a ensuite vu le jour, le « Efficient Mixed-Model Association » ou EMMA (Kang et al. 2008). Il permet d'augmenter nettement la vitesse de calcul, qui pouvait être longue en utilisant la combinaison Q+K avec le modèle plus classique de MLM de Yu et al. 2006.

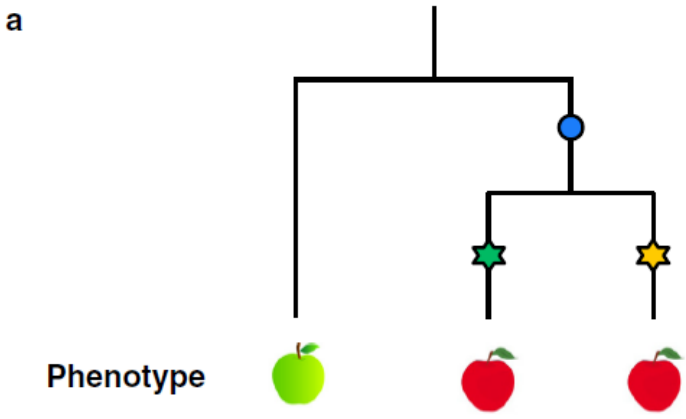
Cette K-matrix peut se calculer via les données SNP, en déterminant les portions du génome qui sont identiques entre deux accessions possédant un lien de parenté ; c'est l'identité par ascendance (identity by descent en anglais ou IBD) (Powell et al. 2010). Une des méthodes disponibles aujourd'hui est la méthode « Kinship-based INference for Genome-wide association studies » ou KING qui permet l'estimation des liens de parenté au sein d'une population structurée (Manichaikul et al. 2010). Une autre méthode permet également d'estimer ces liens de parenté au sein d'une population structurée avec une ascendance mélangée (admixed ancestry en anglais) ; c'est la méthode « Relatedness Estimation in Admixed Populations » ou REAP (Thornton et al. 2012).

Une version plus rapide de EMMA, le « EMMA eXpedited » ou EMMAX, permet de réduire encore plus le temps de calcul (Kang et al. 2010). Ce nouvel algorithme repose sur l'hypothèse que chaque marqueur ne contribue qu'à une faible proportion de la variance génétique totale. Ainsi, la prise en compte de la structure de la population peut être approximée par le biais d'un calcul d'estimations d'un modèle plus réduit, sans ajuster chaque marqueur, en utilisant le critère d'information bayésien (Huang et al. 2018). Plus tard, ce modèle s'est vu encore être amélioré pour devenir le « Genome-Wide Efficient Mixed-Model Association » ou GEMMA (Zhou et al. 2013).

Dans cette course de l'amélioration du modèle linéaire mixte appliqué à la GWAS, la création du « Compressed MLM » ou CMLM constitue une étape importante puisque dans ce modèle, les individus sont groupés selon la « K-matrix », réduisant ainsi encore un peu plus le temps de calcul, particulièrement dans les dispositifs comprenant beaucoup d'accessions (Zhang et al. 2010). Plus tard, le « Factored Spectrally Transformed Linear Mixed Model Select » ou « FaST-LMM-Select » (Lippert et al. 2011; Listgarten et al. 2013), utilise une « K-matrix simplifiée », créée à partir d'un nombre plus restreint de marqueurs.

Toutes ces améliorations du MLM avaient pour objectif principal d'améliorer la vitesse de calcul, mais pas toujours d'améliorer la puissance statistique du modèle. De plus, si ces corrections diminuent le risque d'obtention de faux-positifs, elles peuvent néanmoins conduire à l'obtention de faux-négatifs, en particulier si les effets contribuent de manière importante à la variation du phénotype et si le SNP causal est masqué par cette correction.

Une manière de tenter d'augmenter l'efficacité de la GWAS est d'augmenter le nombre d'accessions et le niveau de diversité, en se basant notamment sur l'origine géographique de ces accessions. Néanmoins, il risque de se passer un phénomène d'introduction d'hétérogénéité génétique désignant le fait que plusieurs polymorphismes nucléotidiques peuvent causer un même phénotype (Korte and Farlow 2013). En effet, pour plusieurs raisons, comme l'adaptation locale, différents SNP peuvent être associés à un même caractère au sein d'accessions qui proviennent de différentes régions (Fournier-Level et al. 2011). Ce phénomène vient alors affaiblir l'association entre le SNP causal et le caractère mais il peut aussi faire en sorte qu'un SNP non causal devienne un « meilleur descripteur » du caractère que le causal (Platt et al. 2010), soit un faux-positif, comme expliqué en figure 10.



b

individual	phenotype	blue	green	orange
#1		0	0	0
#2		1	1	0
#3		1	0	1

Figure 10 : Représentation fictive d'une association induite par l'hétérogénéité génétique.
D'après Korte and Farlow 2013. La figure (a) représente un arbre phylogénétique fictif de trois individus, dans lequel trois mutations ont survenu. Les deux mutations les plus récentes (étoiles) causent un changement de phénotype (pomme rouge). La figure (b) montre que la mutation plus ancienne (rond bleu) n'affecte pas le phénotype « pomme rouge » mais est en parfaite corrélation avec ce caractère. Ainsi, ni l'une ni l'autre des mutations causales ne sont de bons prédicteurs du phénotype « pomme rouge ».

Toutes ces précédentes méthodes sont basées sur le test d'un seul locus à la fois mais pour les caractères complexes contrôlés par de nombreux loci, ces modèles demeurent insuffisants. En effet, le modèle infinitésimal devient obsolète dans le cas de caractères contrôlés par plusieurs loci à effet fort. Heureusement, il est maintenant possible de s'affranchir de ce problème en utilisant le « Multi-Locus Mixed-Model » ou MLMM (Segura et al. 2012) qui inclue notamment les polymorphismes en « compétition » en covariable de ce nouveau modèle linéaire mixte. En incluant plusieurs marqueurs dans un même modèle, il devient possible d'estimer les effets additifs de plusieurs loci (Chen et al. 2019). Le MLMM ajuste alors plusieurs marqueurs associés comme effet fixe dans le modèle pour fractionner la variance expliquée par la proximité génétique en une régression par étapes (Huang et al. 2018).

Un cas de figure supplémenaire peut subvenir lorsque le caractère étudié est corrélé à d'autres, à cause d'un effet pléiotropique par exemple. Le « Multi-Trait Mixed-Model » ou MTMM (Korte et al. 2012) permet notamment de prendre en compte cette intéraction. Aussi, l'interaction entre certains allèles à deux ou plus de deux loci, l'épistasie, peut être prise en compte dans un nouveau modèle nommé Stepwise Procedure for constructing an Additive and Epistatic Multi-Locus model ou SPAEML (Chen et al. 2019). Mais récemment, le « Fixed And Random Model Circulating Probability Unification » ou FarmCPU (Liu et al. 2016) semble être la méthode statistique la plus aboutie dans l'univers de la GWAS pour prendre en compte tous les « confounding effects ». En fait, des pseudos « Quantitative Trait Nucleotides » ou pseudo QTN sont intégrés dans un modèle mixte à effets fixes et sont choisis parmi les marqueurs potentiellement associés pour être évalués par un modèle à effet aléatoire prenant en compte la « K-matrix » définie pour ces pseudos QTN. FarmCPU combine donc l'utilisation itérative de covariables à effets fixes et aléatoires, gagnant ainsi en performance pour éviter au maximum l'obtention de faux-positifs. Le créateur de FarmCPU, en détaillant le concept de son modèle, fait un état de l'art plus visuel de l'évolution des grands modèles de GWAS (Figure 11).

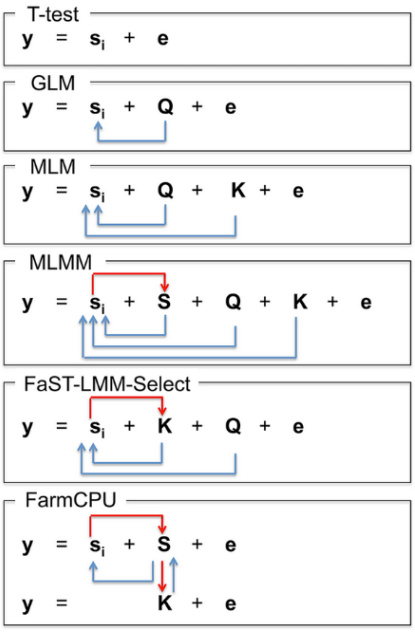


Figure 11 : Concept de développement de FarmCPU.
D'après Liu et al. 2016. Le modèle naïf (T-test) teste l'effet d'un marqueur à la fois (s_i) sur le phénotype (y) avec les résidus (e). Le modèle GLM contrôle les faux-positifs en incluant la structure (Q) comme covariable, ajustant le test sur les marqueurs (flèche bleue). Le modèle MLM inclue Q et la parenté (K) comme covariables mais ne sont pas elles-mêmes ajustées par les résultats des tests d'association. Le modèle MLMM ajoute les pseudos QTN (S) comme covariables qui sont estimés selon une procédure de régression par étapes. Ils sont donc ajustés par les résultats des tests d'association (flèche rouge), mais les covariables Q et K ne sont pas ajustées. Le modèle FaST-LMM est similaire au modèle MLM mais la covariable K est ajustée. FarmCPU innove en combinant MLMM et FaST-LMM, puis en utilisant alternativement deux modèles, à effet fixe et à effet aléatoire : le premier modèle à effet fixe prend en compte l'effet du marqueur et les pseudos QTN ; ces pseudos QTN sélectionnés au sein des marqueurs associés sont ensuite évalués dans un second modèle à effet aléatoire dans lequel K est définie par ces pseudos QTN.

Un des moyens d'avoir plus de certitudes quant aux résultats obtenus reste de tester plusieurs modèles. Il est commun d'observer dans les articles, la comparaison des différents modèles disponibles comme chez l'abricot, notamment via l'utilisation du Q-Q plot (Mariette et al. 2016). Aujourd'hui par exemple, un des packages R dédié à la GWAS, GAPIT, possède une version 3 permettant une certaine interactivité dans la comparaison des résultats donnés par les différents modèles qu'il est possible d'utiliser (Wang and Zhang 2019). Il est néanmoins communément admis que les futurs modèles développés et utilisés doivent être des modèles linéaires mixtes, notamment des modèles « multi-locus », pour limiter l'obtention de faux-positifs (Feng et al. 2016). Il est dit également que lorsque des QTN sont identifiés par plusieurs modèles « multi-locus » à partir d'un même jeu de données, ceux-ci sont généralement fiables (Zhang et al. 2019). Par exemple, une étude du déterminisme génétique des propriétés de collage de l'amidon chez le maïs a montré la supériorité des modèles « multi-locus » comme FarmCPU en comparaison au modèle « simple-locus » GEMMA pour déterminer des QTN (Xu et al. 2018).

Combiner les approches pour une meilleure puissance

Si les deux principales méthodes en génétique quantitative se résument à leur niveau de phénomènes de recombinaison, il est possible de dire que la détection de QTL est une analyse de la liaison génétique au sein d'individus reliés (linkage analysis ou LA), tandis que la GWAS passe par une analyse du déséquilibre de liaison au sein d'individus peu ou non reliés (linkage disequilibrium mapping ou LD). Cette ressemblance a automatiquement entraîné une certaine convergence des deux approches (Charmet 2011) : ainsi, la détection de QTL a été étendue à plusieurs populations et la GWAS a fait rapidement l'objet de corrections pour prendre en compte la parenté entre les individus d'un panel bien souvent construit à partir d'une collection de RPG. La GWAS et la détection de QTL, conduites en parallèle, peuvent être donc complémentaires et peuvent atténuer les limitations de chacunes. Cette approche porte parfois le nom de LD-LA mapping (Meuwissen et al. 2002; Pikkuhookana and Sillanpää 2014). Une des approches les plus connues est l'approche NAM pour Nested Association Mapping (Yu et al. 2008). Dans cette approche, il est question de choisir un panel très diversifié de parents dans un germplasm qui seront génotypés de manière dense et ensuite de générer des descendances en croisant ces parents avec un seul individu pivot. Au final, les lignées pures recombinantes obtenues seront génotypées de manière moins dense. L'atout majeur de cette approche réside dans le fait que les recombinaisons observées dans les descendances vont venir faire disparaître certains faux-positifs (Charmet 2011). Néanmoins, cette approche est principalement appliquée aux grandes cultures, avec un temps de génération plus court.

Sans aller jusqu'à la création de matériel spécifique comme dans le cas des NAM, le LD-LA mapping peut se faire sur une famille et un panel d'association, de façon conjointe. Chez *A. thaliana*, cette approche appliquée sur près de 20 000 plantes avec plus de 200 000 SNP, a permis d'augmenter la capacité de distinguer les associations vraies des fausses. La détection de QTL a permis également d'identifier les faux-négatifs qui sont les associations « perdues » lors de l'application des méthodes classiques de correction pour la structure de la population (Brachi et al. 2010). Similairement, un autre travail a encore démontré la présence de faux-négatifs « perdus » après correction de la structure et du niveau de proximité génétique des accessions en combinant les deux méthodes, toujours chez *A. thaliana* (Zhao et al. 2007).

Similaire à l'approche NAM dans son concept, l'approche MAGIC pour Multi-parent Advanced Generation Inter-Cross (Cavanagh et al. 2008) consiste à choisir un panel diversifié

dans un germplasm (comme les NAM) et de croiser plusieurs parents, puis de croisier entre eux les individus issus de ces premiers croisements, ce sont les inter-croisements. Les lignées pures recombinantes obtenues au final peuvent permettre d'augmenter la résolution de détection de QTL, comme cela a été le cas chez le riz (Bandillo et al. 2013). Mais encore une fois, cette approche s'applique plus facilement aux grandes cultures.

Le potentiel de la transcriptome-wide association study

Puisque la GWAS identifie des loci souvent en dehors des régions codantes et que la variation de leur régulation joue un rôle important dans le caractère, l'intégration des données de l'expression des gènes a été proposée récemment (Yeung et al. 2019). La transcriptome-wide association study (TWAS) est une étude intégrée de GWAS et de données d'expression. Elle repose sur « l'entraînement » d'un modèle prédictif d'expression à partir des données de génotypage et d'expression. Ce modèle prédictif sert alors à prédire l'expression des gènes en lien avec les SNP proches d'un gène, chez les individus du panel de GWAS. Au final, l'expression des gènes prédite est une covariable supplémentaire du modèle linéaire mixte. L'association entre l'expression ainsi prédite et le caractère est alors recherchée ; ce sont les expression QTLs en anglais ou eQTLs (Wainberg et al. 2019). Un nouveau modèle a d'ailleurs vu le jour, le « Collaborative Mixed Model » ou CoMM (Yeung et al. 2019), pour contrebalancer la perte de puissance statistique due à cette méthode en deux étapes (expression prédite, puis association avec le caractère).

La validation des résultats

Le passage des résultats de GWAS à la détermination du sens ou de la fonction biologique passe notamment par la recherche de gènes candidats, étape essentielle de la « post-GWAS ». Une fois réalisée la priorisation des associations candidates, il est possible de regarder du côté des annotations du génome (Gallagher and Chen-Plotkin 2018). Classiquement, il est intéressant de regarder toutes les séquences codantes autour des SNP associés, dans une fenêtre possédant un niveau de DL suffisant. Il existe maintenant des outils permettant d'effectuer une interprétation plus avancée des résultats de GWAS et l'un d'entre eux, « Postgwas » (Hiersche et al. 2013), permet par exemple de créer des Manhattan plots plus précis au niveau de la région d'association, ainsi que des analyses de réseaux d'intéractions biologiques en utilisant les annotations du génome.

Il est également nécessaire de répliquer l'expérience dans d'autres populations. La réplication consiste à vérifer une association trouvée durant une première étude au sein d'un échantillon indépendant (Clarke 2011). Ceci est particulièrement vrai chez l'espèce humaine

et ses différents groupes ethniques, mais aussi chez les plantes. La répétition des analyses dans des grands échantillons indépendants, de préférence très divers et sans lien de parenté, est une des façons d'identifier des associations robustes et valables dans un contexte plus général. Il est en effet peu probable qu'un faux-positif le soit réellement dans des échantillons indépendants (Burghardt 2017). Ainsi, la collaboration internationale devient souvent nécessaire pour l'échange de matériel végétal ou de données et l'identification de SNP associés n'est que « the end of the beginning ! » (Ikegawa 2012).

Aujourd'hui, il est de plus en plus aisé de valider la fonction biologique d'un gène candidat, notamment par des approches de génétique inverse ou d'édition du génome. La plus récente et célèbre d'entre elles est la méthode CRISPR/Cas9 qui permet de cibler une paire de bases spécifique. Mais même s'il est évident qu'elle peut ouvrir la porte au domaine de la génomique fonctionnelle, son application au sein des espèces pérennes avec une longue période juvénile est encore sous-représentée (Liu et al. 2016).

Application des « fruits » de la recherche

La sélection assistée par marqueurs est particulièrement utile pour les caractères ne pouvant être mesurés que très tardivement, comme tous ceux liés aux fleurs et aux fruits pour les espèces à longue période juvénile, telle que le noyer (Meuwissen 2003). Sur la base du génotype aux marqueurs, seuls les hybrides ayant les allèles favorables seront plantés. Cette sélection pouvant avoir lieu au stade plantule, il en résulte une optimisation importante du dispositif expérimental. En parallèle ou bien en complément de la GWAS, la sélection génomique nécessite aussi un génotypage à haute-densité et les deux méthodes peuvent être utilisées conjointement pour venir rentre plus performante la sélection assistée par marqueurs (Minamikawa et al. 2018). A l'origine créée dans le monde de l'amélioration du bétail et relativement récente chez les plantes, la sélection génomique est une méthode qui utilise tous les SNP disponibles lors du génotypage à haute-densité (et non pas seulement ceux utilisés en SAM classique) pour prédire les valeurs génétiques (genomic estimated breeding values en anglais ou GEBV) au sein des descendants (Lorenz et al. 2011). Schématiquement, les GEBV sont calculées sur une population « formatrice » à l'aide des données génotypiques, mais aussi phénotypiques, via l'utilisation de modèles linéaires similaires à ceux utilisés en GWAS (Iwata et al. 2016). Ces informations sont ensuite utilisées pour sélectionner des individus candidats uniquement sur leurs données génotypiques dans le but d'additionner les effets génétiques des caractères complexes (Jonas and de Koning 2013). Il a d'ailleurs été avancé le fait que la sélection génomique permettrait de gagner du temps et de l'argent durant les

premiers processus de création variétale, tout en donnant des résultats plus fiables que la SAM classique (Heffnet et al. 2011; Cabrera-Bosquet et al. 2012), notamment chez les espèces arboricoles comme *Pinus taeda*, une des espèces les plus économiquement importantes en Amérique du Nord (Resende et al. 2012). L'efficacité de la sélection génomique a aussi été observée chez l'eucalyptus (Denis and Bouvet 2013), la vigne (Fodor et al. 2014) et le palmier à huile (Wong and Bernardo 2008).

Néanmoins, une des principales difficultés de l'application de la sélection génomique chez les plantes est la part importante de l'environnement dans le contrôle du caractère. Là où l'environnement est plus facilement contrôlé chez le bétail, il est important en agronomie de prendre en compte l'interaction génotype-environnement (G×E). Le succès de la sélection génomique dépend donc de la précision des prédictions, elle-même affectée par de nombreux phénomènes en interaction, que l'on retrouve en GWAS, comme le degré de DL ou la structure de la population (Marulanda et al. 2015). Quoi qu'il en soit, la sélection basée sur l'ADN en général (DNA-informed breeding en anglais) est en passe de devenir une norme. Chez les *Rosaceae* par exemple, le développement des ressources génomiques a quelque peu détourné l'investissement pour la sélection plus traditionnelle comme l'expérimentation de matériel « élite » et l'évaluation phénotypique qui demeuraient les bases pour le développement de nouveaux cultivars (Peace 2017). Mais le transfert des résultats prodigués par la recherche en génétique quantitative à l'application directe en sélection est une étape longue et critique. Finalement, durant cette dernière décennie, les marqueurs moléculaires ont été utilisés en SAM principalement pour des caractères à déterminisme génétique simple tel que le caractère plat chez la pêche (Lambert et al. 2009), mais encore assez peu pour des caractères complexes. Un certain déséquilibre est observé entre les articles d'études d'association publiés et ceux concernant l'application en SAM (Jonas and de Koning 2013; Bliss 2010). Toutefois, compte tenu du fait que le cout du génotypage diminue et que celui de l'expérimentation ne cesse d'augmenter, nous pouvons nous attendre à une augmentation de l'utilisation de la SAM dans les années à venir.

La génétique non mendélienne

Les caractères quantitatifs héritables suivent donc des lois de transmission précises mais il existe une part de l'hérédité qui ne suit pas ces lois. Certains caractères sont notamment sous la dépendance du génome de certains organites, c'est l'hérédité extra-nucléaire ou hérédité cytoplasmique. Le cas le plus célèbre chez les plantes est celui de la stérilité cytoplasmique mâle provoquée par des mutations dans le génome mitochondrial.

Mais une autre forme d'hérédité se traduit par l'absence de modification de la séquence du génome. Il s'agit de l'hérédité épigénétique, entraînant une modification de la chromatine.

La récente prise en compte de l'épigénétique

L'intégration des caractères souhaités durant le siècle dernier au sein des cultures a surtout été réalisée en s'appuyant sur la diversité génétique naturelle ou induite par les biotechnologies mais cette diversité ne peut expliquer la totalité de la variabilité phénotypique observée. Il est maintenant clair que les mécanismes épigénétiques peuvent contribuer à la variation phénotypique et donc jouer un rôle dans le phénomène d'héritabilité manquante (Banta and Richards 2018). Une part de cette héritabilité manquante réside dans des modifications de la chromatine, la structure protéique au sein de laquelle l'ADN est empacté. Ce phénomène appelé « régulation épigénétique » joue un rôlé clé dans la formation de la chromatine et dans l'activité transcriptionnelle des gènes. Cette régulation passe notamment par des modifications post-traductionnelles des histones (principaux constituants protéiques des chromosomes) ou la méthylation de l'ADN et peut contribuer à la variabilité phénotypique (Gallusci et al. 2017). L'information épigénétique est ainsi impliquée dans de nombreux processus physiologiques, notamment de réponse adaptative à des changements environnementaux (Meyer 2015), mais aussi dans des caractères plus communément étudiés en sélection comme le rendement chez le riz (Chen and Zhou 2013).

Concernant la phénologie, chez le châtaignier par exemple, il a été montré une augmentation du niveau de méthylation de l'ADN dans la mise en place des bourgeons à l'automne, puis une diminution lors du débourrement au printemps (Santamaría et al. 2009). Pour le pêcher, certains gènes DAM (pour Dormancy-Associated MADS-Box) impliqués dans la dormance s'enrichissent en H3K27me3, marque chromatinienne associée à l'extinction de gène, durant la levée de la dormance (De la Fuente et al. 2015). Ce phénomène est également observé chez le cerisier (Vimont et al. 2019). Les chercheurs·euses se sont donc demandé comment exploiter cette diversité épigénétique dans les programmes d'amélioration, notamment chez les espèces fruitières et forestières qui sont souvent multipliées de façon végétative, limitant la diversité génétique.

L'importance de développer des approches intégratives en génétique quantitative

Ainsi, durant ces dernières années, il a été montré le fait que les régions différemment méthylées peuvent se comporter comme des QTL (epigenetic QTLs en anglais ou epiQTLs) puisque qu'elles peuvent compter jusqu'à 90% de l'héritabilité de certains caractères comme la date de floraison et la longueur des racines chez des lignées épimutantes d'*Arabidopsis*

thaliana (Cortijo et al. 2014). Chez *Brassica napus*, il a été démontré que 97% des épiallèles se sont transmis durant cinq méioses au sein de deux lignées parentales distinctes, indiquant une relative stabilité dans leur héritabilité. Ces epiQTLs controllaient notamment des caractères liés au développement de la plante et à la qualité de la graine (Long et al. 2011). Néanmoins, si ces états épigénétiques peuvent être transmis à la descendance, ils peuvent aussi être « réinitialisés » avant d'atteindre la lignée germinale (Banta and Richards 2018). Il semble donc dorénavant nécessaire de développer des méthodes qui incorporent les effets épigénétiques dans les modèles de génétique quantitative (Tal et al. 2010).

5. Les objectifs de la thèse et la démarche scientifique

5.1. Les objectifs

Chez les espèces fruitières arboricoles avec une longue période juvénile comme le noyer, la création variétale est un processus long. Il est cependant nécessaire de renouveler les variétés françaises actuelles en raison de la concurrence mondiale grandissante et parce que ces variétés ne semblent pas pleinement adaptées aux futures conditions environnementales.

La dynamique autour des questions concernant le matériel végétal s'est essouflée à la suite de la décision de l'INRAE de Nouvelle-Aquitaine-Bordeaux, depuis 2007, de réduire le nombre d'espèces fruitières étudiées et de mettre un terme au programme d'innovation variétale qui était de 1977 à 2007 sous la direction d'Eric Germain. L'INRAE, possédant du matériel végétal « noyer » riche et unique, a donc cherché un partenaire pour procéder à un transfert. Le CTIFL, centre opérationnel de Lanxade, s'est ainsi proposé de reprendre la collection de ressources génétiques de l'INRAE. En parallèle, sur la demande de la filière nucicole, le CTIFL souhaiterait initier un nouveau programme de création variétale sur la noix, via la sélection assistée par marqueurs.

Avant d'initier un programme de création variétale, il est nécessaire d'acquérir de larges connaissances sur le matériel végétal disponible, notamment en étudiant un grand nombre de caractères agronomiques. Cet état des lieux du matériel végétal nécessite en conséquence de travailler sur de nombreuses thématiques et de répondre à plusieurs questions scientifiques. La thèse s'inscrit dans ce cadre avec des objectifs de deux types :

- Des objectifs fondamentaux avec d'une part, l'évaluation de la diversité génétique d'une collection de ressources génétiques et d'autre part, l'identification du déterminisme génétique des principaux caractères d'intérêt agronomique tels que ceux liés à la phénologie, à la qualité du fruit en coque et du cerneau et à la sensibilité à différents bioagresseurs. Concernant la phénologie, une descendance de cartographie est également disponible.
- Des objectifs appliqués avec le développement d'outils nécessaires à la mise en place d'une sélection assistée par marqueurs, tels que l'obtention de marqueurs moléculaires, la construction d'une base de données phénotypiques et génotypiques et la création d'une core collection.

Ce projet repose sur un certain nombre d'hypothèses :

- Les caractères d'intérêt agronomique étudiés sont héritables.
- Le phénotypage pour ces caractères est robuste.
- La collection de ressources génétiques comprend une variabilité phénotypique suffisamment large pour ces caractères.
- L'étendue de la diversité génétique et la puissance du génotypage de la collection permettent l'identification des régions génomiques impliquées.
- La descendance de cartographie ségrège pour les caractères liés à la phénologie et contient un nombre suffisant d'individus pour mener une étude de détection de QTL.

5.2. La démarche scientifique

La démarche scientifique retenue se décompose en trois grands axes ; l'un basé sur la collection de ressources génétiques, l'autre sur la descendance de cartographie et le dernier sur le développement d'outils nécessaires à la mise en place de la sélection assistée par marqueurs.

Le premier axe exploratoire se base sur la collection de ressources génétiques pour étudier l'architecture génétique des principaux caractères agronomiques de la fleur au fruit. Il se base sur :

- La collecte et la valorisation de données disponibles issues des archives du précédent programme de création variétale de l'INRAE de Nouvelle-Aquitaine-Bordeaux. Ces données concernent l'évaluation d'un certain nombre de caractères phénotypiques évalués pendant plusieurs années.
- Le phénotypage pendant deux à trois années des caractères d'intérêt agronomique liés à la phénologie (date de débourrement, dates de floraisons mâle et femelle), à la qualité du fruit en coque (calibre, poids, force de cassage) et du cerneau (poids, taux de remplissage, teneur en acides gras et tocophérols) et à la sensibilité aux deux principaux bioagresseurs du noyer en France (bactériose liée à *Xanthomonas arboricola* pv. *juglandis* et une forme d'anthracnose liée à *Colletotrichum acutatum*).
- Le génotypage à l'aide de marqueurs microsatellites d'une part, pour éliminer les accessions génétiquement très proches et à l'aide de marqueurs SNP d'autre part, sur le sous-ensemble retenu de la collection.

- Une analyse en génétique d'association.

Le deuxième axe se base sur la descendance F_1 de cartographie issue d'un croisement entre deux géniteurs contrastés en termes de phénologie pour valider l'architecture génétique des caractères liés à cette phénologie. Il se base sur :

- Le phénotypage des caractères liés à la phénologie pendant deux années.
- Le génotypage à l'aide de marqueurs SNP.
- Une réalisation de cartographie génétique et une recherche de QTL liés aux caractères phénologiques.

Le troisième axe consiste en le développement d'outils nécessaires à la mise en place d'une sélection assistée par marqueurs. Après création d'une core collection, il sera possible de déterminer des géniteurs·trices d'intérêt pour de futurs croisements. En parallèle, il sera possible d'obtenir des marqueurs exploitables en sélection assistée par marqueurs, soit grâce à la descendance de cartographie, soit grâce à l'analyse en génétique d'association sur la collection. En combinant marqueurs et géniteurs·trices d'intérêt, les premiers outils nécessaires à l'initiation d'un nouveau programme de création variétale seront réunis.

Ces grands axes s'inscrivent dans une démarche commune à l'équipe d'accueil « Adaptation du Cerisier au Changement Climatique » de l'UMR BFP, de l'aspect fondamental jusqu'à l'aspect appliqué. En effet, l'équipe possède une collection de ressources génétiques de cerisier maintenue par le Centre de Ressources Biologiques « *Prunus-Juglans* ». L'équipe possède en parallèle de nombreuses descendances de cartographie. Ainsi, des travaux de GWAS puis des travaux de détection de QTL sont menés sur des caractères en commun, comme ceux liés à la phénologie (dates de débourrement et de floraison) et au fruit (poids, calibre, etc.), en lien avec le changement climatique. Aussi, l'équipe mène un programme d'innovation variétale et développe donc des outils pour la mise en place d'une sélection assistée par marqueurs, qui est utilisée en routine notamment pour le poids du fruit.

Chapitre 2 – Matériels et Méthodes

1. Matériel végétal
   - 1.1. La collection de ressources génétiques *Juglans* spp. de l'INRAE de Nouvelle-Aquitaine-Bordeaux
   - 1.2. La descendance F_1 de cartographie génétique

2. Caractérisation phénotypique
   - 2.1. Caractérisation de la phénologie et du type de fructification
   - 2.2. Caractérisation de la noix sèche
   - 2.3. Caractérisation de la coque et du cerneau
   - 2.4. Caractérisation de la qualité du cerneau
   - 2.5. Visualisation et analyse des données de caractérisation phénotypique
   - 2.6. Caractérisation de la sensibilité aux deux principaux pathogènes

3. Analyses moléculaires
   - 3.1. Prélèvement du matériel végétal et extraction d'ADN
   - 3.2. Amplification par PCR de marqueurs microsatellites
   - 3.3. Génotypage à l'aide de la puce de génotypage Axiom™ *J. regia* 700K SNP

4. Analyses génétiques
   - 4.1. Calcul des paramètres de diversité
   - 4.2. Analyse de la structure de la population
   - 4.3. Détermination des arbres de distance génétique
   - 4.4. Cartographie génétique et détection de QTL
   - 4.5. Analyse en génétique d'association

Chapitre 2 – Matériels et Méthodes

1. Matériel végétal

1.1. La collection de ressources génétiques *Juglans* spp. de l'INRAE de Nouvelle-Aquitaine-Bordeaux

La collection de ressources génétiques *Juglans* spp. de l'INRAE de Nouvelle-Aquitaine-Bordeaux est une collection *in vivo* et *ex situ* maintenue au Centre de Ressources Biologiques *Prunus-Juglans*. Elle se situe au sein de l'Unité Expérimentale Arboricole (UEA) de Toulenne près de Bordeaux, selon les coordonnées GPS suivantes : latitude 44°34'37.442''N – longitude 0°16'51.48''W. Les arbres adultes sont âgés de 20 à 30 ans et sont greffés sur différents types de porte-greffe, comme du semis de *J. regia*, de *J. nigra* ou des hybrides. Certaines accessions existent en deux exemplaires, l'un à côté de l'autre. Les arbres ont été conduits classiquement, de façon similaire à ce que l'on peut trouver chez le·la producteur·trice, à l'exception :

- de la taille, très faiblement effectuée pour conserver le port naturel de l'arbre,
- des traitements limités, dont les substances cancérogènes, mutagènes et toxiques pour la reproduction (CMR) qui sont bannies puisque l'unité accueille du public,
- de la fréquence des interventions plus grande, en raison de la grande fenêtre de stades phénologiques présente,
- et des distances de plantation plus grandes, en moyenne de 10 x 10 ou 9 x 10 mètres, puisque les arbres sont voués à rester plus longtemps en verger (environ 100 arbres par hectare contre 190 en production).

Par exemple, afin de lutter contre la bactériose, 500 grammes par hectare de cuivre ont été appliqués à huit reprises entre avril et juin pour l'année 2019. Aussi, 500 kilogrammes par hectare de matière organique ont été apportés sous forme de fumure la même année. L'UEA de Toulenne possède un sol de type sablo-limoneux, avec 12 mètres d'alluvions fertiles, non sensible à l'asphyxie racinaire car drainant et offrant un certain niveau de tolérance à la sécheresse. Le climat est de type océanique, avec en moyenne 850 mm/an de précipitations, avec peu de risques de gelées et possède une résistivité homogène.

Cette collection résulte d'un important travail de prospection réalisé par Eric Germain, responsable du précédent programme de création variétale de l'INRAE de Nouvelle-

Aquitaine-Bordeaux entre 1988 et 2000, dans de nombreux pays du monde. E. Germain savait que la création d'un réservoir de diversité génétique constitue une base essentielle dans l'initiation d'un programme d'amélioration. Ainsi, cette collection représente une grande partie des espèces du genre *Juglans*. Elle possède plus de 400 individus au sein de 16 espèces. Parmi le noyer cultivé *J. regia*, la collection comprend presque 200 variétés provenant de 23 pays d'Amérique (Etats-Unis, Chili, Canada), de toute l'Europe (France, Espagne, Portugal, Royaume-Uni, Bulgarie, Roumanie, Hongrie, Ukraine, Russie, Grèce), du Proche et Moyen-Orient (Turquie, Iran, Afghanistan) et d'Asie Centrale et Orientale (Chine, Inde, Japon). L'importance de ce matériel végétal a fait l'objet d'une double page dans la revue Nutfruit Magazine publiée en juillet 2019 sous le nom de « The INRA's Walnut Germplasm Collection, a Reservoir of Diversity Essential for the Future of French Walnut Industry », disponible en Annexe 2.

Dans le cadre de la thèse, 253 accessions ont été étudiées en termes de diversité génétique et de structure de population détaillées en figure 12. Parmi ces 253 accessions, 217 appartiennent à l'espèce *J. regia*, avec 194 variétés et/ou cultivars et 23 hybrides créés par E. Germain en cours d'évaluation, puis 36 accessions appartiennent à 14 espèces apparentées. Les informations concernant ces 253 accessions sont disponibles dans les données supplémentaires de l'article publié chapitre 3, paragraphe 2.

Pour réaliser des travaux de génétique d'association, un sous-ensemble de 170 accessions de *J. regia* a été retenu, représentant la plus grande diversité sur la base des études précédentes. Ce sont ces 170 accessions qui ont été caractérisées phénotypiquement. La stratégie de construction de ce sous-ensemble est décrit à l'issue de l'article publié chapitre 3, paragraphe 2.

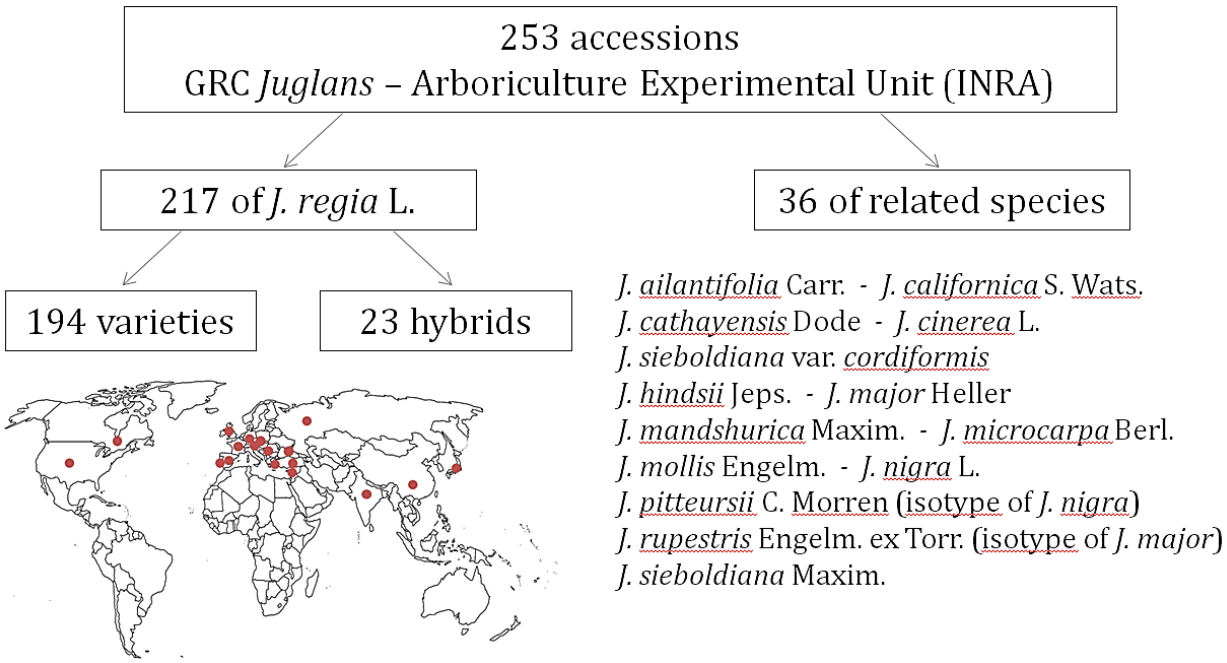


Figure 12 : Description globale du matériel végétal étudié au sein de la collection de ressources génétiques.
Une sélection de 253 individus de la collection de ressources génétiques a été réalisée afin de mener une étude de diversité génétique et de structure de population.

1.2. La descendance F_1 de cartographie génétique

La descendance F_1 de cartographie génétique est une descendance en ségrégation issue d'un croisement réalisé entre deux variétés contrastées pour des caractères liés à la phénologie. Le parent femelle est un individu de la variété 'Franquette', une variété ancienne issue de la sélection massale, représentant actuellement 70% du verger français et originaire de la rive droite de la basse vallée de l'Isère, dans la région de Vinay, en Auvergne-Rhône-Alpes. 'Franquette' possède une date de débourrement considérée comme étant plutôt tardive, en moyenne vers fin avril. Le parent mâle est l'accession 'UK 6-2', rapportée du jardin botanique de Kiev en Ukraine par E. Germain, mais qui est originaire du centre d'origine de *J. regia*, dans une zone s'étendant aujourd'hui sur trois pays d'Asie Centrale, l'Ouzbékistan, le Tadjikistan et le Kirghizistan. 'UK 6-2' possède une date de débourrement considérée comme étant plutôt intermédiaire à précoce, en moyenne vers début avril.

Cette descendance F_1 est constituée de 85 individus plantés en 2003 sur leurs propres racines à l'UEA de Toulenne. D'abord créée pour les besoins du programme de création variétale par E. Germain, cette descendance a été conservée pour sa deuxième utilité, à savoir la création d'une cartographie génétique du noyer.

2. Caractérisation phénotypique

2.1. Caractérisation de la phénologie et du type de fructification

La phénologie

Au sein de la communauté mondiale du noyer, les stades phénologiques décrits par E. Germain en figures 13 et 14 sont les stades communément admis et habituellement observés.

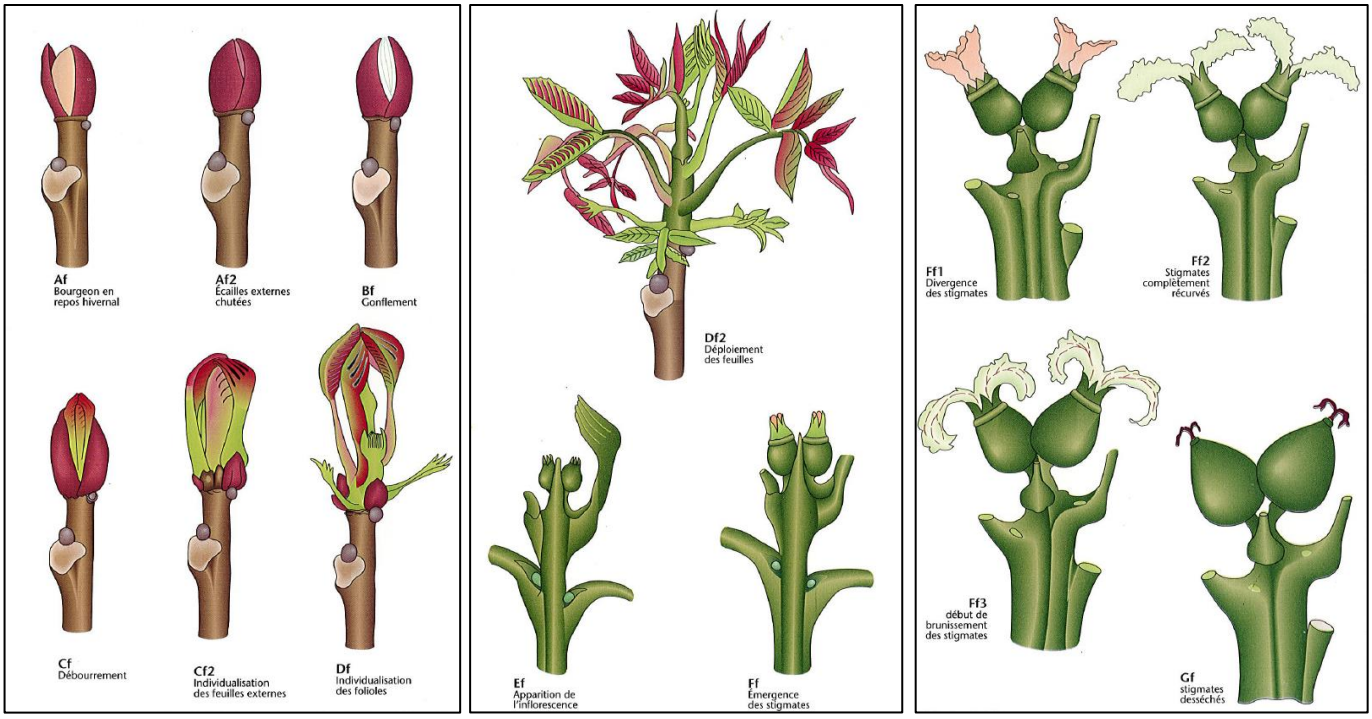


Figure 13 : Stades phénologiques de l'inflorescence femelle du noyer *J. regia* proposés par Germain et al. 1999.

Ces stades phénologiques se rapportent au bourgeon recouvert d'écailles qui donne naissance, après débourrement soit à un rameau végétatif, soit à une pousse fructifère terminée par une inflorescence femelle. Au printemps, après éclatement du bourgeon, les fleurs femelles apparaissent au sommet d'un axe feuillé. Les stades repères sont :

- Af – durant la période hivernale, le bourgeon recouvert d'écailles est à l'état dormant
- Af2 – dès janvier, les écailles dures du premier ordre tombent et le bourgeon est encore enveloppé par d'autres écailles peu différenciées semi-membraneuses
- Bf – le bourgeon gonfle, les enveloppes externes se desserrent et les extrémités des bractées sous-jacentes recouvertes d'un duvet blanchâtre apparaissent : c'est le stade dit « laineux » ou « bourgeon blanc »
- Cf – le bourgeon s'allonge et l'extrémité des folioles terminales des feuilles les plus extérieures se distinguent : c'est le débourrement
- Cf2 – les écailles et bractées s'écartent, les premières feuilles commencent à s'individualiser
- Df – le bourgeon est ouvert, les premières feuilles se séparent et leurs folioles sont bien individualisées
- Df2 – les premières feuilles sont complètement déployées ; d'abord dressées, elles prennent ensuite un port plus ou moins oblique laissant apparaître en leur centre les fleurs femelles
- Ef – apparition des fleurs femelles
- Ff – apparition des stigmates
- Ff1 – les stigmates de couleur jaune orangé sont divergents et leur réceptivité est optimale : c'est le début de la pleine floraison femelle
- Ff2 – les stigmates prennent une couleur vert-jaune pâle et sont complètement récurvés
- Ff3 – les papilles des stigmates commencent à se nécroser et ceux-ci se strient de fins filets bruns
- Gf – dessèchement et noircissement des stigmates

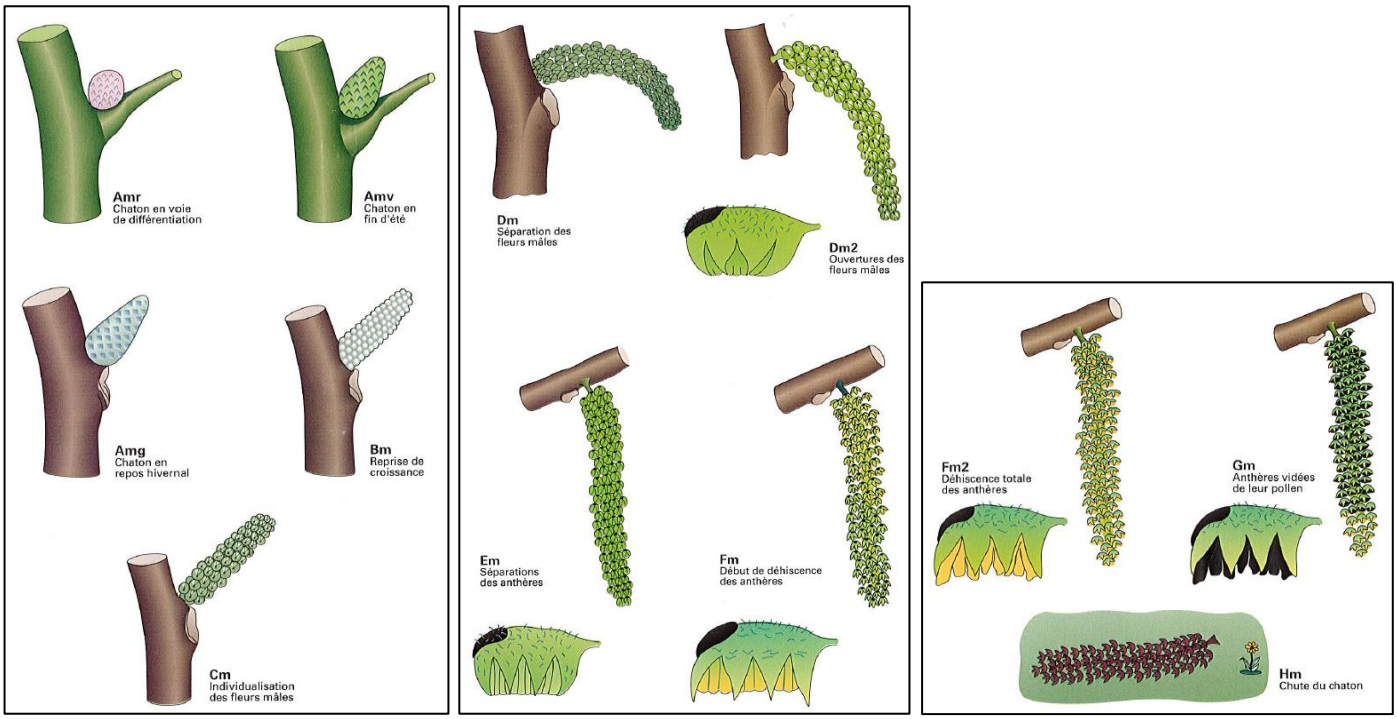


Figure 14 : Stades phénologiques de l'inflorescence mâle du noyer *J. regia* proposés par Germain et al. 1999.

Ces stades phénologiques se rapportent à l'inflorescence mâle, petit chaton qui apparait en début d'été, qui grossit très lentement jusqu'au repos hivernal et qui s'allonge rapidement au printemps pour libérer son pollen. Le chaton n'étant pas enveloppé d'écailles, les stades caractérisent l'inflorescence elle-même :

- Amr – au début de l'été, le chaton mâle différencié de forme globuleuse a une teinte rosée, alors que les bourgeons restent verts
- Amv – durant l'été, le chaton croît légèrement et devient conique pour atteindre environ 5 mm de longueur avec une coloration verte
- Amg – début octobre, le chaton cesse de croître, il mesure 5 à 8 mm et prend une coloration grise qu'il conservera tout l'hiver
- Bm – environ 3 semaines avant le débourrement, sa croissance reprend, le chaton gonfle et s'allonge pour atteindre 13 à 20 mm de longueur
- Cm – le chaton, raide et oblique, atteint la grosseur d'un crayon à papier et mesure 30 à 40 mm. Sa couleur passe progressivement du vert-brun au vert-clait et les glomérules des fleurs sont bien distincts
- Dm – le chaton perd sa rigidité et devient semi-retombant ; les glomérules se séparent
- Dm2 – les glomérules s'espacent et commencent à s'ouvrir ; le chaton pend
- Em – ouverture complète des glomérules et séparation des anthères qui commencent à jaunir
- Fm – début de déhiscence des anthères à partir de la base du chaton
- Fm2 – déhiscence totale des anthères et pleine émission du pollen
- Gm – les anthères vidées de leur pollen noircissent
- Hm – le chaton tombe à terre et se dessèche

Ces stades sont ceux officiellement reconnus comme descripteurs de la phénologie du noyer par Bioversity International, par le passé International Plant Genetic Resources Institute (IPGRI 1994). L'échelle BBCH du noyer suit d'ailleurs ces mêmes stades phénologiques.

Durant toute la période allant des stades Af2 pour la fleur femelle et Bm pour la fleur mâle, jusqu'aux stades Gf et Hm respectivement, soit de début mars à début juin, deux passages par semaine ont été effectués pour les 170 accessions de *J. regia* de la collection et

pour les 85 individus du croisement ‘Franquette’ × ‘UK 6-2’ pour observer et noter les stades des deux inflorescences ; et ce, durant les trois printemps 2018, 2019 et 2020.

A la suite de ces deux passages hebdomadaires, les principaux stades sont convertis en jours julien (nombre de jours écoulés depuis le 1er janvier de l’année en question). Ces principaux stades sont pour l’inflorescence femelle : Cf (*débourrement*), Ff1 (*début de floraison femelle*), Ff2 (*pleine floraison femelle*) et Ff3 (*fin de floraison femelle*), et pour l’inflorescence mâle : Fm (*début de floraison mâle*), Fm2 (*pleine floraison mâle*) et Gm (*fin de floraison mâle*).

Ainsi, la *durée de la floraison femelle* est calculée par Ff3 - Ff1 et la *durée de la floraison mâle* par Gm - Fm. En soustrayant Fm2 et Ff2 (les pleines floraisons mâle et femelle), il est aussi possible d’obtenir la *dichogamie* selon cette échelle catégorielle : 1- très protandre (supérieur à -10 jours), 3- protandre (entre -4 et -10 jours), 5- homogame (entre -3 et +3 jours), 7- protogyne (entre +4 et +10 jours) et 9- très protogyne (supérieur à +10 jours).

Le type de fructification

Aussi, le type de fructification est noté pour les 170 accessions de *J. regia* de la collection, mais pas pour les 85 individus du croisement, ce caractère ne ségrégeant pas. Ce caractère est noté une seule fois puisqu’il est stable au cours du temps. Il existe quatre principaux types de fructification chez *J. regia* avec une graduation catégorielle entre ces types, qu’il n’est pas simple de déterminer :

a) *la fructification en position terminale* – sur les rameaux d’un an, les bourgeons à fruits sont uniquement localisés en position terminale, du fait d’une forte tendance acrotone de ces rameaux ; peu de ramifications secondaires sont donc observées et la fructification est alors essentiellement localisée en périphérie de l’arbre ; il est représentatif des populations de noyer de France et de l’Est de l’Europe.

b) *la fructification de type intermédiaire* – comme prédédemment, l’induction florale femelle n’intéresse que les bourgeons en position terminale mais au printemps, le démarrage d’un nombre non négligeable de pousses latérales est observé en position apicale mais aussi en position médiane ; l’arbre est plus ramifié et la productivité est plus importante ; la variété californienne « Hartley » est représentative de ce type.

c) *la fructification sur brindilles latérales* – les bourgeons à fruits sont cette fois répartis le long des pousses d’un an et ils débourrent au printemps en majorité pour donner naissance le long du rameau à de courtes brindilles fructifères, du fait d’un faible gradient acrotone ; le potentiel de production est nettement supérieur et ce type de fructification est

plutôt rare chez *J. regia*, sauf dans les populations du centre d'origine en Asie Centrale ; la variété californienne « Payne » est représentative de ce type.

d) un quatrième type un peu à part est *la fructification en grappes*, caractéristique de quelques accessions de la collection provenant du jardin botanique de Kiev en Ukraine.

Le *type de fructification* a été noté selon cette échelle catégorielle : 1- terminale, 3- intermédiaire, 5- latérale faible, 7- latérale forte et 9- grappes.

2.2. Caractérisation de la noix sèche

Au préalable à leur caractérisation, les noix fraîches sont séchées grâce au principe du séchage par ventilation. C'est un phénomène d'extraction de l'eau présente dans les noix par pulsion d'un courant d'air chaud et sec autour d'elles. Cet air a un pouvoir séchant variable en fonction de sa température et de son hygrométrie. Un séchoir à faux-fond a été utilisé ; c'est le type de séchoir le plus répandu. Il s'agit d'une cellule remplie de noix avec un faux-fond grillagé laissant circuler l'air chaud. La noix fraîche possède en moyenne une teneur en eau de 40%. Il est conseillé d'atteindre une teneur de 12% pour une noix sèche. Ainsi, en conditions classiques d'un entrepôt à 25°C pour une hygrométrie de 45%, il a été déterminé selon un mémo disponible dans l'ouvrage « Le séchage des noix, guide pratique » du CTIFL destiné à la profession, qu'il fallait 4 jours de séchage pour un air pulsé à 25°C.

La noix sèche a été caractérisée au sein des 170 accessions de *J. regia* de la collection pour différents caractères pour les trois récoltes 2017, 2018 et 2019. Ce phénotypage a pu être réalisé notamment grâce à l'aide précieuse de Morgane Vincent, apprentie au sein du CTIFL, et de Julie Crabier, stagiaire Master 2 au sein de l'Université de Bordeaux et financée par le GIS Fruits. L'encadrement de Morgane et de Julie s'est inscrit dans le déroulement de la présente thèse. D'abord, un lot homogène de 100 noix a été constitué pour chaque accession. Sur chacun des lots de 100 noix ont été mesurés les caractères suivants.

Le poids

Le poids de 100 noix sèches pour les récoltes 2017, 2018 et 2019 est déterminé en grammes selon une amélioration du descripteur « 6.5.12 Inshell nut weight (g) » de Bioversity International (IPGRI, 1994) préconisant la pesée de 20 noix. La pesée est effectuée à l'aide d'une balance de précision à 10^{-1} grammes.

Le calibre

Le calibre de 100 noix sèches pour les récoltes 2017, 2018 et 2019 est déterminé en pourcentage du poids total de 100 noix et en pourcentage du nombre total de noix. Chaque noix est calibrée à l'aide d'un dispositif mis au point par la Station Expérimentale de la Noix de Creysse permettant de séparer les noix selon 7 calibres : <28mm, 28-30mm, 30-32mm, 32-34mm, 34-36mm, 36-38mm, >38mm. Ensuite, les noix sont pesées en lot par calibre et comptabilisées pour chaque calibre.

La hauteur

La hauteur de 100 noix sèches pour la récolte 2017 est déterminée en millimètres selon une amélioration du descripteur « 6.5.3 Nut length (mm) » de Bioversity International (IPGRI, 1994) préconisant la mesure de 20 noix. Chaque noix est mesurée une par une à l'aide d'un pied à coulisse électronique de précision à 10^{-2} millimètres (Figure 16). Pour la récolte 2018, cette mesure a été obtenue sur un sous-lot randomisé de 50 noix sèches via l'utilisation de la tomographie à rayons X, analyse réalisée par le BioGEVES*.

Les largeurs de face et de profil

Les largeurs de face (section longitudinale au niveau de la suture de la coque) et de profil (section perpendiculaire à la suture de la coque) de 100 noix sèches pour la récolte 2017 sont déterminées en millimètres selon une amélioration du descripteur « 6.5.2 Nut diameter (mm) » de Bioversity International (IPGRI, 1994) préconisant la mesure de 20 noix. Chaque noix est mesurée une par une de face et de profil à l'aide d'un pied à coulisse électronique de précision à 10^{-2} millimètres (Figure 15). Pour la récolte 2018, ces mesures ont été obtenues sur un sous-lot randomisé de 50 noix sèches via l'utilisation de la tomographie à rayons X, analyse réalisée par le BioGEVES*.

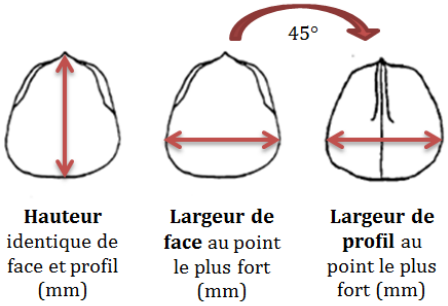


Figure 15 : Mesures de la hauteur et des largeurs d'une noix.
La hauteur de la noix est évidemment identique qu'elle soit mesurée de face ou de profil. A partir de la face, une rotation de 45° de la noix permet d'avoir le profil. Les mesures sont effectuées au niveau le plus « bombé », représentatif du calibre.

Les formes de face et de profil

Les formes de face et de profil sont déterminées en faisant la moyenne visuelle de 20 noix sèches au hasard pour la récolte 2017 selon le descripteur « 6.5.1 Nut shape » de Bioversity International (IPGRI, 1994) (Figure 16). Pour la récolte 2018, la forme globale de la noix est donnée par un indice de forme obtenu sur un sous-lot randomisé de 50 noix sèches via l'utilisation de la tomographie à rayons X, analyse réalisée par le BioGEVES*. La tomographie a également permis d'obtenir un indice de forme de « Feret », défini par D/d où d est le dimètre de Feret minimum et D le diamètre de Feret maximum dans une direction orthogonale direction.

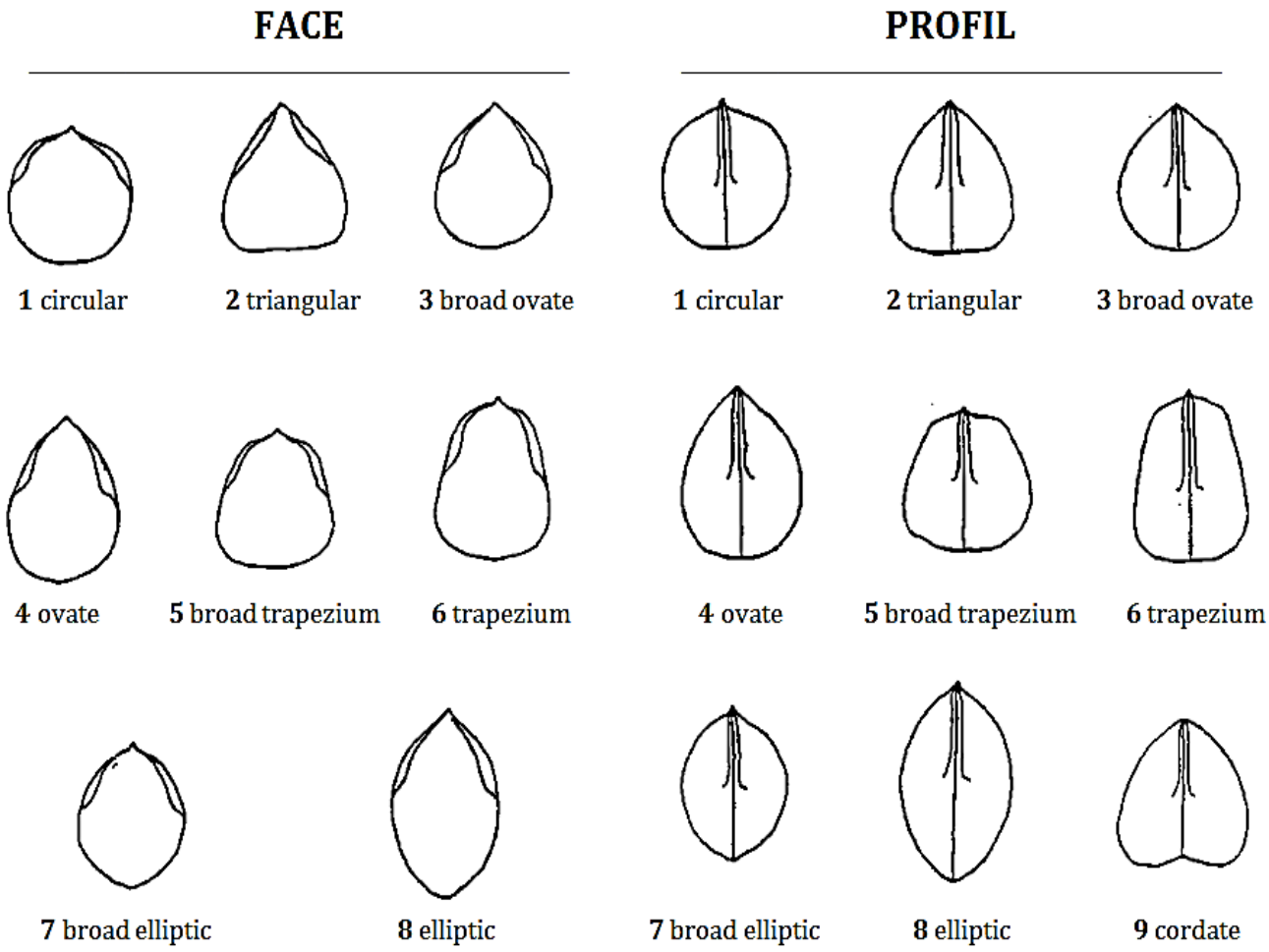


Figure 16 : Dessins des différentes formes majeures de noix de face et de profil.

La sphéricité

Uniquement pour la récolte 2018, la sphéricité exprimée sur une échelle allant de 0 à 1 a été mesurée sur un sous-lot randomisé de 50 noix sèches via l'utilisation de la tomographie à rayons X, analyse réalisée par le BioGEVES*. Elle indique également indirectement la forme de la noix (1 = sphère parfaite).

La rugosité

Uniquement pour la récolte 2018, la rugosité exprimée sur une échelle allant de 1 à 2,5 a été mesurée sur un sous-lot randomisé de 50 noix sèches via l'utilisation de la tomographie à rayons X, analyse réalisée par le BioGEVES*. Elle indique le niveau de veinure de la noix.

2.3. Caractérisation de la coque et du cerneau

La coque et le cerneau ont été caractérisés au sein des 170 accessions de *J. regia* de la collection pour différents caractères pour les trois récoltes 2017, 2018 et 2019. Sur le lot de 100 noix qui a servi à caractériser la noix sèche (partie 2.2.), un sous-lot randomisé de 50 noix a été réalisé. Sur chacun des lots de 50 noix, ont été mesurés les caractères suivants.

Les forces de rupture de la coque et des valves

Les forces de rupture de la coque pour les 25 premières noix et des valves pour les 25 autres noix pour les récoltes 2017, 2018 et 2019 sont déterminées à l'aide d'un texturomètre de la marque Lloyd Materials Testing™, de la société Ametek®, série TA1 Texture Analyzer, modèle TA-PLUS. Un texturomètre est un appareil permettant d'étudier les propriétés physiques d'un matériau en appliquant sur celui-ci différentes forces comme la traction, la pénétration, la friction ou la compression. Dans le cadre d'une étude de force de compression, le capteur de force se compose de quatre jauges de déformation (Figure 17). Le principe du capteur de force consiste à traduire la déformation du support sur lequel sont positionnées ces jauges, due à la compression sur le matériau étudié, par la variation d'une résistance électrique. La noix est déposée sur un plateau en bois créé spécifiquement par les techniciens du CTIFL pour ce phénotypage, afin de bien la maintenir. Le capteur de force est relié au programme de pilotage « NexygenPlus3 » de la société Ametek® via lequel il est possible de paramétrer les essais de cassage.

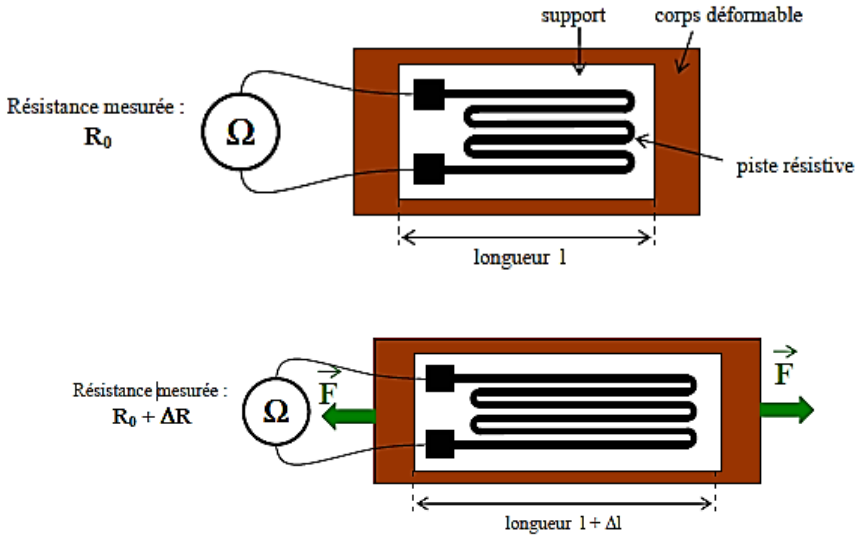


Figure 17 : Principe de la jauge de déformation.
La jauge de déformation est constituée d'un fil conducteur très fin collé sur un support. Ce fil subit les mêmes déformations que le support. La résistance d'un fil conducteur est définie par la loi de Pouillet : $R = \rho \times (L/S)$ où R est la résistance électrique, ρ la résistivité du matériau, L la longueur du fil et S la section du fil. La déformation de la jauge – ici, un étirement – modifie la longueur L et donc, la résistance R.

Un protocole spécifique pour un matériau du type noix en coque a été développé grâce au formidable travail de Nicolas Jammes, stagiaire du 9 avril au 22 juin 2018, dans le cadre de la validation de son DUT Mesures Physiques, à l'Université de Limoges. L'encadrement de Nicolas s'est inscrit dans le déroulement de la présente thèse. L'intitulé de son stage était : « Création d'un protocole permettant la détermination de la dureté des noix ». Il a montré que la force de rupture est grandement différente selon la position (test de la coque ou de la valve) et selon la variété (test de trois variétés). Ceci a donc conforté l'idée de mesurer ces deux forces de rupture. En revanche, le calibre n'a pas montré d'influence sur ces forces mesurées. Ainsi, les noix peuvent être choisies au hasard dans le lot. Les paramètres qui ont été retenus sont une vitesse d'allongement de 75 mm/min (vitesse de compression de la noix) avec une détection de la rupture lorsque la force chute brusquement de 80%. C'est évidemment une mesure destructive et l'unité de la mesure est en Newtons.

L'épaisseur de la coque

L'épaisseur de la coque sur 25 coques choisies aléatoirement pour la récolte 2017 est déterminée après ouverture des 50 noix. Chaque coque est mesurée à l'équateur sur le point le plus éloigné des pôles et à l'opposé des valves pour obtenir l'épaisseur de la coque côté face à l'aide d'un pied à coulisse électronique de précision à 10^{-2} millimètres. Pour la récolte 2018, ces mesures ont été obtenues sur un sous-lot randomisé de 50 noix sèches via l'utilisation de la tomographie à rayons X, analyse réalisée par le BioGEVES*.

Le poids des 100 demi-cerneaux

Le poids des 100 demi-cerneaux issus du cassage des 50 noix pour les récoltes 2017, 2018 et 2019 est déterminé en grammes selon une amélioration du descripteur « 6.5.13 Kernel weight (g) » de Bioversity International (IPGRI, 1994) préconisant la pesée de 20 cerneaux. Les 100 demi-cerneaux sont pesés ensemble à l'aide d'une balance de précision à 10^{-1} grammes.

Le rendement au cassage

Le rendement au cassage est défini pour les récoltes 2017, 2018 et 2019 par le ratio entre le poids des cerneaux hors déchets et le poids des noix en coque. Sont considérés comme déchêts les cerneaux qui ne sont pas commercialisables, comme les cerneaux flétris ou mal-nourris, les cerneaux moisis et les cerneaux noirs suite à l'action d'un pathogène (Figure 18). La formule est :

Rendement au cassage commercial =

[(poids de 100 demi-cerneaux hors déchêts) / (poids des 50 noix en coque)] x 100

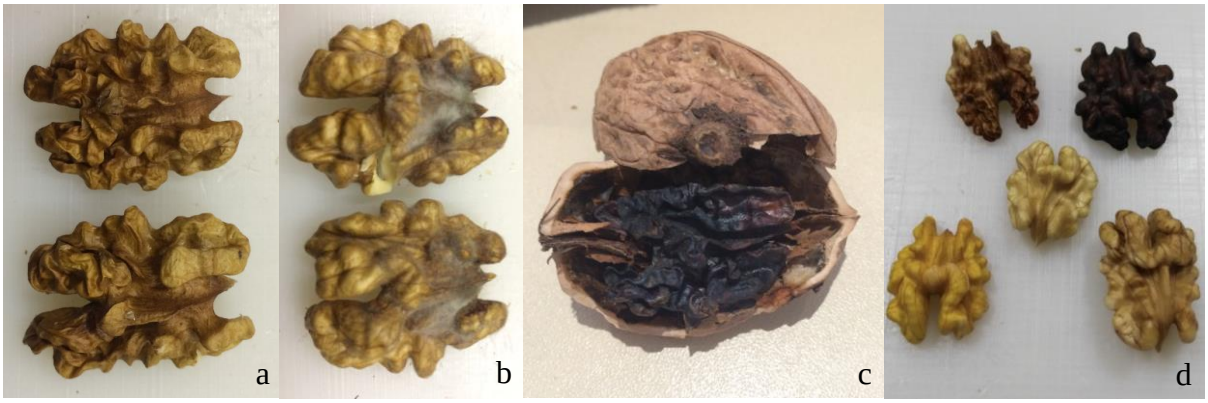


Figure 18 : Photographies de cerneaux non-commercialisables.
Exemples de déchets : a. Cerneau mal-nourri ou flétri, b. Cerneau moisi, c. Cerneau noir dû à la bactériose, d. En bas, trois cerneaux commercialisables – En haut, cerneau flétri à gauche, cerneau noir à droite.

Pour la récolte 2018, le rendement au cassage global est également défini comme le ratio entre le volume occupé par les cerneaux et le volume des noix en coque, en mm^3. Ces mesures ont été obtenues sur un sous-lot randomisé de 50 noix sèches via l'utilisation de la tomographie à rayons X, analyse réalisée par le BioGEVES*.

Les volumes de la noix et l'aire de sa surface

Pour la récolte 2018, l'utilisation de la tomographie à rayons X a également permis d'obtenir tous les volumes des différentes parties de la noix en mm^3, comme le volume total

de la noix, le volume de la coque, le volume du cerneau et le volume de l'espace vide. L'aire de la surface totale de la noix en mm² a également été mesurée.

**Utilisation de la tomographie à rayons X*

Dans les grandes lignes, la tomographie à rayons X repose sur l'absorption de rayons X par un objet et permet la reconstruction d'images en coupe de cet objet en trois dimensions (Figure 19). Il s'agit d'une technique non destructive et précise. Néanmoins, son coût est élevé, la mise au point d'un protocole n'est pas aisée et la taille des données acquises est conséquente. En ce qui concerne le matériel utilisé, le BioGEVES possède un tomographe 3D à rayons X, modèle X50 de la marque North Star Imaging.

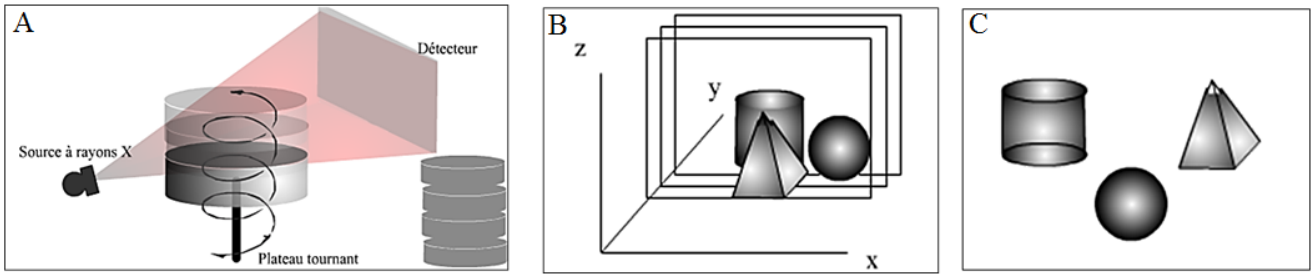


Figure 19 : Principe de la tomographie 3D.
(A) Acquisition des images selon plusieurs projections. (B) Reconstruction du volume 3D à partir des projections. (C) Traitement et extraction d'informations à partir des images.

Un total de 2 164 images (ou radiographies) a été obtenu, avec une résolution de 992x992x2991 voxels (taille de voxel : 0,1x0,1x0,1 mm), un courant électrique de 300 µA et une vitesse de rotation de 4,99 degrés/s, donnant une durée de scan de 14 minutes et 34 secondes. Au niveau de la méthodologie, il est d'abord nécessaire de séparer virtuellement chaque noix. En effet, plusieurs noix d'un même lot sont analysées ensemble pour diminuer le nombre de scans. Pour cela, le programme « Avizo™ v9.0.0 » de Thermo Scientific™ a été utilisé pour l'étape de reconstruction 3D. Une première étape de séparation a été réalisée afin de pouvoir mesurer chaque noix individuellement. Ensuite, de multiples étapes de traitement d'images ont été appliquées afin de séparer les trois éléments du fruit, à savoir la coque, le cerneau et l'espace vide. Ces étapes consistent grossièrement à faire apparaître ou disparaître différents niveaux d'arrière-plan ou d'avant-plan.

Les différentes mesures indiquées précédemment par un astérisque ont été obtenues grâce aux fonctions automatiques du programme « Avizo™ v9.0.0 » et à celles du programme « MATLAB® ». Grâce aux mesures de largeur et hauteur, des caractères sont automatiquement calculés comme les volumes et les aires des surfaces. Deux caractères supplémentaires ont été calculés avec une méthode non standard puisque non quantifiables

par ces programmes. Ainsi, le calcul de la sphéricité et de la rugosité ont été mis au point au préalable par le BioGEVES dans le cadre d'un projet ANR visant à améliorer la compétitivité de la filière de la betterave sucrière (projet AKER ANR-11-BTBR-0007). Ainsi, les formules sont :

Sphéricité = $\frac{\pi^{\frac{1}{3}}(6V)^{\frac{2}{3}}}{A}$ et Rugosité = $\frac{A}{\sqrt[3]{36\pi V^2}}$ où V est le volume de la noix et A son aire en 3D. La sphéricité d'une sphère parfaite est égale à 1 et tout objet qui n'est pas une sphère parfaite aura une sphéricité inférieure à 1.

2.4. Caractérisation de la qualité du cerneau

Après cette caractérisation physique, 100 grammes de cerneaux de chacune des 170 accessions de *J. regia* de la collection ont été analysés au niveau biochimique par l'ITERG, l'institut des corps gras, pour la récolte 2018. Des molécules « marqueurs » de la qualité nutritionnelle et de l'aptitude à la conservation de l'huile de noix ont été quantifiées, telles que les acides gras et la vitamine E.

Une extraction qualitative non altérante de la matière grasse a été réalisée selon une méthode interne : broyage des cerneaux avec un broyeur de la marque Retsch, modèle ZM 200, extraction de la matière grasse par un mélange hexane/isopropanol 3/2 (v/v), miscella lavé au chlorure de potassium KCl à 0,8% et évaporation sous-vide avec un évaporateur rotatif de la marque Buchi. Les esters méthyliques d'acides gras ont été analysés par chromatographie en phase gazeuse selon les normes NF EN ISO 12966-2 et 12966-4. Les tocophérols et tocotriénols ont été analysés par chromatographie en phase liquide à haute performance selon la norme NF EN ISO 9936. Ces analyses ont permis d'obtenir les caractères suivants.

La *teneur en acides gras saturés*, la *teneur en acides gras mono-insaturés* et la *teneur en acides gras-polyinsaturés*, données en pourcentage de la teneur en acides gras totaux.

La *teneur totale en tocophérols*, formes naturelles de la vitamine E (α-tocophérol, β-tocophérol, γ-tocophérol et δ-tocophérol), donnée en mg/kg de cerneaux. L'*activité vitaminique E*, donnée en mg d'alpha-tocophérol-équivalent/100 g de cerneaux).

Le tableau 1 synthétise les détails de l'ensemble des caractères présentés dans les parties 2.1. à 2.4 au cours des différentes campagnes de phénotypage. Pour plus de clarté, y sont indiquées pour les caractères physiques les mesures effectuées à la main ou par analyse

par tomographie à rayons X. En raison du confinement à la suite de la pandémie de Covid-19, la troisième campagne de phénotypage de la phénologie au premier trimestre 2020 a été annulée.

2.5. Visualisation et analyse des données de caractérisation phénotypique

L'ensemble de ces données de phénotypage nouvellement acquises durant la thèse ont été visualisées grâce à la réalisation d'une ACP. Le programme et langage R a été utilisé de nombreuses fois tout au long des analyses. Ce programme est disponible à cette adresse : http://www.R-project.org/. Les packages R « FactoMineR » (Lê et al. 2008), « factoextra », « corrplot » (Wei and Simko 2017) et « ggplot2 » ont été utilisés pour produire l'ACP, mais aussi pour créer des nuages de points, des diagrammes de densité ainsi que des matrices de corrélation.

2.6. Caractérisation de la sensibilité aux deux principaux pathogènes

En raison du nombre important de caractères cités précédemment et étudiés durant la thèse, la caractérisation de la sensibilité aux deux principaux pathogènes du noyer a été confiée chaque année à des stagiaires. L'encadrement de ces stagiaires s'est inscrit dans le déroulement de la présente thèse. En effet, aucun protocole de routine n'a encore été développé à ce jour pour ces caractérisations. Le but de ces différents stages a été donc de mettre au point des méthodes de culture de ces pathogènes et de leur inoculation sur différents organes détachés du noyer ; organes détachés puisque des tests en conditions de verger ne sont pas réalisables. Les deux principaux pathogènes du noyer sont d'une part, le complexe d'espèces *Colletotrichum acutatum*, ascomycètes responsables d'une forme d'anthracnose et d'autre part, la bactérie *Xanthomonas arboricola* pv. *juglandis*, responsable de la bactériose. Aucun moyen de lutte efficace n'existe à ce jour pour ces deux maladies. Ainsi, plusieurs travaux ont été réalisés sur ces pathogènes durant les saisons de printemps/été en 2017, 2018 et 2019. Une synthèse des travaux réalisés en pathologie se situe en Annexe 3.

		2017	2018	2019	2020
Phénologie	Début de floraison mâle	/	Deux observations hebdomadaires - donnée en jour julien	Deux observations hebdomadaires - donnée en jour julien	Deux observations hebdomadaires - donnée en jour julien
	Pleine floraison mâle	/	Deux observations hebdomadaires - donnée en jour julien	Deux observations hebdomadaires - donnée en jour julien	Deux observations hebdomadaires - donnée en jour julien
	Fin de floraison mâle	/	Deux observations hebdomadaires - donnée en jour julien	Deux observations hebdomadaires - donnée en jour julien	Deux observations hebdomadaires - donnée en jour julien
	Durée de floraison mâle	/	Calculée par différence	Calculée par différence	Calculée par différence
	Débourrement	/	Deux observations hebdomadaires - donnée en jour julien	Deux observations hebdomadaires - donnée en jour julien	Deux observations hebdomadaires - donnée en jour julien
	Début de floraison femelle	/	Deux observations hebdomadaires - donnée en jour julien	Deux observations hebdomadaires - donnée en jour julien	Deux observations hebdomadaires - donnée en jour julien
	Pleine floraison femelle	/	Deux observations hebdomadaires - donnée en jour julien	Deux observations hebdomadaires - donnée en jour julien	Deux observations hebdomadaires - donnée en jour julien
	Fin de floraison femelle	/	Deux observations hebdomadaires - donnée en jour julien	Deux observations hebdomadaires - donnée en jour julien	Deux observations hebdomadaires - donnée en jour julien
	Durée de floraison femelle	/	Calculée par différence	Calculée par différence	Calculée par différence
	Dichogamie	/	Calculée par différence	Calculée par différence	Calculée par différence
	Type de fructification	/	/	Une seule observation	/
Noix sèche	Poids de la noix	100 noix - donné en dixième de grammes	100 noix - donné en dixième de grammes	100 noix - donné en dixième de grammes	/
	Calibre de la noix	100 noix - donné en dixième de grammes	100 noix - donné en dixième de grammes	100 noix - donné en dixième de grammes	/
	Hauteur de la noix	100 noix - donnée en centième de millimètres	50 noix - donnée en centième de millimètres	/	/
	Largeur de face	100 noix - donnée en centième de millimètres	50 noix - donnée en centième de millimètres	/	/
	Largeur de profil	100 noix - donnée en centième de millimètres	50 noix - donnée en centième de millimètres	/	/
	Forme de face	20 noix - donnée selon grille	50 noix - donnée par un indice de forme globale	/	/
	Forme de profil	20 noix - donnée selon grille		/	/
	Sphéricité de la noix	/	50 noix - donnée par un indice de sphéricité	/	/
	Rugosité de la noix	/	50 noix - donnée par un indice de rugosité	/	/
Coque et cerneau	Force de rupture de la coque	25 noix - donnée en Newtons	25 noix - donnée en Newtons	25 noix - donnée en Newtons	/
	Force de rupture de la valve	25 noix - donnée en Newtons	25 noix - donnée en Newtons	25 noix - donnée en Newtons	/
	Epaisseur de la coque	25 morceaux de noix - donnée en centième de millimètres	50 noix - donnée en centième de millimètres	/	/
	Poids du cerneau	100 demi-cerneaux - donné en dixième de grammes	100 demi-cerneaux - donné en dixième de grammes	100 demi-cerneaux - donné en dixième de grammes	/
	Rendement au cassage	ratio poids des cerneaux hors déchets et poids des noix en coque	ratio poids des cerneaux hors déchets et poids des noix en coque	ratio poids des cerneaux hors déchets et poids des noix en coque	/
Qualité	Teneur en acides gras saturés	/	100 grammes de cerrneaux - donnée en % du total	/	/
	Teneur en acides gras mono-insaturés	/	100 grammes de cerrneaux - donnée en % du total	/	/
	Teneur en acides gras poly-insaturés	/	100 grammes de cerrneaux - donnée en % du total	/	/
	Teneur totale en tocophérols	/	100 grammes de cerrneaux - donnée en mg/kg de cerneau	/	/
	Activité vitaminique E	/	100 grammes de cerneaux - donnée en mg d'alpha TE/100 g de cerneau	/	/

En noir, phénotypage réalisé par Anthony Bernard

En bleu et gras, phénotypage réalisé par le BioGEVES

En vert et gras, phénotypage réalisé par l'ITERG (TE: tocophérol équivalent)

En rouge, phénotypage annulé pour cause du confinement à la suite de la pandémie de Covid-19

Tableau 1 : Synthèse des campagnes de caractérisation phénotypique.

3. Analyses moléculaires

3.1. Prélèvement du matériel végétal et extraction d'ADN

Pour chacun des 253 individus de la collection de ressources génétiques, un des deux exemplaires a été sélectionné et identifié. Pour la descendance de cartographie, chacun des 85 individus est unique et sur ses propres racines. Entre les mois d'avril et juin 2017, 5 à 10 jeunes folioles ont été prélevées pour chaque individu et insérées dans un sachet kraft. Entre le trajet de l'UEA de Toulenne et de l'INRAE de Nouvelle-Aquitaine-Bordeaux, les sachets kraft ont été stockés dans une glacière réfrigérée. À l'INRAE, les folioles ont été stockées une nuit en chambre froide à +4°C et le lendemain, elles ont été envoyées par courrier express en 24h au laboratoire du BioGEVES. Une fois au BioGEVES, les échantillons ont été congelés durant une nuit à -80°C pour ensuite être lyophilisés.

Pour l'extraction d'ADN en plaque de 96 puits, le laboratoire du BioGEVES a utilisé le kit d'extraction d'ADN NucleoSpin® 96 Plant II Core Kit du fabricant Macherey-Nagel, selon les recommandations du fabricant. Ce kit utilise la technologie de séparation de l'ADN par adsorption sur une membrane de silice et permet un rendement compris entre 1 à 30 µg d'ADN pour 100 mg de matériel végétal. Le volume d'élution est compris entre 100 à 200 µL et le temps d'extraction moyen pour une plaque est de 60 minutes sans compter le temps de lyse.

La normalisation de la concentration en ADN a été réalisée grâce au lecteur de microplaques LVis en absorbance SPECTROstar® Omega du fabricant BMG Labtech, permettant un dosage par spectrophotométrie dans un volume de 2 µL de la solution mère d'ADN. Les extraits sont ensuite normalisés à une concentration de 2,5 ng/µL.

3.2. Amplification par PCR de marqueurs microsatellites

Tous les individus (253 de la collection et 85 de la descendance de cartographie) ont été génotypés au laboratoire du BioGEVES grâce à 15 marqueurs microsatellites sélectionnés dans la littérature (Dangl et al. 2005; Woeste et al. 2002; Dang et al. 2016). Les séquences des amorces, les motifs de répétition et les références se trouvent dans le Tableau 1 de l'article du chapitre 3, paragraphe 2.1.

La séquence des amorces « Forward » a été additionnée d'une queue 35S constituée de 19 paires de bases à l'extrémité 5'. Les réactions PCR ont été réalisées dans un volume de 10 µL contenant 4 µL d'ADN dilué à 2,5 ng/µL et 6 µL de mix PCR. Le mélange PCR est

composé de 3,80 µL d'eau ultra-pure, de 1 µL de tampon 10X, de 0,2 µL de dNTP à une concentration de 10 mM, de 0,60 µL de $MgCl_2$ à une concentration de 25 mM, de 0,10 µL de polymérase Ampli Taq Gold, de 0,10 µL d'amorces « Reverse » et de 1 µM d'amorces « Forward ».

Les étapes de la PCR sont : 10 min à 94°C, 30 ou 35 cycles (30 sec à 94°C, 1 min à 58°C, 40 sec à 7°C) et 10 min à 72°C. Les amplicons sont dilués et 5,8 µL de formamide, 0,2 µL de marqueur de taille et 4 µL d'amplicons sont additionnés dans chaque puits. Un analyseur par électrophorèse sur capillaires Applied Biosystem 3730 DNA a été utilisé et les données ont été traitées grâce au programme GeneMapper™, tous les deux de la marque ThermoFisher®.

3.3. Génotypage à l'aide de la puce de génotypage Axiom™ *J. regia* 700K SNP

Après le travail réalisé avec les marqueurs microsatellites, un sous-ensemble d'accessions de la collection de ressources génétiques a été sélectionné pour réaliser d'une part, une étude de génétique d'association et d'autre part, une cartographie génétique avec détection de QTL. 200 accessions ont donc été génotypées au sein de la collection de ressources génétiques, ainsi que 85 individus de la descendance de cartographie, à l'aide d'une puce de génotypage à haute densité de 700K SNP spécifique de *J. regia*, mise au point grâce à l'équipe de l'Université de Davis en Californie (Marrano et al. 2019a) et la société Affymetrix®. Le programme « Axiom Analysis Suite 3.1 » lié à la technologie Affymetrix® a été utilisé pour effectuer le filtre qualité des données de génotypage.

4. Analyses génétiques

4.1. Calcul des paramètres de diversité

La diversité génétique d'une population, naturelle ou artificielle, est très souvent présentée en premier lieu grâce au calcul des paramètres de diversité. Ces derniers sont nombreux et rendent compte, à différents niveaux, du comportement des allèles au sein de l'échantillon. Les 253 individus de la collection ont fait l'objet d'analyses de leur diversité via l'utilisation des 15 marqueurs microsatellites.

L'utilisation des packages R « adegenet 2.1.0 » (Jombart 2008) et « diveRsity 1.9.90 » (Keenan et al. 2013) a permis de calculer le *nombre d'allèles total révélés au locus étudié* (A), l'*hétérozygotie attendue* (H_e) calculée sous les conditions d'Hardy-Weinberg [explique la relation entre les fréquences génotypiques et les fréquences alléliques en panmixie, pour un organisme diploïde qui se reproduit sexuellement, avec des générations non-chevauchantes, pour une taille de population large, avec des forces évolutives comme la migration et la sélection négligeables], l'*hétérozygotie observée* (H_o) qui peut donc être inférieure ou supérieure, le *coefficient de consanguinité* (F_{IS}) et l'*indice de fixation* (F_{ST}), paramètres de Wright donnant respectivement une indication de l'écart à la panmixie et du degré de différenciation des sous-populations, où $F_{IS} = (H_e-H_o)/H_e$.

Ensuite, le *nombre d'allèles effectifs par locus* (A_E), qui est le nombre d'allèles également fréquents qui seraient nécessaires pour obtenir la même hétérozygotie attendue que dans la population étudiée, a été calculé grâce au module complémentaire Excel « GenAlEx 6.5 » (Peakall and Smouse 2006; Peakall and Smouse 2012), disponible à cette adresse : http://biology-assets.anu.edu.au/GenAlEx/Welcome.html.

Aussi, le *Polymorphism Information Content* (PIC) de Botstein, a été calculé grâce au programme « Cervus 3.0 » (Kalinowski et al. 2007), disponible à cette adresse : http://www.fieldgenetics.com/pages/aboutCervus_Using.jsp.

4.2. Analyse de la structure de la population

Les 253 accessions initiales de la collection ont fait l'objet d'analyses de leur structure via l'utilisation des 15 marqueurs microsatellites. La structure d'une population peut d'abord être appréhendée grâce à la réalisation d'une analyse en coordonnées principales (ACoP) (Principal Coordinates Analysis en anglais). Il s'agit d'une ACP mais appliquée à une matrice de dissimilarités. C'est une méthode de visualisation de données qui consiste à transformer

des variables liées entre elles en nouvelles variables décorrélées les unes des autres. Ces nouvelles variables sont les composantes principales. L'ACP et l'ACoP sont proches car permettent de reprensenter des individus dans un espace de faible dimension avec des axes optimaux en termes de variabilité mais l'ACoP s'applique à une matrice de distances euclidiennes, donc de dissimilarités.

Il est possible de calculer cette matrice à l'aide du programme « DARwin 6.0.14 » (Perrier 2006), disponible à cette adresse : http://darwin.cirad.fr/. Pour ce programme, il faut préparer deux fichiers en parallèle : un fichier « .var » qui contient les données de génotypage et un fichier « .don » qui contient des données informatives comme le nom de l'accession et son origine. La matrice de dissimilarités est obtenue en cliquant d'abord sur Dissimilarity → Calculate from allelic data. Ensuite, une transformation « puissance » permet d'obtenir des distances euclidiennes, en cliquant sur Dissimilarity → Transformation et en choisissant Power 0.5. La matrice est ensuite calculée en cliquant sur Factorial Analysis → Analysis. En récupérant les coordonnées selon le nombre d'axes étudiés (souvent 3 axes, x, y et z), il est possible de visualiser l'ACoP en trois dimensions grâce à l'utilisation du package R « scatterplot3d ».

Une autre solution complémentaire pour appréhender la structure d'une population consiste à déterminer le meilleur « K », paramètre représentant le nombre théorique de populations ancestrales desquelles appartiennent les individus. Pour cela, le programme « STRUCTURE 2.3.4 » (Pritchard et al. 2000) a été utilisé, disponible à cette adresse : https://web.stanford.edu/group/pritchardlab/structure.html. Dans ce programme, il faut choisir les paramètres de l'analyse et une des méthodes possibles est de faire l'analyse en deux grandes étapes, puisque la puissance des paramètres choisis influe sur la durée de l'analyse. La première étape consiste à « dégrossir » le meilleur K, en testant K de 1 à 20 par exemple, avec des paramètres statistiques peu puissants mais permettant de faire une analyse relativement rapide ; quelques heures en moyenne (length of burnin period 5000, number of MCMC reps after burnin 50 000). Une fois le K le plus probable déterminé, par exemple 5, il est possible de tester K de 1 à 10 (pour encadrer la valeur), avec des paramètres plus puissants (length of burnin period 100 000, number of MCMC reps after burnin 750 000). L'analyse prend en moyenne deux jours. Le meilleur K est ensuite déterminé en utilisant le programme en ligne « Structure Harvester » (Earl and VonHoldt 2012), disponible à cette adresse : http://taylor0.biology.ucla.edu/structureHarvester/, et la méthode du deltaK (Evanno et al. 2005). Il est alors possible de visualiser les diagrammes en bâtons de l'appartenance des

individus aux K populations ancestrales. Un individu est considéré comme étant « admixed » s'il n'appartient pas à une population à hauteur d'un minimum de 80%.

4.3. Détermination des arbres de distance génétique

Une étude complémentaire de la diversité génétique consiste à réaliser des arbres de distance génétiques des individus via la méthode du neighbor-joining (Saitou and Nei 1987). Ces arbres phylogénétiques permettent alors de visualiser quels individus sont proches génétiquement (distance entre les branches de l'arbre) et ceux qui présentent une grande diversité génétique (longueur de la branche). Pour cela, il faut utiliser à nouveau le programme « DARwin 6.0.14 », en cliquant sur Tree construction → Neighbor Joining.

4.4. Cartographie génétique et détection de QTL

Grâce à la descendance F_1 disponible, il a été possible de réaliser une cartographie génétique grâce à l'utilisation de la puce de génotypage à haute densité de 700K SNP. Un tri des « aliens » a d'abord été réalisé via les données SSR et SNP. Un tri des SNP est ensuite réalisé grâce au programme « PLINK 1.9 » (Purcell et al. 2008), disponible à cette adresse : https://www.cog-genomics.org/plink/1.9/. Ce tri permet d'exclure notamment les SNP qui ne ségrègent pas de façon mendélienne et ceux qui possèdent moins de 90% de taux de génotypage. Ensuite, la stratégie du « pseudo-testcross » est utilisée (Grattapaglia and Sederoff 1994).

Au final, 849 SNP ont été retenus pour la création de la carte parentale femelle 'Franquette' et 1 088 pour la carte parentale mâle 'UK 6-2'. Les cartes génétiques ont été construites grâce au programme « JoinMap® 4.0 » (Van Ooijen 2006) après retrait des marqueurs distordus. Un LOD minimal de 16 a été choisi pour l'assignement dans les groupes de liaison et la fonction de cartographie de Kosambi (Kosambi 1944) a été retenue pour convertir les fréquences de recombinaison en distances génétiques. La représentation graphique des groupes de liaison a été réalisée grâce au programme « MapChart 2.32 » (Voorips 2002).

Les QTL ont été déterminés grâce au programme « MultiQTL 2.6 », disponible à cette adresse : http://www.multiqtl.com/. La méthode du Multiple Interval Mapping (MIM) a été retenue (Kao et al. 1999) et tous les groupes de liaison ont été scannés en utilisant le modèle « one QTL, 1,000 permutation runs », comparant l'hypothèse H1 « un QTL est présent dans le groupe de liaison » contre l'hypothèse H0 « pas de QTL dans le groupe de liaison ». Le

seuil pour le MIM est de 5% et le seuil de risque de faire une erreur de type 1 pour chaque QTL a été calculé comme suit :

$$\alpha_{chr} = 1 - \{1 - [1 - (1 - 0.05)^{1/M}\}^{m}$$

où M est le nombre total de marqueurs sur chaque carte parentale et m est le nombre de marqueurs sur le groupe de liaison du QTL (Saintagne 2004).

4.5. Analyse en génétique d'association

Tout d'abord, la génétique d'association étant relativement complexe à mettre en œuvre, ce qui suit est un résumé des méthodes utilisées. Il est nécessaire de se référer aux « Matériels & Méthodes » des articles relatifs aux résultats de GWAS d'une part et à l'introduction du manuscrit d'autre part, pour comprendre tous les aspects de cette analyse.

Comme pour le choix des marqueurs utilisés en cartographie, un tri des SNP s'impose avant de réaliser la GWAS sur les 170 individus retenus de la collection ; il s'agit du contrôle qualité des SNP. De façon identique à la démarche retenue pour la cartographie, sont conservés les SNP qui ségrègent de façon mendélienne et ceux qui possèdent plus de 90% de taux de génotypage ; mais cette fois, il faut aussi ne conserver que les SNP dont la fréquence de l'allèle mineur est d'au moins 5%. Au final, 364 275 SNP sont retenus pour la GWAS.

La structure du panel de GWAS a été cette fois étudiée grâce aux programme « fastSTRUCTURE » (Raj et al. 2014) et « CLUMPP » (Jakobsson and Rosenberg 2007), qui fonctionnent sur les mêmes modèles statistiques que « STRUCTURE » mais qui sont plus adaptés aux importants jeux de données de génotypage engendrés par les SNP. La représentation graphique des diagrammes en bâtons a été réalisée avec le programme « distruct » (Rosenberg 2004). En complément, une ACP a été conduite grâce au package R « SNPRelate » et le degré d'apparentement entre les individus a été analysé grâce au package R « gdsfmt » (Zheng et al. 2012).

Les niveaux de déséquilibre de liaison ont été appréciés autour de chaque SNP trouvé associé en GWAS grâce au programme « HaploView 4.2 » (Barrett et al. 2005). Ces blocs de DL ont été construits en utilisant la méthode de l'intervalle de confiance (Gabriel et al. 2002) et celle de « solid spine of LD », pour laquelle le premier et le dernier SNP dans un dit bloc sont en DL, avec tous les autres intermédiaires non nécessairement en DL.

Les données phénotypiques utilisées en GWAS ont été obtenues via le calcul des Best Linear Unbiased Predictions (BLUPs), en considérant un effet « année », grâce au package R « lme4 » (Bates et al. 2015). Le modèle global peut s'écrire comme suit :

$$P_{ik} = \mu + Y_i + g_k + e_{ik}$$

où P_{ik} est le phénotype observé de l'accession k durant l'année i, μ est la valeur moyenne du caractère, Y_i est l'effet fixe de l'année i, g_k est l'effet aléatoire de l'accession k et e_{ik} sont les résidus du modèle.

Au préalable, les travaux classiques de statistiques descriptives comme le calcul de matrices de corrélations, ainsi que la visualisation des données, ont été menés grâce aux packages R « tidyverse » (Wickham 2017) et « corrplot » (Wei and Simko 2017). L'héritabilité au sens large pour chaque caractère a été estimée selon la formule générale suivante :

$$H^2 = \sigma^2_G / [(\sigma^2_G + (\sigma^2_\varepsilon / n_{obs/g})]$$

où σ^2_G est la variance de l'effet génotypique, σ^2_ε est la variance des résidus et $n_{obs/g}$ est le nombre d'observations par génotype. Evidemment, des effets ont pu être ajoutés, comme l'effet de l'année ou du site, selon les données utilisées pour effectuer la GWAS.

La GWAS a été réalisée grâce aux packages R issus du programme « GAPIT » (Lipka et al. 2012). Différents modèles linéaires ont été testés comme le MLMM (Segura et al. 2012) et FarmCPU (Liu et al. 2016). Le nombre optimal de composantes principales à inclure dans le modèle (prise en compte de la structure de la population) a été déterminé pour chaque caractère en se basant sur le Bayesian Information Criterion (BIC). Différents seuils de significativité d'une association ont été comparés, comme le False Discovery Rate (FDR) (Benjamini and Hochberg 1995) et la correction de Bonferroni.

Enfin, les annotations de gènes ont été recherchées grâce aux données disponibles relatives à la version 2 du génome de 'Chandler', basée sur le séquençage Oxford Nanopore MinION (https://www.hardwoodgenomics.org/Genomeassembly/2539069) et une annotation de ce génome basée sur le séquençage Pacific Bioscience Isoform (https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Juglans_regia/100/) (Marrano et al., 2020 Preprint).

Chapitre 3 – Résultats et Discussion

1. Caractérisation phénotypique de la collection de ressources génétiques *J. regia* de l'INRAE
   - 1.1. Données phénotypiques chronologiques issues des archives d'Eric Germain de l'INRAE
   - 1.2. Données phénotypiques acquises pendant les 3 années de thèse, de la fleur au fruit
   - 1.3. Caractérisation de la noix en coque via l'utilisation de la microtomographie à rayons X
2. Analyse de la diversité génétique et de la structure de la collection de ressources génétiques *Juglans* de l'INRAE via l'utilisation de marqueurs SSR
3. Etude du déterminisme génétique du type de fructification et des caractères liés à la phénologie
4. Etude du déterminisme génétique des caractères liés au fruit
5. Contribution au management des ressources génétiques du noyer

Chapitre 3 – Résultats et Discussion

1. Caractérisation phénotypique de la collection de ressources génétiques *J. regia* de l'INRAE

1.1. Données phénotypiques chronologiques issues des archives d'Eric Germain de l'INRAE

Eric Germain et Francis Delort ont mené un programme de création variétale sur le noyer de 1977 à 2007 à l'INRAE de Nouvelle-Aquitaine-Bordeaux, dont est issue la collection de ressources génétiques. Ce matériel végétal a donc été caractérisé phénotypiquement pour les besoins du programme. Cependant, ce programme a pris fin en 2007 suite à de nouvelles orientations de l'INRAE, qui a décidé de restreindre le nombre d'espèces modèles étudiées. Jusqu'à maintenant, ces données chronologiques étaient restées archivées. Elles regorgent pourtant d'informations précieuses sur plusieurs caractères étudiés durant la thèse, notamment ceux liés à la phénologie et à la noix en coque. Afin de partager ces données avec la communauté scientifique et en écho au protocole de Nagoya, celles-ci ont été publiées suivant les principes FAIR (Findable, Accessible, Interoperable, Reusable). Ainsi, une question peut se poser dans le cadre de ce travail :

Les données d'archives d'Eric Germain sont-elles exploitables pour réaliser une analyse en génétique d'association ?

Ces données phénotypiques, accompagnées des données climatiques, concernent 150 accessions de la collection et sur trois sites différents : l'UEA de Toulenne et les deux stations expérimentales de la noix, Creysse et SENuRA. Elles sont disponibles publiquement sur le portail d'entrepôt de données de l'INRAE, « Portail Data INRAE », de l'Unité de Recherche Génomique Info (URGI Plant and Fungi Dataverse). Ces données ont notamment permis de construire la figure 5 du chapitre 1, qui met en évidence une tendance à l'avancée de la date de débourrement du bourgeon végétatif et de l'infloresence femelle entre 1989 et 2017 des variétés 'Lara' et 'Franquette', habituellement utilisées en expérimentation comme témoins en France. Ces résultats ont été présentés au XXX International Horticultural à Istanbul en Turquie, du 12 au 16 août 2018, sous forme de poster. Le « Full Article » lié au poster se trouve en Annexe 3.

Ces données sont accompagnées par l'ontologie utilisée pour acquérir ces données. Leur observation est principalement basée sur les recommandations des standards de Bioversity International (anciennement IPGRI), donc Eric Germain en a été un contributeur (IPGRI 1994), toutefois certains caractères ont été observés spécifiquement pour les besoins du programme de création variétale qu'il conduisait.

La « data note » suivante a été publié en ligne le 17 octobre 2019 dans le journal BMC Research Notes, volume 12:662, sous le nom de « The walnut genetic resources of INRA: chronological data and ontology » (DOI 10.1186/s13104-019-4678-1). Elle donne l'accès au fichier de phénotypage contruit en format « comma-separated values ».

Bernard *et al. BMC Res Notes* (2019) 12:662
https://doi.org/10.1186/s13104-019-4678-1

BMC Research Notes

DATA NOTE Open Access

The walnut genetic resources of INRA: chronological phenotypic data and ontology

Anthony Bernard[1,4], Teresa Barreneche[1], Marine Delmas[2], Sophie Durand[3], Cyril Pommier[3], Fabrice Lheureux[4], Eloïse Tranchand[5], Marianne Naudin[6] and Elisabeth Dirlewanger[1*]

Abstract

Objectives: Persian walnut (*Juglans regia* L.), the walnut species cultivated for nut production, is grown worldwide in temperate areas. In this work, chronological phenotypic data have been collected regarding a part of the walnut genetic resources of the French National Institute for Agricultural Research (INRA) of Bordeaux. Using a well described ontology, these data have been collected in order to assess the phenotypic variations among the accessions, and to better manage the germplasm collection. These data can also be helpful for any breeding program as they provide a clear phenotypic characterization of the main cultivars.

Data description: This paper introduces a dataset collected for 150 *J. regia* accessions for a period from 1965 to 2016, and for 3 observation sites, released as comma separated value spreadsheet. It includes observations about phenological traits (e.g. flowering dates), traits related to in-shell walnut (e.g. weight and size), and traits related to kernel (e.g. color). It can be used by other researchers particularly for multi-site phenological studies in the context of climate change since climate data files are also available. In addition, a complete walnut ontology was deposited in this repository and can assist to standardize the management of any walnut germplasm collection.

Keywords: Persian walnut, *Juglans regia* L., Phenotyping, Phenology, Ontology, Genetic resources, Germplasm

Objective

Genetic resources constitute an essential reservoir of allelic diversity for traits currently being used and those not yet used in breeding programs, their management is crucial [1]. In France, Eric Germain and Francis Delort from the Institut National de la Recherche Agronomique (INRA—French National Institute for Agricultural Research) led a walnut breeding program from 1977 to 2005 [2]. The diversity of INRA's walnut genetic resources with regards to geographical origin is a result of international cooperation established by Eric Germain. He travelled widely and collected diverse plant materials maintained by the *Prunus*/*Juglans* Genetic Resources Center. These materials have been phenotypically characterized for the needs of his breeding program. Nevertheless, it ended in 2007 due to a new orientation for INRA, following a decision to focus on a smaller number of model species, such as *Prunus* and *Malus*.

Until now, these chronological phenotypic data have remained archived. However, they contain valuable information about traits related to phenology, in-shell nut and kernel of walnut cultivars with worldwide origins. Furthermore, a public ontology has been used and adapted during the breeding program. Regarding the context of global competition and climate change, a new French research project called INNOV'noyer started in 2017 in cooperation between the Centre Technique Interprofessionnel des Fruits et Légumes (Ctifl—Fruit and Vegetable Interprofessional Technical Centre) and the INRA Bordeaux-Aquitaine. The objective being to start a new breeding program using marker-assisted selection, and to provide a larger choice of walnut cultivars better adapted to this context. In this way, these data represent a valuable aid for the project and for any researcher who wants to obtain information regarding phenology of a major tree crop species [3]. To our knowledge, these data also provide the most extensive dataset available publicly

*Correspondence: elisabeth.dirlewanger@inra.fr
[1] UMR 1332 BFP, INRA, Université de Bordeaux, 33140 Villenave d'Ornon, France
Full list of author information is available at the end of the article

© The Author(s) 2019. This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

particularly for the phenology in a walnut germplasm collection, and a complete walnut ontology is reported.

Data description

Here we present a raw dataset of chronological phenotypic observations of 150 *J. regia* accessions named Ephesis_export (Table 1, Data file 1) publicly available in the "Portail Data INRA" repository. All accessions are grafted trees located in the arboriculture experimental unit in Toulenne (*Prunus*/*Juglans* Genetic Resources Center) near Bordeaux (latitude 44°34′37.442″N–longitude 0°16′51.48″O) and a part of them are also located in the two French walnut experimental stations of Creysse (latitude 44°53′16.529″N–longitude 1°36′23.838″E) and SENuRA (latitude 45°8′38.565″N–longitude 5°17′34.323″E).

Toulenne and Creysse locations in southwestern France are both characterized by an oceanic climate and a sandy-silt soil. In contrast, SENuRA location in southeastern France, also with a sandy-silt soil, is in a continental climate. Both Creysse and SENuRA locations belong to the main walnut production areas in France. The 150 *J. regia* accessions represent the maximum genetic diversity of the available collection [4].

This dataset includes mainly observations from Toulenne (50%) and other from Creysse and SENuRA particularly for control cultivars. In the first row, the header shows the accession number (e.g. "RA 0311"), the accession name (e.g. "Franquette"), the trial site (e.g. "Toulenne"), the campaign (e.g. "1991"), and the phenotyped traits (e.g. "BUD_Est_Jd: Date of bud break"). The dataset contains information particularly about traits related to phenology (e.g. Date of bud break, female/male bloom dates), traits related to in-shell walnut (e.g. weight of 100 dry walnuts, shell thickness) and those related to kernel (e.g. weight of 100 kernels, percentage of light colored kernels).

For the range of observation dates, climate data recorded by on-site weather stations are also available in the "Portail Data INRA" repository for each site: Creysse from 1992 to 2017 (Table 1, Data file 2), SENuRA from 1988 to 2016 (Table 1, Data file 3) and Toulenne from 1980 to 2011 (Table 1, Data file 4). These climate data files show daily (except for SENuRA site which gives climate data monthly) the minimum, the mean and the maximum temperature in degrees Celsius. They also include the rainfall level expressed in millimeters.

Phenotype measurements have been performed using a well described ontology based on already existing methodologies [5] and INRA internal protocols. This file named WATO-Walnut-Trait-Ontology (Table 1, Data file 5) is also publicly available in "Portail Data INRA" repository. In the first row, the header shows the variable name and synonyms (e.g. "BUD_Est_Jd", "Date of bud break"), the reference of the methodologies and the institution (e.g. "Descriptors for walnuts, IPGRI, 1994, 6.1.2", "INRA"), the trait description (e.g. "when over 50% of terminal buds have enlarged and bud scales have split exposing the green of the leaves inside") and the scale name and class ("Julian days of the year of evaluation", "Time").

Both chronological phenotypic data and ontology could be used by any researcher wishing to obtain phenotypic description of a particular walnut cultivar. It could be also used to conduct a phenotyping campaign using descriptors recognized within the community working on *J. regia*. Finally, it could be useful to assess the effect of climate change on phenology for tree crop species.

Limitations

Although these observations represent a complete database concerning a large number of accessions which have been phenotyped on more than 30 traits, they contain

Table 1 Overview of data files

Label	Name of data file	File types (file extension)	Data repository and identifier (DOI)
Data file 1	Ephesis_export.csv	Comma separated values file (.csv)	Portail Data INRA repository (URGI Plant and Fungi Dataverse, INRA's Walnut Genetic Resources Observation) (https://doi.org/10.15454/58SYJV) [6]
Data file 2	Climate data Creysse from 1992 to 2017.tab	Tabular data file (.tab)	Portail Data INRA repository (URGI Plant and Fungi Dataverse, INRA's Walnut Genetic Resources Observation) (https://doi.org/10.15454/58SYJV)
Data file 3	Climate data SENuRA from 1988 to 2016.tab	Tabular data file (.tab)	Portail Data INRA repository (URGI Plant and Fungi Dataverse, INRA's Walnut Genetic Resources Observation) (https://doi.org/10.15454/58SYJV)
Data file 4	Climate data Toulenne from 1980 to 2011.tab	Tabular data file (.tab)	Portail Data INRA repository (URGI Plant and Fungi Dataverse, INRA's Walnut Genetic Resources Observation) (https://doi.org/10.15454/58SYJV)
Data file 5	WATO-Walnut-Trait-Ontology.csv	Comma separated values file (.csv)	Portail Data INRA repository (LovINRA, Walnut Trait Ontology) (https://doi.org/10.15454/AV5RT2) [7]

missing data left blank as many phenotyping data on tree crop species. Those missing data are notably due to particular climatic conditions some years. Then, regarding climate data from the SENuRA observation site, they are not as accurate as the other two because temperature and rainfall data are not given daily but with monthly average. There is also an imbalance in the amount of data for 'Franquette' and 'Lara' accessions. Indeed, these two cultivars are usually used as controls for experimentation in France. In that respect, more data are available for controls. Finally, the dates of observation ("Campaign" column in Table 1, Data file 1) are heterogeneous. When an accession has been introduced in orchards at the *Prunus*/*Juglans* Genetic Resources Center in year n, it has been usually observed in years $n+1$, $n+2$ and $n+3$. However, the year n of introduction can vary a lot due to the long process of germplasm establishment.

The "Portail Data INRA" repository is mainly a landing page which hosts the entire dataset. But for more details, readers have the possibility to manage the data with a pleasant interface (e.g. to choose a particular accession) using the "link to data" available on the page. This link brings to the URGI Plant and Fungi platform hosted by the Unité de Recherche Génomique Info (URGI) which is a research unit in genomics and bioinformatics at INRA, dedicated to plants and crop parasites. This platform also provides the passport descriptors such as the country of origin or the pedigree for each accession.

Abbreviations

ANRT: Agence Nationale de la Recherche et de la Technologie; Ctifl: Centre Technique Interprofessionnel des Fruits et Légumes; DOI: Digital Object Identifier; INRA: Institut National de la Recherche Agronomique; SENuRA: Station d'Expérimentation Nucicole en Rhône-Alpes; URGI: Unité de Recherche Génomique Info.

Acknowledgements

First, our special thanks are addressed to Eric Germain and Francis Delort, former heads of the walnut breeding program at INRA of Bordeaux. We also gratefully thank the entire staff of the arboriculture experimental unit of Toulenne and that of the two French walnut experimental stations of Creysse and SENuRA for their work during many years to characterize the accessions. Finally, we thank Peter Prince (Lab technician, Ctifl) for English language revision of Objective part.

Authors' contributions

AB was in charge of the digitization of the data previously stored in archives. AB sorted the data and wrote the manuscript. TB was, and MD is in charge of the genetic resources management and they were really helpful to check the passport information. SD and CP checked and published the data in the URGI Plant and Fungi Dataverse, and the walnut ontology in LovINRA, respectively. ET and MN provided phenotypic data from the two French walnut experimental stations from Creysse and SENuRA, respectively. FL and ED, respectively the initiator of the project INNOV'noyer and the thesis supervisor of AB designed the data management. All authors reviewed the manuscript. The data integration into the repository was performed using the facilities of the URGI platform (https://urgi.versailles.inra.fr/). All authors read and approved the final manuscript.

Funding

The Ctifl, holder of the project "INNOV'noyer", in partnership with INRA, wants to thank the "Région Nouvelle-Aquitaine" for funding a part of this project and the "Cifre" convention by "ANRT" (Agence Nationale de la Recherche et de la Technologie" for funding half the cost of the doctoral fellowship of Anthony Bernard. Both INRA and Ctifl, in partnership with the two French walnut experimental stations of Creysse and SENuRA, were in charge of the collection of data. Both INRA and Ctifl contributed to the writing process of this manuscript.

Availability of data materials

The phenotypic data described in this Data Note can be freely and openly accessed on the "Portail Data INRA" repository via the identifier INRA's Walnut Genetic Resources Observation [6] and the following Digital Object Identifier (DOI): https://doi.org/10.15454/58SYJV. The walnut ontology is available on the same repository via the identifier Walnut Trait Ontology [7] and the following DOI: https://doi.org/10.15454/AV5RT2. Please see Table 1 and reference list for details and links to the data.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1] UMR 1332 BFP, INRA, Université de Bordeaux, 33140 Villenave d'Ornon, France. [2] UEA 0393, INRA, 33210 Langon, France. [3] URGI, INRA, Université Paris-Saclay, 78026 Versailles, France. [4] Ctifl, centre opérationnel de Lanxade, 24130 Prigonrieux, France. [5] Station expérimentale de la noix de Creysse, 46600 Creysse, France. [6] Station d'expérimentation nucicole en Rhône-Alpes, 38160 Chatte, France.

Received: 29 November 2018 Accepted: 28 September 2019
Published online: 17 October 2019

References

1. National Research Council. Managing global genetic resources: agricultural crop issues and policies. Washington, D.C: The National Academies Press; 1993.
2. Bernard A, Lheureux F, Dirlewanger E. Walnut: past and future of genetic improvement. Tree Genet Genomes. 2018;14:1.
3. Fitchett JM, Grab SW, Thompson DI. Plant phenology and climate change: progress in methodological approaches and application. Progr Phys Geogr Earth Environ. 2015;39(4):460–82.
4. Bernard A, Barreneche T, Lheureux F, Dirlewanger E. Analysis of genetic diversity and structure in a worldwide walnut (*Juglans regia* L.) germplasm using SSR markers. PLoS ONE. 2018;13(11):e0208021.
5. IPGRI. Descriptors for walnut (*Juglans* spp.). Rome: International Plant Genetic Resources Institute; 1994.
6. Bernard A, Dirlewanger E, Delmas M, Barreneche T. INRA's walnut genetic resources observations. Portail Data INRA. 2018. https://doi.org/10.15454/58SYJV.
7. Bernard A, Germain E. Walnut trait ontology. Portail Data INRA. 2018. https://doi.org/10.15454/AV5RT2.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1.2. Données phénotypiques acquises pendant les 3 années de thèse, de la fleur au fruit

Si elles représentent un large jeu de données concernant les accessions étudiées, les données chronologiques issues des archives d'Eric Germain possèdent un certain nombre de limitations pour réaliser des analyses en génétique d'association. Elles contiennent d'abord de nombreuses données manquantes probablement à cause des conditions climatiques de certaines années. Mais la principale limitation réside dans leur caractère hétérogène. En effet, les accessions ont d'abord été introduites en verger au fur et à mesure qu'Eric Germain les collectait ; bien souvent grâce à l'aide d'autres sélectionneurs du monde entier ou par le biais de ses propres prospections. Lorsqu'une accession a été introduite une année « n » en verger, celle-ci a bien souvent été observée les années « n+1 », « n+2 » et « n+3 ». Néanmoins, cette année « n » d'introduction peut varier considérablement en raison du long processus nécessaire à la création d'une collection. L'objectif premier d'Eric Germain était l'évaluation de matériel végétal pour l'amélioration donc il n'avait pas besoin d'une collection du même âge. Le Centre de Ressources Biologiques *Prunus-Juglans* a, quant à lui, été créé plus tard. Cette constatation a mené à la conclusion suivante :

Il est nécessaire de réaliser de nouvelles campagnes de phénotypage, de la fleur au fruit, afin d'acquérir de nouvelles données concernant les principaux caractères d'intérêt agronomique.

Ainsi, pour réaliser les analyses en génétique d'association sur des données phénotypiques plus homogènes et contemporaines, un nouveau phénotypage des caractères liés à la phénologie et au fruit a été réalisé. La variabilité phénotypique de la collection, basée sur les observations effectuées durant les 3 années de thèse, est appréhendée par des nuages de points, des diagrammes de densité ainsi que des matrices de corrélation disponibles dans les données supplémentaires des différents articles publiés : chapitre 3, paragraphe 3 pour les caractères liés à la phénologie, puis chapitre 3, paragraphe 4 pour les caractères liés au fruit. Afin d'apporter un complément et une vision globale des relations entre tous les caractères observés de la fleur au fruit, une ACP a été réalisée sur l'ensemble des données obtenues pendant les 3 années de thèse. Ces données résultent de l'évaluation des 170 individus *J. regia* de la collection qui ont été retenus pour la GWAS.

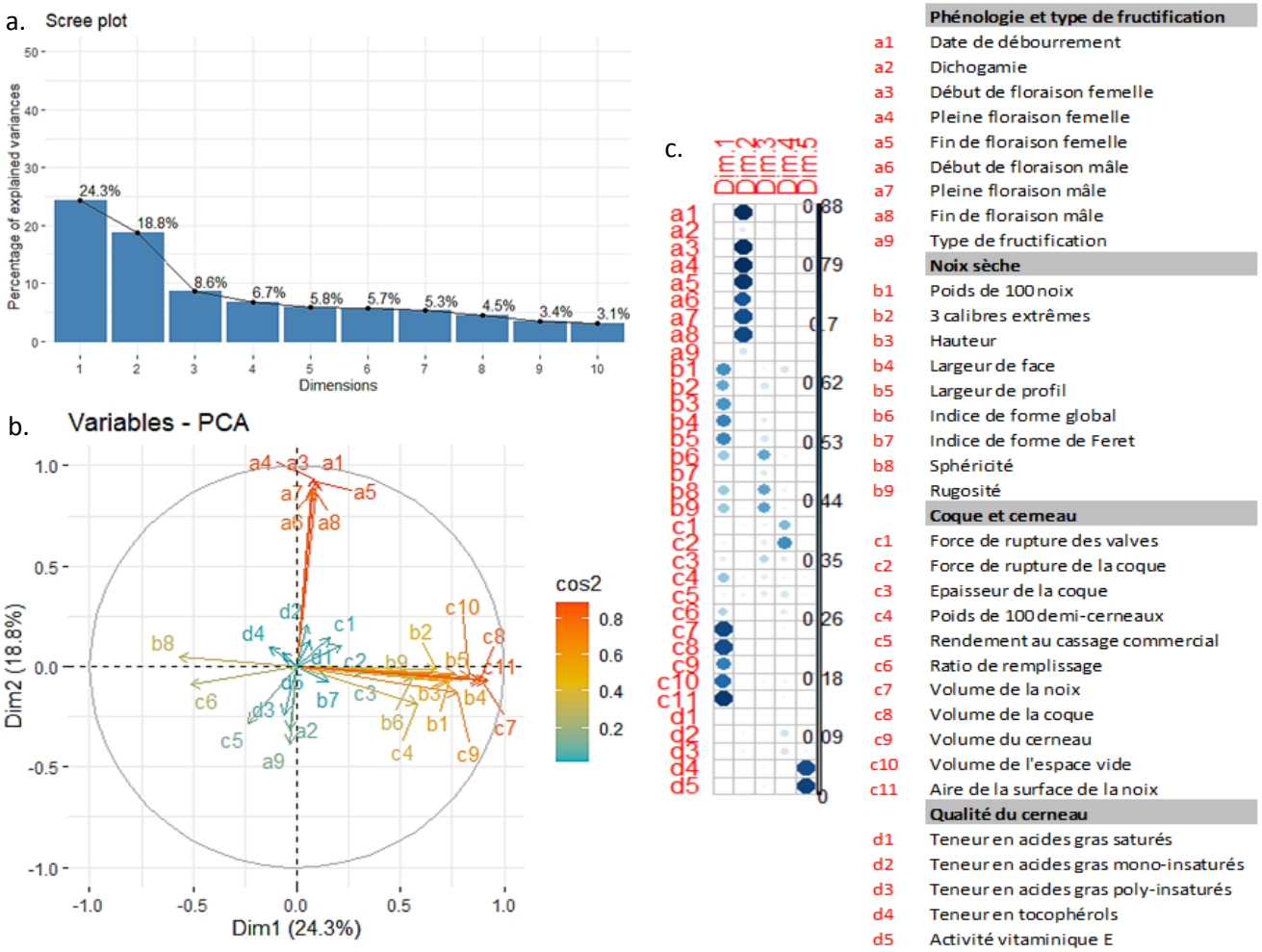


Figure 20 : Résultats de l'ACP réalisée sur l'ensemble des données de phénotypage acquises pendant la thèse.

La figure 20 résulte de l'analyse des données des 34 variables décrites dans le chapitre 2 sur les 170 individus phénotypés. La première étape de l'ACP consiste en la standardisation des données puisqu'au départ, les variables sont mesurées et exprimées dans des unités différentes. Les variables ont été normalisées de façon à ce que l'écart-type de la variable soit égal à 1 et la moyenne nulle. Pour se faire, chaque valeur se voit soustraire la moyenne de la variable et est divisée par l'écart-type. Les données deviennent centrées-réduites et l'ACP est alors normée.

La deuxième étape de l'ACP comprend l'examen de la variance expliquée par les composantes principales (graphique « scree plot », figure 20a). Il est classique de voir que les premiers axes de l'ACP, ou dimensions, expliquent la plus grande proportion de la variance du jeu de données. La somme des proportions est égale à 100% si l'on considère toutes les dimensions existantes. Ici, les 10 premières dimensions expliquent 86,2% de la variance

totale. Une méthode communément retrouvée pour déterminer le nombre de dimensions à retenir est de vérifier que la proportion de variance expliquée par la dimension dépasse les 10%. Lors de l'étude des 34 variables au sein des 170 accessions sélectionnées, les dimensions 1 et 2 expliquent respectivement 24,3% et 18,8% de la variance totale (cumul 43,1%). Une autre méthode consiste à retenir le nombre de dimensions qui retient une proportion satisfaisante de la variance totale. Ainsi, le cumul des 5 premières dimensions dans ce cas atteint 64,2% de la variance totale ; ce qui semble satisfaisant afin de ne pas perdre trop d'informations.

La troisième étape de l'ACP est la représentation graphique des variables en fonction des corrélations entre ces variables et de la qualité de leur représentation. Le graphique des variables se présente donc sous la forme d'un cercle de corrélation (graphique « variables – PCA », figure 20b). La qualité de représentation est traduite par leurs coordonnées dans le plan au carré, le cosinus carré (cos^2). Un cos^2 élevé indique une part expliquée élevée de la variable sur les dimensions considérées. La variable sera donc positionnée loin du centre du cercle. Il est présenté en échelle colorée du cercle des corrélations et par la matrice de cos^2, pour les 5 premières dimensions (figure 20c). L'analyse sur les dimensions 1 et 2 des caractères étudiés révèle plusieurs résultats intéressants :

- La dimension 1, expliquant 24,3% de la variance totale, est principalement définie par les variables liées à la noix sèche de b1 à b9 (à l'exception de b7) et par certaines variables liées à la coque et au cerneau avec c4, c7, c8, c9, c10 et c11. Les variables de c7 à c11 (volumes des différentes parties de la noix et aire de la surface) sont particulièrement bien représentées car proches du cercle, tandis que les variables de b1 à b9 (poids, calibre, dimensions, sphéricité, rugosité et dimensions de la noix) sont moins bien représentées. L'angle plutôt fermé que forme toutes ces variables montre qu'elles sont corrélées positivement entre elles. En revanche, les variables b8 (sphéricité) et c6 (ratio de remplissage) sont corrélées négativement avec les précédentes. Ainsi, les accessions produisant des noix ayant un plus grand volume et poids auront tendance à être moins sphériques et à avoir un ratio de remplissage inférieur.
- La dimension 2, expliquant 18,8% de la variance totale, est principalement définie par les variables liées à la phénologie et au type de fructification (a1 à a9). Les dates de débourrement et de floraison sont bien représentées, tandis que la dichogamie (a2) et le type de fructification (a9) sont des variables moins bien

représentées. L'angle quasiment fermé formé par les dates de débourrement et de floraison montre qu'elles sont positivement corrélées entre elles. En revanche, la dichogamie et le type de fructification sont des variables corrélées négativement avec les précédentes. Ainsi, les accessions tardives auront tendance à avoir un type de fructification en position terminale.

- Les variables expliquées d'une part par la dimension 1 (fruit) et celles expliquées d'autre part par la dimension 2 (fleur) définissent globalement un angle droit. Ainsi, il n'y a pas nécessairement de lien entre la phénologie et les caractéristiques du fruit. Il est donc possible de trouver des accessions précoces produisant des petits ou des gros fruits. Inversement, les accessions produisant des gros fruits peuvent être aussi bien précoces que tardives.
- Les autres dimensions sont constituées de variables globalement moins bien représentées. La dimension 3 abrite principalement l'autre part de la variance des variables de la noix, sphéricité (b8) et rugosité (b9), mais surtout l'indice global de forme. La dimension 4 est constituée des variables force de rupture des valves (c1) et de la coque (c2). Enfin, la dimension 5 décrit surtout les variables teneur en tocophérols (d4) et activité vitaminique E (d5). Les variables liées à la qualité du cerneau, comme les teneurs en les différents acides gras (d1 à d3), n'ont pas pu être bien représentées dans cette ACP.

L'ACP sur les deux premières dimensions montre ainsi qu'il n'est peut-être pas nécessaire d'observer tous les caractères à l'avenir, dans le cadre par exemple de la gestion de collections de ressources génétiques ou de caractérisation de géniteurs pour l'amélioration. En ce qui concerne la phénologie, les dates de débourrement et de début de floraison mâle et femelle peuvent être suffisantes pour décrire la variabilité globale de la phénologie de la collection. Le même constat peut être fait pour certains des caractères liés au fruit. Ce choix permettrait d'économiser du temps de phénotypage passé en verger et à la paillasse.

La quatrième étape de l'ACP est la représentation graphique des individus. Puisque les dimensions 1 et 2 abritent l'essentiel de l'inertie du jeu de données et que ces dimensions représentent mal les variables liés à la qualité du cerneau (d1 à d5), celles-ci ont été écartées pour créer le nuage de points des individus. Ce nuage de points se trouve en figure 21 et représente les données de l'année 2018, pour laquelle l'ensemble des caractères de la fleur et du fruit sont disponibles (voir tableau 1).

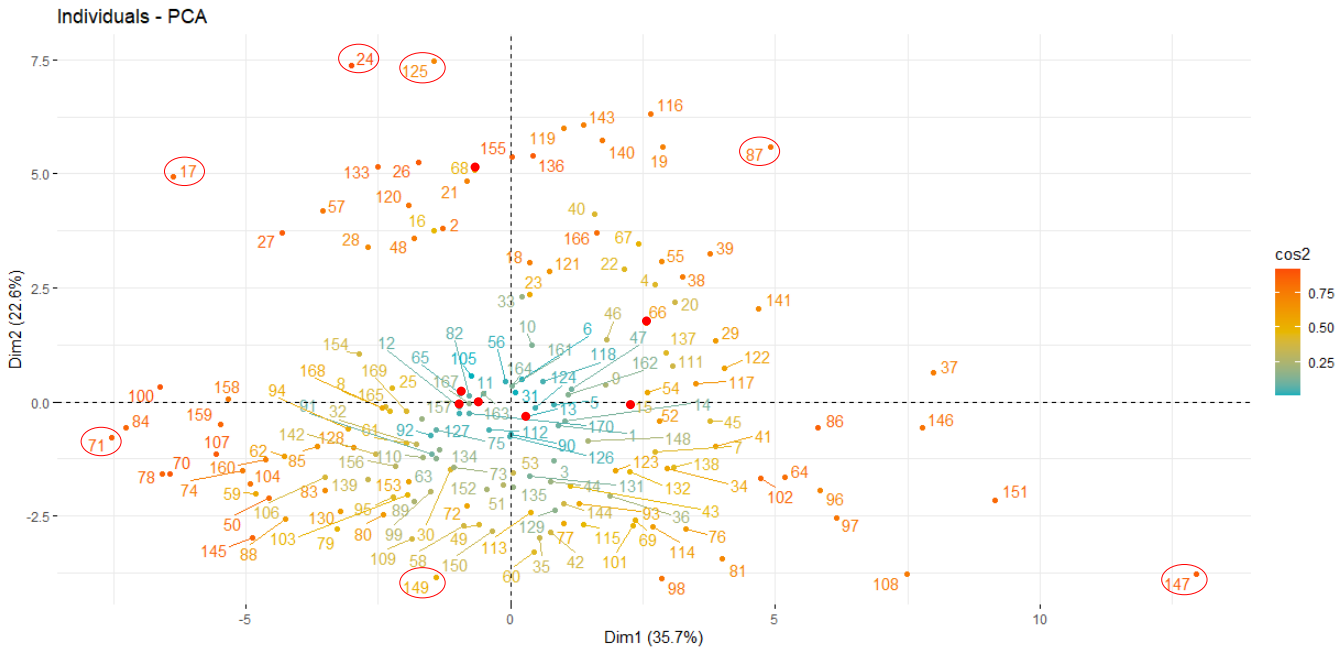


Figure 21 : Graphique des individus de l'ACP sur les données de 2018.

Comme pour les variables, les individus ont été colorés selon leur cos². Ce graphique doit se lire dans les mêmes « directions » que celui des variables. Ainsi, les individus dans la partie supérieure suivant la dimension 2 sont les accessions présentant préférentiellement une phénologie tardive (valeurs positives), tandis que ceux de la partie inférieure sont les accessions présentant préférentiellement une phénologie précoce (valeurs négatives). De la même façon, si l'on considère la dimension 1, les individus dans la partie de droite correspondent aux accessions produisant des fruits plus volumineux. En 2018, 'Saint Martial' (24) et 'Le Bordelais' (125) sont les accessions les plus tardives, tandis que 'Early Ehrhardt' (149) est l'accession la plus précoce. 'Carmelo' (147) possède le fruit le plus volumineux contrairement à 'UK 6-2' (71) dont le fruit est le moins volumineux. L'intérêt de ce graphique est de pouvoir, par exemple, trouver rapidement deux accessions tardives ayant respectivement des petits fruits et des gros fruits. Ainsi, 'Lozeronne' (17) et 'Hybride H 188-13' (87) sont tous les deux tardifs mais le premier possède des petits fruits, tandis que le second porte des gros fruits.

La position des 7 variétés hybrides françaises issues du programme d'amélioration de l'INRAE dirigé par E. Germain est très intéressante. Elles sont représentés par un point rouge épais sur la figure 22. 'Fernette' (15), 'Fernor' (11), 'Ferjean' (12) inscrits entre 1995 et 1999 se trouvent au centre de l'ACP. Ces hybrides ont été sélectionnés d'abord pour leur capacité à fructifier de façon latérale mais leurs caractéristiques liées à la phénologie et au fruit restent classiques. Idem pour 'Feradam' (13) et 'Ferbel' (65) inscrits en 2010. La variété 'Fertignac'

(68) a été créée pour devenir un pollinisateur tardif en verger donc il est logique qu'elle se retrouve proche des accessions tardives. En revanche, 'Ferouette' (66) semble intéressante. Cette variété est sensiblement tardive et possède des fruits relativement gros. Mais ces résultats montrent que ces nouvelles variétés ne répondent pas pleinement aux attentes de la profession et qu'il est nécessaire de proposer une gamme plus large.

En conclusion, l'ACP représente une étape essentielle pour appréhender la variabilité disponible dans toute collection de ressources génétiques. Elle est complémentaire des études de structure et de diversité, tous présentés dans les chapitres suivants. C'est un des outils incontournables contribuant à la définition de la meilleure stratégie de management des ressources génétiques. Elle facilite également le choix de futurs géniteurs pour une étape de pre-breeding.

1.3. Caractérisation de la noix en coque via l'utilisation de la microtomographie à rayons X

Traditionnellement, la caractérisation phénotypique de la noix en coque se pratiquait à l'aide d'une simple règle ou bien grâce à l'utilisation d'un pied à coulisse électronique. Il va de soi que cette méthode présente des limites, tant sur la précision que sur le temps demandé à l'opérateur pour, par exemple, phénotyper des dizaines de noix sur l'ensemble d'une collection de ressources génétiques. La question suivante s'est posée :

Comment améliorer les techniques de phénotypage de la noix en coque ?

Les techniques d'imagerie sont encore peu appliquées chez les plantes mais peuvent apporter de grands avantages, notamment pour la caractérisation des fruits à coque, puisqu'il s'agit d'une technique non destructive. Ainsi, grâce à l'expertise du GEVES et de la plateforme Phenotic, une méthode a été mise au point dans le but de mesurer différents caractères morphologiques sur plusieurs noix simultanément, tels que la longueur de la noix, son diamètre de face ou de profil. Des caractères qui nécessitent habituellement l'ouverture de la noix ont pu également être appréciés, comme l'épaisseur de la coque, le volume du cerneau, ainsi que des caractères qu'il était auparavant difficile à quantifier, tels que la rugosité de la coque ou la sphéricité de la noix.

Ces mesures ont été obtenues à partir d'images en 3D reconstituées et acquises par tomographie à rayons X, à la suite de plusieurs étapes qui constituent un « workflow » : élimination du « bruit de fond », individualisation des noix et extraction des différentes parties de la noix (coque, cerneau et vide). La méthode a été appliquée pour caractériser 50 noix sur 161 accessions de la collection, obtenues durant la récolte de l'année 2018. Cette méthode s'est montrée être adaptée à toutes les tailles et toutes les formes de noix et pourrait être transférable facilement à d'autres fruits à coque.

L'article suivant a été publié en ligne le 26 août 2020 dans le journal Plant Methods, volume 16:115, sous le nom de « 3D characterization of walnut morphological traits using X-ray computed tomography » (DOI 10.1186/s13007-020-00657-7).

Bernard *et al. Plant Methods* (2020) 16:115
https://doi.org/10.1186/s13007-020-00657-7

Plant Methods

RESEARCH Open Access

3D characterization of walnut morphological traits using X-ray computed tomography

Anthony Bernard[1,2†], Sherif Hamdy[3†], Laurence Le Corre[3], Elisabeth Dirlewanger[1*] and Fabrice Lheureux[2]

Abstract

Background: Walnuts are grown worldwide in temperate areas and producers are facing an increasing demand. In a climate change context, the industry also needs cultivars that provide fruits of quality. This quality includes satisfactory filling ratio, thicker shell, ease of cracking, smooth shell and round-shaped walnut, and larger nut size. These desirable traits have been analysed so far using calipers or micrometers, but it takes a lot of time and requires the destruction of the sample. A challenge to take up is to develop an accurate, fast and non-destructive method for quality-related and morphometric trait measurements of walnuts, that are used to characterize new cultivars or collections in any germplasm management process.

Results: In this study, we develop a method to measure different morphological traits on several walnuts simultaneously such as morphometric traits (nut length, nut face and profile diameters), traits that previously required opening the nut (shell thickness, kernel volume and filling kernel/nut ratio) and traits that previously were difficult to quantify (shell rugosity, nut sphericity, nut surface area and nut shape). These measurements were obtained from reconstructed 3D images acquired by X-ray computed tomography (CT). A workflow was created including several steps: noise elimination, walnut individualization, properties extraction and quantification of the different parts of the fruit. This method was applied to characterize 50 walnuts of a part of the INRAE walnut germplasm collection made of 161 unique accessions, obtained from the 2018 harvest. Our results indicate that 50 walnuts are sufficient to phenotype the fruit quality of one accession using X-ray CT and to find correlations between the morphometric traits. Our imaging workflow is suitable for any walnut size or shape and provides new and more accurate measurements.

Conclusions: The fast and accurate measurement of quantitative traits is of utmost importance to conduct quantitative genetic analyses or cultivar characterization. Our imaging workflow is well adapted for accurate phenotypic characterization of a various range of traits and could be easily applied to other important nut crops.

Keywords: Walnut, Germplasm collection, Morphological traits, X-ray computed tomography, 3D characterization, Image analysis

Background

Persian walnut (*Juglans regia* L.), the walnut species cultivated for nut production, is one of the oldest food sources known and is grown worldwide [1]. According to the Food and Agriculture Organization of the United Nations (www.fao.org, 2017 data), worldwide in-shell walnut production exceeds 3.8 M tons. The three largest producers are China, USA and Iran. France is the 9th largest producer, the 2nd in Europe, with 40,000 tons. French walnut orchard area reached approximately 21,000 hectares in 2017 (https://agreste.agriculture.gouv.fr/), making it the most important French fruit crop other than apple. The production is mainly exported in-shell to Europe thanks to its high quality, especially due to the well-known Protected Designations of Origin 'Noix du

*Correspondence: elisabeth.dirlewanger@inrae.fr
†Anthony Bernard and Sherif Hamdy contributed equally to this work
[1] Univ. Bordeaux, INRAE, Biologie du Fruit et Pathologie, UMR 1332, 33140 Villenave d'Ornon, France
Full list of author information is available at the end of the article

© The Author(s) 2020. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Périgord' and 'Noix de Grenoble', the two main walnut production areas. However, the number of cultivars is low and the French walnut industry needs new cultivars well adapted to French climatic conditions, with high nut and kernel quality. The quality traits include round-shaped walnut, smooth and shell easy to crack, larger nut size and high nut/kernel weight ratio [2].

Nowadays, agriculture is facing challenges for crop production and the plant research community needs to perform quantitative analyses of numerous plant traits in order to accelerate progress in breeding [3, 4]. This is why crop germplasm collections are of tremendous importance since no production area is fully self-sufficient in genetic diversity to cover all producer and consumer demand [5]. A germplasm collection is ordinarily evaluated with morphological descriptors which is usually the first step for describing accessions and selecting genitors for breeding programs [6, 7]. For obvious cost reasons, the International Plant Genetic Research Institute (IPGRI, now called Bioversity International) gives ontologies to manage germplasm collections based on morphological measurements, such as fruit length, diameter or thickness in millimeter without any measuring tools mentioned, and visual appreciation for traits that are difficult to quantify such as fruit shape or rugosity.

In walnut, since mid-1970s numerous studies report correlations between various traits related to morphological properties of the nut [8–13]. For instance, to evaluate promising genotypes mainly originated from seed in Iran, shell thickness was measured with a micrometer and shell roughness was recorded by assigning values from 1 to 7 [14, 15] or from 1 to 9 [16, 17] based on visual appreciation recommended by IPGRI [18]. Other works in Iran, using a caliper, also focused on nut length and diameter [19–23]. In Turkey, similar studies were conducted still using a caliper or even a compass to measure walnut diameter, length and/or shell thickness [24–28]. Finally, walnut germplasm collections from Europe were also characterized using similar tools in Albania [29], Serbia [30], Bulgaria [31], Romania [32] and Italy [33].

Overall, the measurements using a caliper, a micrometer or even by simple visual observation are until now the classical methods in walnut assessment. However, this kind of evaluation is painstaking, time-consuming and can lead to inaccuracy and low resolution. Fortunately, imaging techniques applied for plant phenotyping and food quality determination have been developed over the past decades using a wide range of methodologies, mainly for field crops and for various traits such as growth dynamics, shoot structure and morphometric parameters [34]. We can quote the use of visible light [35] and fluorescence imaging [36] on barley; the thermal infrared [37], near-infrared [38] and hyperspectral imaging [39] on rice; or 3D imaging on soybean [40] and magnetic resonance imaging on bean [41]. With the genomics era allowing researchers to unravel the genetic architecture of complex traits, we clearly observe a shift to a need in high-throughput phenotyping for crop improvement [42].

X-ray computed tomography (CT) is a non-destructive imaging technique based on computer-processed X-rays to acquire tomographic slice images of the scanned sample and generate a 3D reconstruction [34]. Used at first for medical purposes, X-ray CT has been lately applied in various agricultural products in particular to evaluate internal quality [43], especially in fruits and vegetables. In apples, X-ray CT was used to evaluate the density and the water content under varying moisture conditions [44], while in pears this technology was used to study the core breakdown development [45]. This method was also used in nuts and few works are reported, such as the detection of pinhole insect damage in almonds [46], the segmentation and classification in hazelnuts [47] and the behaviour study of fourth-instar weevil in pecan nuts [48]. In grapevine, a recent study aimed to characterize inflorescence architecture using X-rays [49]. The authors found correlations between 24 morphological traits among 392 samples of 10 wild *Vitis* species. They were able then to perform a multivariate discriminant analysis to classify the different species.

However, there is not much work showing the application of such methods on walnut. A technical report is available on 3D reconstruction of a walnut using X-ray CT [50] and also a data collection providing an image reconstruction pipeline [51]. However, in this study, the authors aimed to develop a method adapted for machine learning and they used 42 walnuts as models because they have variability, hard shell, softer kernel and empty space which are characteristics similar to the human head. Here, we present the development of a robust method that extracts for the first time complete morphological measurements of walnut using X-ray CT within a worldwide germplasm collection. This method offers the possibility to quantify several traits such as rugosity, sphericity and shape indexes previously really difficult to quantify but essential for French walnut industry. It also allows to evaluate the filling ratio which is the volume occupied by the kernel over the total nut volume, until now impossible to know without cracking the walnut. Our results will be helpful for new breeding programs by selecting the best accessions as genitors in order to release tomorrow's varieties. Our method could be also used as a reference for walnut or other nut crop germplasm investigation in any breeding program and could

pave the way for future application in industry, particularly for internal quality control.

Methods

Plant materials and sample preparation

A panel of 161 unique *J. regia* accessions from worldwide was analysed. All the accessions are maintained at the *Prunus* and *Juglans* Genetic Resources Center and located in the Fruit Experimental Unit of INRAE in Toulenne (latitude 44°34′37.442″N – longitude 0°16′51.48″W), near Bordeaux, France (Additional file 1). The panel choice was made thanks to a previous work based on genetic diversity and phenotypic variation results [52].

In-shell walnut sampling was performed during harvest season in September 2018 and walnuts were dried following classical French industrial recommendations, for 2 days at 25 °C using a food dryer, and then stored until analyses in a cold room set to 2 °C. For each accession, a selection of 50 walnuts was performed based on their sanitary state and sent to the GEVES laboratory (Beaucouzé, France). All the samples were stored in an environmentally controlled room at 10 °C and 47.75% (±3.4) relative humidity until use. During the preparation, the walnuts were embedded in a floral foam sample holder (9 length × 8 width × 21 cm height) to keep the samples from any abrupt or slight movement during the scanning process in order to avoid producing distorted images. The floral foam was chosen based on preliminary trials on different low-density materials in order to observe the level of attenuation of the X-rays passing through these materials.

The walnuts were scanned in batches, knowing that, the sample size was not fixed due to the huge variation between the walnuts in size and the limited scanning scope of the detector. The sample size ranged from 5 up to 16 walnuts per scan.

X-ray computed tomography imaging system specifications

X-ray CT scans were performed at the GEVES laboratory (Beaucouzé, France) using a 3D X-ray imaging system, the NSI X-50 model from North Star Imaging©, Inc. (Minnesota, USA), which has a focus tube with focal spot up to 1 µ, a voltage range of 10–130 kV, an electric current range of 50–300 µA, a flat-panel detector with a resolution of 256 × 256 and an adjustable rotary stage.

Image acquisition and reconstruction

Scans were obtained at constant electron acceleration energy of 120 kV, an electric current of 300 µA and a rotation speed of 4.99 degrees/s resulting in a scan duration of 14 m34 s. A total of 2,164 images (or radiographs) in a.tif format were used for reconstructing each 3D image using North Star Imaging© reconstruction software EFX-CT (version 1.9.5.12) where the resulting 3D images were exported in a.nsihdr format with a resolution of 992 × 992 × 2991 voxels (voxel size of 0.1 × 0.1 × 0.1 mm).

After 3D reconstruction, a multi-stage workflow was applied to all CT images in order to eventually achieve a quantitative study. This workflow consists of three key steps as illustrated in Fig. 1d-f: preprocessing steps, walnuts individualization, and morphological traits extraction and quantification.

(*i*) The preprocessing: it begins with automatically loading each image in our image collection $I_{(1,x,y,z)}, ..., I_{(m,x,y,z)}$ sequentially (where m is the number of images) and then denoising them in order to eliminate the noise and artifacts introduced by the X-ray system during image acquisition. We applied Gaussian filter, successive morphological operations such as opening and closing, and also, we removed the unwanted small spots in the image. Subsequently, all the voxels which represent the sample holder are eliminated and only the voxels which represent the walnuts are preserved by discarding all the voxels below a certain threshold τ resulting in a binary mask $M_{(x,y,z)}$ according to the Eq. (1).

$$M_{(x,y,z)} = \begin{cases} 0 \; if \; I_{(x,y,z)} \leq \tau \\ 1 \; if \; I_{(x,y,z)} > \tau \end{cases} \quad (1)$$

(*ii*) The individualization: the task of extracting the features of the walnuts and quantification is challenging especially if the walnuts are touching. To overcome this difficulty, all the walnuts in the images were separated and individualized (Fig. 2). The individualization step consists of multiple sub-steps such labelling, masking, convex hull estimation and exporting. Labelling is based on voxel connectivity in the whole 3D volume in order to determine the regions of interest which represent the walnuts by assigning identical values to all the voxels that belong to an individual walnut. Each walnut in the image was assigned a unique value starting from 1 to n consecutively where n represents the total number of walnuts in the image. Then, labelling was followed by generating a set of masks using the determined regions of interest then finally, given the original input image and the generated set of masks $K_{(x,y,z)}$ as shown in the Eq. (2), each walnut was exported in a separate sub-image $i_{(x,y,z)}$ in a.nsihdr format after estimating the convex hull of each walnut. Loading, preprocessing, labelling and convex hull calculation take 20 min for an average sized sample.

$$f\left(I_{(x,y,z)}, K_{(x,y,z)}\right) = \left\{i_{(1,x,y,z)}, i_{(2,x,y,z)} \ldots i_{(n,x,y,z)}\right\} \quad (2)$$

(*iii*) The morphological traits extraction and quantification: at this step, the principal components of the

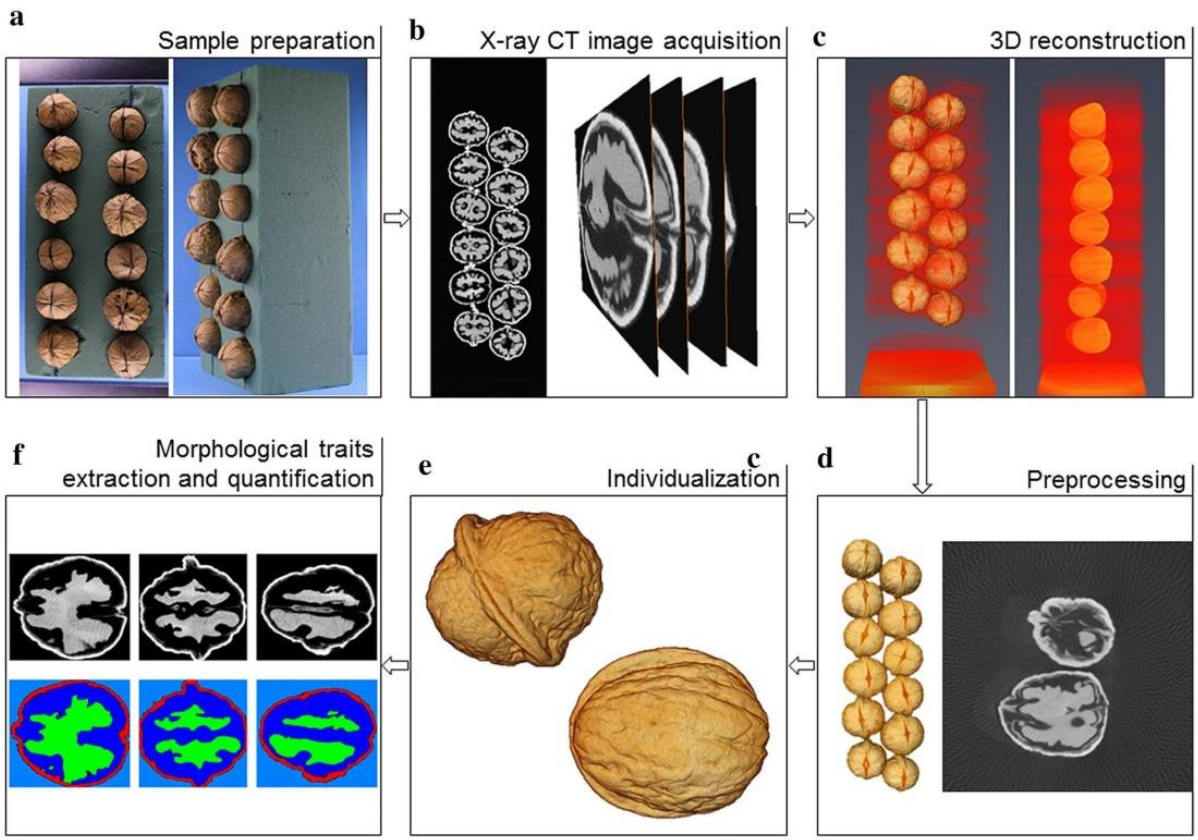


Fig. 1 X-ray CT workflow of walnut measurements. **a** Preparation of walnut samples using floral foam (12 walnuts/batch in average), **b** Acquisition of X-ray CT images with right images of 2D slices, **c** 3D reconstruction, **d** Preprocessing with right image showing noise and artifacts that have to be removed, **e** Individualization of each walnut of the batch and **f** XY, YZ, ZX segmentation and labelling of a walnut leading to segmentation of each different part, with the shell in red, the kernel in green and the empty space in blue

walnuts whose morphological features were segmented using multi-level thresholding and watershed algorithm which is a transformation that treats the image like a topographic map [53]. As a consequence, in our case, each main part of each walnut was segmented and was given a unique label as shown in Fig. 1f, considering the kernel, the shell, and the empty space between the kernel and the shell. The optimum threshold τ and greyscale ranges of the principal parts of the walnuts were estimated experimentally based on the analysis of the histogram, using K Nearest Neighbor clustering method, that corresponds to the distribution of the intensities of the images (Fig. 3). Cropping and exporting take 1m30s/walnut. An additional movie file shows this in more detail (Additional file 2).

A fully automated in-house image processing pipeline was developed using the Thermo Scientific Avizo© software V9.0.0 built-in functions, the MATLAB© version 7.7.0 R2008b image processing toolbox [54] from The MathWorks©, Inc. (Massachusetts, USA), the TCL scripting language and Spyder Python IDE. To use this pipeline, walnuts that have no damage on the shell are required.

Measurements of a total of 14 morphological and shape descriptors were obtained: the nut length, the nut face diameter, the nut profile diameter, the nut volume, the nut shape VA3D, the nut Feret shape 3D (defined by D/d where d is the minimum Feret diameter and D is the maximum Feret diameter in the orthogonal direction, so 90° from the minimum Feret diameter; the maximum Feret diameter is the maximum diameter of an object as if it were freely rotating in three dimensions using a caliper [55]), the nut surface area, the shell volume, shell thickness, the kernel volume, the kernel filling ratio, and the empty space volume (Table 1). Quantification takes 2 min/walnut.

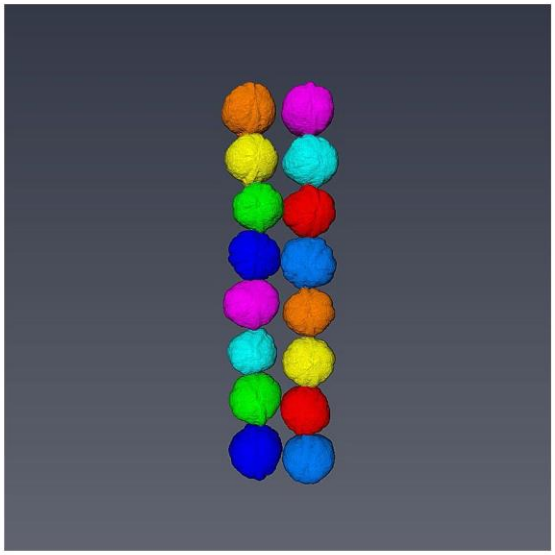


Fig. 2 Separation and individualization of walnuts

In addition, 2 supplementary traits, the nut sphericity (close to roundness) (3) and the shell rugosity (or surface roughness) (4) indexes, were measured using a non-standard calculation since they are not supported by the Avizo© software:

$$Nut\ Sphericity = \frac{\pi^{\frac{1}{3}}(6V_n)^{\frac{2}{3}}}{A} \quad (3)$$

$$Shell\ Rugosity = \frac{A}{\sqrt[3]{36\pi\,{V_n}^2}} \quad (4)$$

where V_n is the nut volume and A is its surface area. The sphericity of a sphere is 1 and any object which is not a sphere will have sphericity less than 1. The nut shape VA3D is defined by $\frac{A^3}{36\pi {V_n}^2}$.

Experiments were run on a workstation equipped with an Intel® Xeon® dual-core processor running at 3 GHz using 64 MB of RAM and running Windows® version 10. The dataset was assessed using R software [56] with the package "tidyverse" [57]. Pearson correlation matrices were performed using the package "corrplot" [58] and Principal Component Analysis (PCA) using the package "FactoMineR" [59].

Results

Building a workflow for assessing phenotypic variation of the germplasm collection from CT images and 3D processing

The X-ray CT imaging workflow permitted to characterize the INRAE germplasm collection for 14 traits whose descriptive statistics (mean, standard deviation,

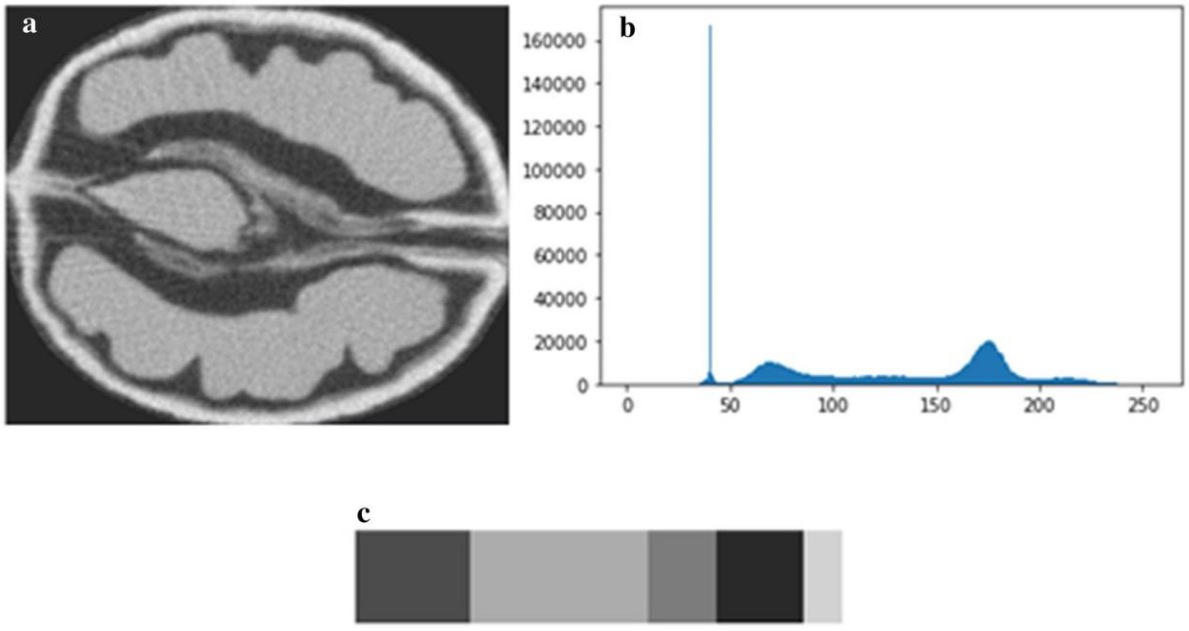


Fig. 3 Greyscale histogram analysis. **a** Example of a 2D slice, **b** the corresponding histogram and **c** a bar that visualizes an approximate percentage of pixels in each cluster

Table 1 Walnut morphological traits measured by the workflow

Morphological trait	Symbol	Description	Unit
Nut			
Nut length	L	The largest length of the nut from the base to the end	mm
Nut face diameter	F	The largest longitudinal section of the nut through suture	mm
Nut profile diameter	P	The largest longitudinal section of the nut perpendicular to suture	mm
Nut volume	V_n	Total volume of the nut, $V_n = V_s + V_k + V_e$	mm^3
Nut shape VA3D	S_1	Shape factor of the nut	–
Nut feret shape 3D	S_2	Feret shape factor of the nut	–
Nut surface area	A	Surface area of the nut	mm^2
Nut sphericity	Ψ	Index of nut roundness	–
Shell			
Shell volume	V_s	Volume of the shell	mm^3
Shell thickness	T	Thickness of the shell	mm
Shell rugosity	Ω	Index of shell surface roughness	–
Kernel			
Kernel volume	V_k	Volume of the kernel	mm^3
Kernel filling ratio	R	Ratio of the kernel volume V_k to the total volume of the nut V_n	%
Empty Space			
Empty space volume	V_e	Volume of the empty space	mm^3

Table 2 Descriptive statistics of walnut morphological traits

Morphological trait	Unit	Mean ± SD[a]	Range
Nut			
Nut length	mm	38.39 ± 2.18	28.57–51.43
Nut face diameter	mm	32.27 ± 1.70	25.99–40.75
Nut profile diameter	mm	33.29 ± 1.70	27.06–42.84
Nut volume	mm^3	19,400.02 ± 2669.03	10,382.05–42,813.08
Nut shape VA3D	–	1.47 ± 0.08	1.24–1.69
Nut feret shape 3D	–	1.25 ± 0.05	1.12–1.48
Nut surface area	mm^2	4019.53 ± 401.91	2622.59–7093.53
Nut sphericity	–	0.88 ± 0.02	0.84–0.93
Shell			
Shell volume	mm^3	4076.78 ± 595.05	2390.66–9051.88
Shell thickness	mm	1.03 ± 0.12	0.73–1.49
Shell rugosity	–	1.14 ± 0.02	1.07–1.19
Kernel			
Kernel volume	mm^3	5723.89 ± 1039.03	3408.85–9548.93
Kernel filling ratio	%	30.02 ± 3.55	20.66–37.42
Empty space			
Empty space volume	mm^3	9599.35 ± 1719.56	4536.51–24,212.21

[a] SD is the abbreviation for standard deviation

minimum and maximum) are given in Table 2. The collection exhibits high phenotypic variation in morphology-related traits, particularly for Nut Volume which ranged from approx. 10,000 for 'UK 56–12' to more than 42,000 mm^3 for 'Carmelo' (Additional file 3). We found between the minimum and the maximum a factor of 1.8 for Nut Length, 1.6 for Nut Face and Nut Profile Diameters, and 2.0 for Shell Thickness. Globally, the Kernel Filling Ratio is low, ranging from approx. 21 to 37%, and the Nut Sphericity has a low variation, from 0.84 to 0.93.

Using Pearson correlation coefficient, we found unsurprisingly significant high positive correlations (p value 0.001) between all the morphometric traits: Nut Face/Profile Diameters, Nut Length, Nut Surface Area, Nut/Shell/Kernel/Empty Space Volumes (Fig. 4). We can observe that Nut Length is positively correlated with Nut Face and Profile Diameters (0.67 and 0.64, respectively), meaning that the longer the walnut is, the larger the diameter is. Then, Nut Volume is unsurprisingly positively correlated with all three previous traits (Nut Length 0.80, Nut Face Diameter 0.91 and Nut Profile Diameter 0.94), since the variation of the volume of an object depends on those three dimensions. Nut Surface Area is also positively correlated with all volumes-related traits (from 0.85 to 0.99). We also observed significant moderate positive correlations (p value 0.001) between all those morphometric traits (except Kernel Volume), and Shell Rugosity and Nut Shape VA3D (from 0.30 to 0.50), indicating that the bigger the fruit is, the rougher the shell is. Those morphometric traits are significantly negatively correlated (p value 0.001) with Nut Sphericity and Kernel Filling Ratio. This means that a big fruit is less spherical and less filled by the kernel. Finally, Nut

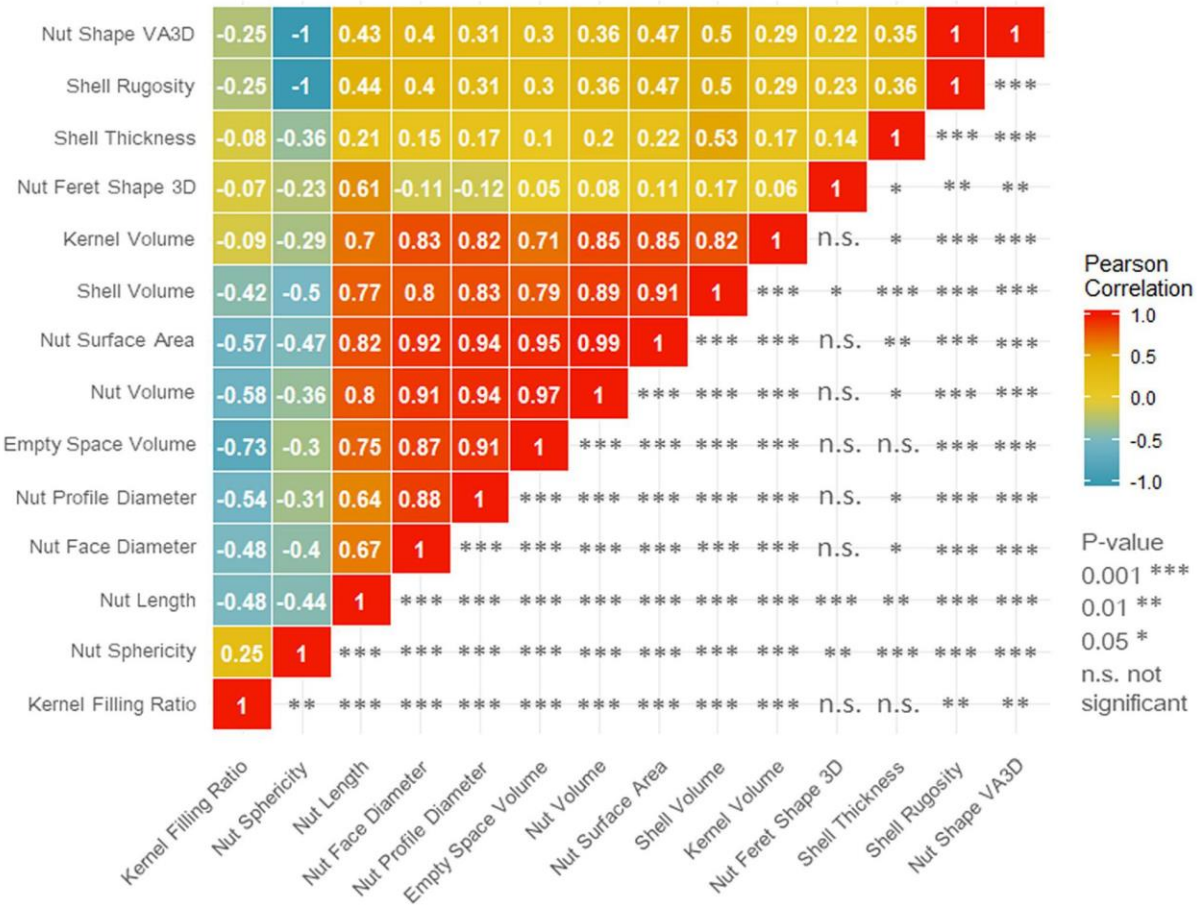


Fig. 4 Pearson correlation matrix for walnut morphological traits

Sphericity is perfectly negatively correlated (p value 0.001) with Shell Rugosity and Nut Shape VA3D showing that the closer the fruit gets to a spherical shape, the smoother the shell is.

The X-ray CT method for the selection of superior genotypes

In French walnut industry, both producers and consumers have particular expectations. For example, they would prefer large round-shaped walnuts, easy to crack and having a high kernel filling ratio. A Principal Component Analysis using the dataset obtained permitted us to select interesting genotypes for the previous important traits (Fig. 5).

The first two dimensions of the PCA explain 75.5% of the total variance (Fig. 5b). The dimension 1 corresponds to the morphometric traits (Nut Length, Nut Face Diameter, Nut Profile Diameter), the volumes (Nut Volume, Shell Volume, Kernel Volume, Empty Space Volume) and the Nut Surface Area (Fig. 5a and d). The dimension 2 is linked to the Nut Shape VA3D, the Nut Sphericity and the Shell Rugosity (Fig. 5a, b). Unfortunately, the Kernel Filling Ratio and the Shell Thickness are both traits badly represented by the PCA (Fig. 5). By visualizing the scatterplot of the individuals, it is now easy to select superior genotypes for most of the traits (Fig. 5c). For example, 'Carmelo' (131) and 'Germisara' (135) are the two accessions giving the biggest walnuts, and 'Milotaï n°10′ (113) and 'Lozeronne n°1′ (7) are the two accessions giving the more round-shaped walnuts.

Since the Shell Thickness and the Kernel Filling Ratio are not well represented by the PCA, we also looked at the ten superior genotypes for both traits (Table 3).

'Lozeronne n°1′, previously identified as highly round-shaped, gives walnuts with a thin shell, more easily cracked, but, this accession is also one of those that gives the smallest walnuts. The two Iranian accessions 'IR 13–1′ and 'IR 100–2′ are among those having

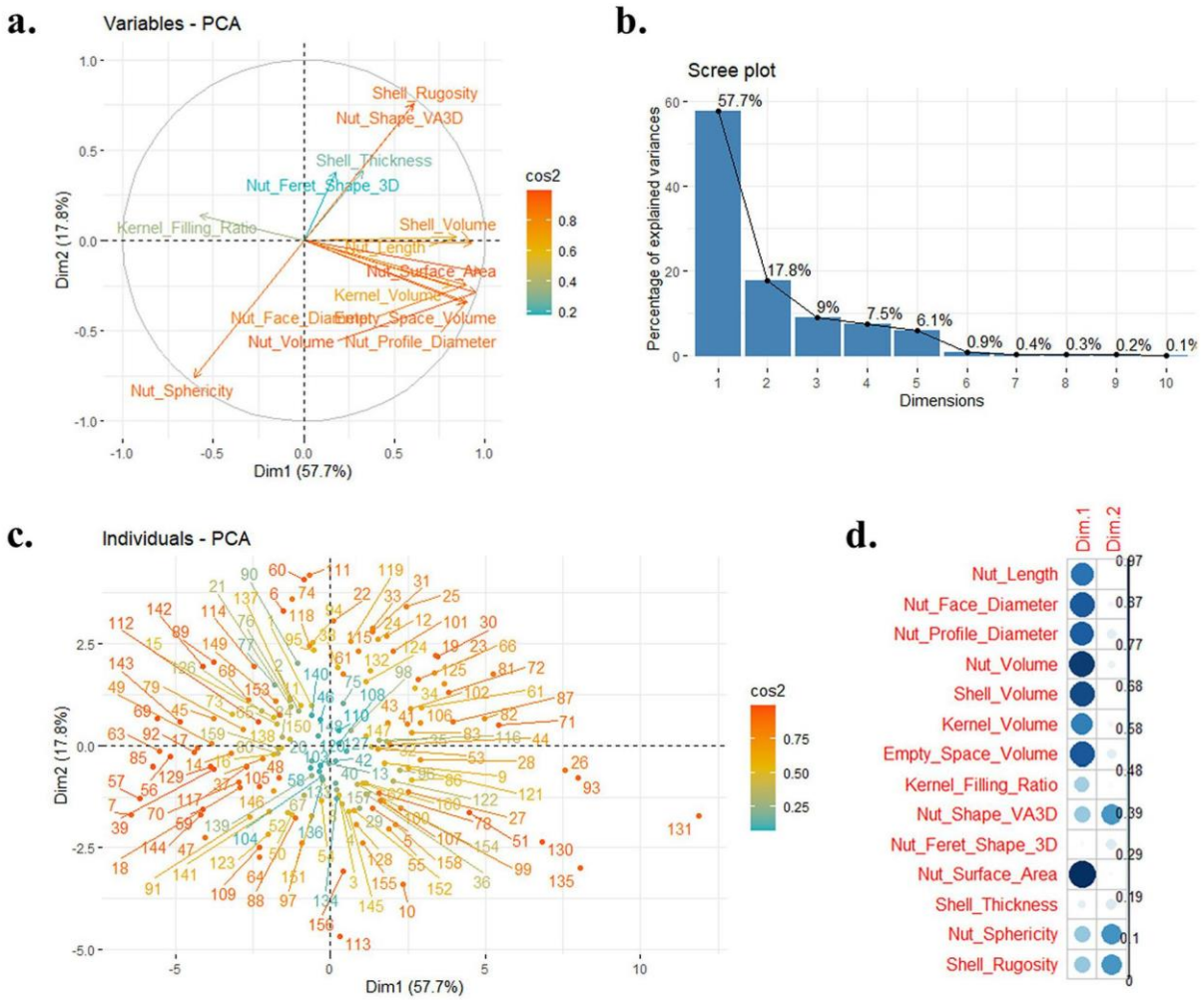


Fig. 5 Principal Component Analysis using the 161 walnut accessions and the 14 traits quantified using the X-ray CT method. **a** PCA correlation circle of the 14 variables (dimensions 1 and 2), **b** Scree plot of the percentage of variances explained by the first ten dimensions, **c** PCA scatterplot of the 161 accessions (dimensions 1 and 2), **d** Correlation plot of the 14 variables (dimensions 1 and 2). For a and c plots, color gradient indicates the quality of the representation of each variable given by the squared cosines cos^2

the highest kernel/nut ratio. In our panel, although we have a large diversity, we do not find the optimal accession for all the traits, but we can identify the best for each trait with this technique.

Discussion

Primarily developed for medical purposes, X-ray CT is now widely applied in food science, for instance to track the microstructural evolution of dairy products or to quantify salt concentrations in pork meat [60]. An overview of applications related to various food products showed the use of X-ray CT in chestnut for post-harvest assessment of internal decay and in pecan nuts for components screening [61–63]. We showed for the first time that X-ray CT is a method of choice also for walnut germplasm analysis of morphological traits. We obtained quantitative data of high accuracy on 14 traits including volumes and shell thickness which are classically obtained by cracking the nuts. The obtained dataset includes information that is crucial for the INRAE walnut germplasm collection, allowing a precise and relevant characterization of the nuts of each accession. It will help to select superior genitors for a breeding program, so that we can hope to combine many favourable traits in a new variety.

Table 3 List of superior genotypes considering the shell thickness and the kernel filling ratio

Superior genotype	Mean ± SD[a]
Shell thickness (mm)	
Lozeronne n°1	0.7316 ± 0.1987
Izvor 10	0.8113 ± 0.1956
H 110–34	0.8293 ± 0.1845
Marchetti	0.8437 ± 0.1855
H 113–21	0.8527 ± 0.0995
Pourpre Hollande	0.8555 ± 0.1752
Sexton	0.8705 ± 0.0270
AS 1	0.8833 ± 0.1520
S 34 B Pyrrus	0.8859 ± 0.1448
H 131–08	0.8865 ± 0.1010
Kernel filling ratio (%)	
IR 13–1	37.42 ± 3.10
H 102–15	35.80 ± 2.77
IR 100–2	35.61 ± 3.20
UK 224–6	35.29 ± 5.53
S 4 B Thétis	35.09 ± 3.59
Lozeronne n°1	35.07 ± 3.03
Wepster W2	34.92 ± 3.19
Ferjean	34.69 ± 3.04
Grappes Suisse	34.51 ± 2.55
Cheinovo	34.37 ± 4.76

[a] SD is the abbreviation for standard deviation

This method is suitable for all types of walnuts, regardless of size and shape. However, the walnuts analysed should not have shell damaged, as, in this case, it is difficult to extract the walnut features with the algorithm developed in our study. The observations show that the shell and the empty space inside those walnuts were incorrectly segmented. The measurements of such walnuts were totally excluded from the results. For future work, it is recommended to discard the damaged walnuts during the sample preparation step, before performing CT scanning in order to avoid any inaccurate calculations. One beneficial direction could be increasing the robustness of the pipeline by designing an artificial intelligence-based task for automatic detection of the walnuts with notably broken and/or damaged shells, and eliminating them before morphological traits extraction and quantification step.

We demonstrate that this method presents numerous advantages compared to classical morphological evaluation, mainly the accuracy of measurement and access to several measures without cracking. For all these reasons, the development of this technology for research scale but also for the industry could be very useful in the future. However, the cost of this technology can constitute a limitation of its use but this method will save time for any lab or industry operator willing to use X-ray CT on walnuts and can be easily transferable to other nuts species. For instance, with the addition of deep learning methods, we can imagine such an application for food security and commercial frauds purposes, in place of molecular biology authentication.

For nuts, it is economically important to know from which cultivar they come from. For French walnut industry, the fruits of the most produced cultivar 'Franquette', old cultivar representing 70% of the orchard surfaces, can be distinguished from those of 'Fernor', a cultivar released in 1995 and sold at a higher price. Based on our results from 50 walnuts, the value of the nut Feret Shape for 'Franquette' is between 1.380 and 1.500 whereas the value for 'Fernor' is between 1.210 and 1.304, considering the standard deviation. The Feret shape is clearly a powerful descriptor to discriminate two cultivars and establish genetic origin, since DNA isolation is often a difficult task on nut materials due to the high lipid content of kernels [64].

Conclusions

We presented a method for better resolution phenotyping of walnuts based on X-ray CT compared to classical measurement methods. The data will be used for INRAE walnut germplasm management, but also for GWAS purposes and for selecting superior genotypes in a future breeding program. This method could be easily adapted for any nut species and potentially moved towards the identification of the first steps of infection by pathogens.

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1186/s13007-020-00657-7.

Additional file 1. List of the 161 accessions studied.

Additional file 2. Video of 3D reconstruction and individualization steps of a walnut.

Additional file 3. Complete dataset.

Abbreviations

3D: Three-dimensional space; ANR: Agence Nationale de la Recherche; CT: Computed tomography; GEVES: Groupe d'étude et de contrôle des variétés et des semences; GWAS: Genome-wide association study; INRAE: Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement; IPGRI: International Plant Genetic Research Institute; kV: Kilovolt; PCA: Principal Component Analysis; µA: Microamper.

Acknowledgements

First, we would like to thank the late Eric Germain, former head of the breeding program at the INRAE of Bordeaux from 1977 to 2007. His remarkable work, then continued by Francis Delort, has given us the opportunity to study a rich set of plant material. We thank the Fruit Tree Experimental Unit of the INRAE in Toulenne and the *Prunus/Juglans* Genetic Resources Center for the maintenance of the collection and for helping us to collect the samples.

We acknowledge the GEVES for X-ray CT analyses. The measurements were performed on a phenotyping platform, PHENOTIC, dedicated to instrumentation and imagery for seeds, seedlings and plants in Angers, France (https://www6.inrae.fr/phenotic/; https://doi.org/10.15454/U2BWFJ). Then, we thank Aurélie Charrier from GEVES for her comprehensive reading of the manuscript. Finally the CTIFL, holder of the project "INNOV'noyer", in partnership with the INRAE of Bordeaux, want to thank the "Région Nouvelle-Aquitaine", and "Cifre" convention of "ANRT" (Agence Nationale de la Recherche et de la Technologie).

Authors' contributions
LL performed the X-ray CT analyses. SH developed the workflow. AB performed the phenotypic statistical analyses. ED, FL and AB conceived and coordinated the study. AB and SH wrote the manuscript. All authors read and approved the final manuscript.

Funding
This work has been mainly funded by the CTIFL (Centre Technique Interprofessionnel des Fruits et Légumes). It has been also funded by the "Région Nouvelle-Aquitaine" in the project "INNOV'noyer", coordinated by the CTIFL, and in partnership with the INRAE of Bordeaux, and funded by the "Cifre" convention number 2016/1558 of ANRT (Agence Nationale de la Recherche et de la Technologie).

Availability of data and materials
The dataset generated during the current study is available in the Additional file 3.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1] Univ. Bordeaux, INRAE, Biologie du Fruit et Pathologie, UMR 1332, 33140 Villenave d'Ornon, France. [2] CTIFL, centre opérationnel de Lanxade, 24130 Prigonrieux, France. [3] GEVES, 49070 Beaucouzé, France.

Received: 21 April 2020 Accepted: 17 August 2020
Published online: 26 August 2020

References

1. Aradhya MK, Potter D, Simon CJ. Cladistic biogeography of Juglans (Juglandaceae) based on chloroplast DNA intergenic spacer sequences. In: Motley TJ, Zerega N, Cross H, editors. Darwins Harvest New Approaches Origins Evolution Conservation Crops. New York: Columbia University Press; 2006. p. 143–170.
2. Bernard A, Lheureux F, Dirlewanger E. Walnut: past and future of genetic improvement. Tree Genet Genomes. 2018;14:1.
3. Rahaman MM, Chen D, Gillani Z, Klukas C, Chen M. Advanced phenotyping and phenotype data analysis for the study of plant growth and development. Front Plant Sci. 2015;6:619.
4. Fiorani F, Schurr U. Future scenarios for plant phenotyping. Ann Rev Plant Biol. 2013;64:267–91.
5. Ghimiray M, Vernooy R. The importance and challenges of crop germplasm interdependence: the case of Bhutan. Food Sec. 2017;9:301–10.
6. Smith JSC, Smith OS. The description and assessment of distance between inbred lines of maize. 2: the utility of morphological—biochemical—and genetic descriptors and a scheme for the testing of distinctiveness between inbred lines. Maydica. 1989;34:151–61.
7. Solar A, Stampar F. Characterisation of selected hazelnut cultivars: phenology, growing and yielding capacity, market quality and nutraceutical value. J Sci Food Agric. 2011;91:1205–12.
8. Hansche PE, Beres V, Forde HI. Estimates of quantitative genetic properties of walnut and their implications for cultivar improvement. J Am Soc Hortic Sci. 1972;97:279–85.
9. Sholokhov LV. Quality characters of fruit in walnut and correlation between them. Tr Novocherkas Inzha Melior Inta. 1974;13:106–12.
10. Komanich IG. Correlations between walnut fruit characters in a seed population. Bull Acad Sci RSS Mold. 1980;2:27–30.
11. Sen SM. Correlations between shell thickness, shell cracking resistance, shell seal and shell upright cracking resistance and some other fruit quality characters in walnut (*Juglans regia* L.). Doga Bilim Derg D2 Tarm Ve Orman. 1985;9:10–24.
12. Sharma OC. Variability in seedlings trees of Persian walnut in Solan area of Himachal Pardesh. PhD. India: Solan University of Horticulture and Forestry; 1996.
13. Akça Y, Sen SM. The relationship between dichogamy and yield-nut characteristics in *Juglans regia* L. Acta Hortic. 1997;442:215–6.
14. Eskandari S, Hassani D, Abdi A. Investigation on genetic diversity of Persian walnut and evaluation of promising genotypes. Acta Hortic. 2005;705:159–66.
15. Khadivi-Khub A, Ebrahimi A, Sheibani F, Esmaeili A. Phenological and pomological characterization of Persian walnut to select promising trees. Euphytica. 2015;205:557–67.
16. Amiri R, Vahdati K, Mohsenipoor S, Mozaffari MR, Leslie C. Correlations between some horticultural traits in walnut. HortScience. 2010;45:1690–4.
17. Abedi B, Parvaneh T. Study of correlations between horticultural traits and variables affecting kernel percentage of walnut (*Juglans regia* L.). J Nuts. 2016;7:35–44.
18. IPGRI. Descriptors for Walnut (Juglans spp.). Rome: International Plant Genetic Resources Institute; 1994.
19. Ghasemi M, Arzani K, Hassani D. Evaluation and identification of walnut (*Juglans regia* L.) genotypes in Markazi province of Iran. Crop Breed J. 2012;2:119–24.
20. Ebrahimi A, Khadivi-Khub A, Nosrati Z, Karimi R. Identification of superior walnut (*Juglans regia* L.) genotypes with late leafing and high kernel quality in Iran. Sci Hortic. 2015;193:195–201.
21. Arzani K, Mansouri-Ardakan H, Vezvaei A, Roozban MR. Morphological variation among Persian walnut (*Juglans regia*) genotypes from central Iran. N Z J Crop Hortic Sci. 2008;36:159–68.
22. Ahandani EA, Ramandi HD, Sarmad J, et al. Evaluation of morphological diversity among some Persian walnut accessions (*Juglans regia* L.) in Guilan, northern Iran. Int J Plant Biol Res. 2014;2:1015–22.
23. Mahmoodi R, Dadpour MR, Hassani D, et al. Development of a core collection in Iranian walnut (*Juglans regia* L.) germplasm using the phenotypic diversity. Sci Hortic. 2019;249:439–48.
24. Bayazit S. Determination of relationships among kernel percentage and yield characteristics in some Turkish walnut genotypes by correlation and path analysis. J Anim Plant Sci. 2012;22:513–7.
25. Aslantaş R. Identification of superior walnut (*Juglans regia*) genotypes in north-eastern Anatolia. Turkey N Z J CropHortic Sci. 2006;34:231–7.
26. Asma BM. Pomological and phenological characterization of promising walnut (*Juglans regia* L.) genotypes from Malatya, Turkey. Acta Sci Pol Horturum Cultus. 2012;11:169–78.
27. Fikret Balta M, Dogan A, Kazankaya A, et al. Pomological definition of native walnuts (*Juglans regia* L.) grown in Central Bitlis. J Biol Sci. 2007;7:442–4.
28. Karadag H, Akça Y. Phenological and pomological properties of promising walnut (*Juglans regia* L.) genotypes from selected native population in Amasya Province. Afr J Biotechnol. 2011;10:16763–8.
29. Zeneli G, Kola H, Dida M. Phenotypic variation in native walnut populations of Northern Albania. Sci Hortic. 2005;105:91–100.
30. Miletic R, Zikic M, Mitic N, Nikolic R. Biological and pomological characteristics of superior walnut selections. Genetika. 2003;35:123–30.
31. Gandev S, Dzhuvinov V. Evaluation of some walnut cultivars under the climatic conditions of South Bulgaria. Agroznanje. 2015;15:5–16.
32. Cosmulescu S, Botu M. Walnut biodiversity in South-Western Romania-resource for perspective cultivars. Pak J Bot. 2012;44:307–11.
33. Poggetti L, Ermacora P, Cipriani G, et al. Morphological and carpological variability of walnut germplasm (*Juglans regia* L.) collected in North-Eastern Italy and selection of superior genotypes. Sci Hortic. 2017;225:615–9.
34. Li L, Zhang Q, Huang D. A review of imaging techniques for plant phenotyping. Sensors. 2014;14:20078–11111.

35. Golzarian MR, Frick RA, Rajendran K, et al. Accurate inference of shoot biomass from high-throughput images of cereal plants. Plant Methods. 2011;7:1–11.
36. Munns R, James RA, Sirault XR, Furbank RT, Jones HG. New phenotyping methods for screening wheat and barley for beneficial responses to water deficit. J Exp Bot. 2010;61:3499–507.
37. Jones HG, Serraj R, Loveys BR, et al. Thermal infrared imaging of crop canopies for the remote diagnosis and quantification of plant responses to water stress in the field. Funct Plant Biol. 2009;36:978–89.
38. Shibayama M, Sakamoto T, Takada E, et al. Estimating paddy rice leaf area index with fixed point continuous observation of near infrared reflectance using a calibrated digital camera. Plant Product Sci. 2011;14:30–46.
39. Huang J, Liao H, Zhu Y, et al. Hyperspectral detection of rice damaged by rice leaf folder (*Cnaphalocrocis medinalis*). Comput Electr Agric. 2012;82:100–7.
40. Biskup B, Scharr H, Schurr U, Rascher U. A stereo imaging system for measuring structural parameters of plant canopies. Plant Cell Environ. 2007;30:1299–308.
41. Rascher U, Blossfeld S, Fiorani F, et al. Non-Invasive approaches for phenotyping of enhanced performance traits in bean. Funct Plant Biol. 2011;38:968–83.
42. Mir AA, Reynolds M, Pinto F, Khan MA, Bhat MA. High-throughput phenotyping for crop improvement in the genomics era. Plant Sci. 2019;282:60–72.
43. Kotwaliwale N, Singh K, Kalne A, et al. X-ray imaging methods for internal quality evaluation of agricultural produce. J Food Sci Technol. 2014;51:1–15.
44. Tollner EW, Hung YC, Upchurch BL, Prussia SE. Relating X-ray absorption to density and water content in apples. Trans ASAE. 1992;35:1921–8.
45. Lammertyn J, Jancsok P, Dresselaers T, et al. Analysis of the time course of core breakdown in 'Conference' pears by means of MRI and X-ray CT. Postharvest Biol Technol. 2003;29:19–28.
46. Kim S, Schatzki T. Detection of pinholes in almonds through X-ray imaging. Trans ASAE. 2001;44:997–1003.
47. Khosa I, Pasero E. Feature extraction in X-ray images for hazelnuts classification. 2014 International Joint Conference on Neural Networks (IJCNN), Beijing, 2014, pp. 2354–2360. https://doi.org/10.1109/IJCNN.2014.6889661.
48. Harrison RD, Gardner WA, Tollner WE, Kinard DJ. X-ray computed tomography studies of the burrowing behaviour of fourth-instar pecan weevil (*Coleoptera*: *Curculionidae*). J Econ Entomol. 1993;86:1714–9.
49. Li M, Klein LL, Duncan KE, et al. Characterizing 3D inflorescence architecture in grapevine using X-ray imaging and advanced morphometrics: implications for understanding cluster density. J Exp Bot. 2019;70:6261–76.
50. Austin L, Kumar R, Kousar B, Lampadaris CH, Lucas MM. Tomography of a walnut. Tech Rep. 2018. https://doi.org/10.13140/RG.2.2.25130.52161.
51. Der Sarkissian H, Lucka F, van Eijnatten M, et al. A cone-beam X-ray computed tomography data collection designed for machine learning. Sci Data. 2019;6:215.
52. Bernard A, Barreneche T, Lheureux F, Dirlewanger E. Analysis of genetic diversity and structure in a worldwide walnut (Juglans regia L) germplasm using SSR markers. PLoS ONE. 2018;13(11):e0208021.
53. Beucher S, Meyer F. The morphological approach to segmentation: the watershed transformation. In: Dougherty ER, editor. Mathematical morphology in image processing. Boca Raton: CRC Press; 1993. p. 433–482.
54. Mathworks T. MATLAB [Internet]. 2013. https://www.mathworks.com/products/matlab.html. Accessed 20 Jan 2020.
55. Russ JC, Neal FB. The Image Processing Handbook. Raleigh: Seventh Edition. CRC Press; 2017.
56. R Development Core Team. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2008. https://www.R-project.org/. Accessed 10 Apr 2020.
57. Wickham H. Tidyverse: Easily Install and Load the 'Tidyverse'. R package version 1.2.1. 2017. https://CRAN.R-project.org/package=tidyverse. Accessed 10 Apr 2020.
58. Wei T, Simko VR. Package "corrplot": Visualization of a Correlation Matrix (Version 0.84). 2017. https://github.com/taiyun/corrplot. Accessed 10 Apr 2020.
59. Lê S, Josse J, Husson F. FactoMineR: an R package for multivariate analysis. J Stat Softw. 2008;25:1–18.
60. Vidhya M, Varadharaju N, John Kennedy Z, Amirtham D, Manohar JD. Applications of X-ray computed tomography in food processing. J Food Process Technol. 2017;8:673.
61. Schoeman L, Williams P, du Plessis A, Manley M. X-ray micro-computed tomography (μCT) for non-destructive characterisation of food. Trends Food Sci Technol. 2016;47:10–24.
62. Donis-González IR, Guyer DE, Fulbright DW, Pease A. Postharvest noninvasive assessment of fresh chestnut (*Castanea* spp.) internal decay using computer tomography images. Postharvest Biol Technol. 2014;94:14–25.
63. Kotwaliwale N, Weckler PR, Brusewitz GH. X-ray attenuation coefficients using polychromatic X-ray imaging of pecan components. Biosyst Eng. 2006;94:199–206.
64. Dang PM, Chen CY. Modified method for combined DNA and RNA isolation from peanut and other oil seeds. Mol Biol Rep. 2013;40:1563–8.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

BMC

2. Analyse de la diversité génétique et de la structure de la collection de ressources génétiques *Juglans* de l'INRAE via l'utilisation de marqueurs SSR

Analyse de la diversité génétique et de la structure des 253 accessions *Juglans* spp. via l'utilisation de 13 marqueurs SSR

La collection de ressources génétiques *Juglans* spp. de l'INRAE est le fruit du travail d'Eric Germain et de son équipe entre 1988 et 2000. Les prospections et échanges menés dans différentes régions du monde ont permis de rassembler une grande partie de la diversité du genre *Juglans* spp. Si la collection a longtemps été caractérisée au niveau phénotypique pour les besoins du programme d'amélioration, les informations quant à sa diversité génétique demeuraient manquantes.

Afin de réaliser les analyses de génétique d'association, il est important de connaître les liens de parenté entre les accessions et la structure. S'ils ne sont pas pris en compte dans les analyses, ces « confounding effects » peuvent mener à de fausses associations marqueur/phénotype, donc des faux-positifs. Ainsi, deux questions peuvent être formulées :

Existe-t-il une structure au sein de la collection de ressources génétiques *Juglans* de l'INRAE ?

Quels sont les liens de parenté entre les accessions de cette collection ?

Ainsi, 253 accessions de 14 espèces du genre *Juglans* ont été analysées à l'aide de 13 marqueurs microsatellites sélectionnés dans la littérature. La diversité génétique et la structure de la collection ont été étudiées.

D'une manière générale, les résultats de cette étude ont montré, pour l'ensemble des 253 accessions, un déficit en hétérozygotie pour certains marqueurs microsatellites et des spécificités alléliques pour les espèces apparentées au noyer cultivé, en lien avec la classification botanique. Aussi, les 10 marqueurs dérivés de *J. nigra* montrent une excellente transférabilité au sein de *J. regia*.

Les paramètres de diversité génétique classiquement utilisés en génétique des populations ont été calculés au sein de *J. regia*, l'espèce du noyer culitvé représentée par 217 accessions. Les 3 marqueurs EST-SSR ont notamment donné les valeurs d'hétérozygotie

observée, d'hétérozygotie attendue et de nombre moyen d'allèles les plus faibles. Par ailleurs, l'analyse de la structure du germplasm montre la présence de deux groupes selon l'origine géographique ; à savoir les accessions provenant du continent américain et d'Europe de l'Ouest d'une part, et celles d'Europe de l'Est et du continent asiatique d'autre part. Ces groupes d'abord déterminés à l'aide du programme STRUCTURE ont été confirmés par l'analyse de positionnement multidimensionnel et par la construction d'un arbre phylogénétique. De plus, le groupe provenant d'Europe de l'Est et du continent asiatique montre un niveau de diversité génétique plus important, traduit par des valeurs d'hétérozygoties observée et attendue supérieures à l'autre groupe.

L'analyse de chacun des deux groupes séparément, en retirant les accessions « admixed », montre également une répartition des accessions en sous-groupes en lien avec leur origine géographique. Il a également été possible de différencier les variétés traditionnelles françaises en fonction de leur bassin d'origine. Ainsi, les variétés originaires de Dordogne et de Corrèze, bassin Sud-Ouest, se séparent de celles originaires d'Isère, bassin Sud-Est.

Par ailleurs, une core collection de 50 accessions a été proposée en prenant en compte les données génotypiques et la connaissance préalable du phénotype. Cette core collection représente un outil potentiel pour le management des ressources génétiques du noyer cultivé à l'INRAE.

L'article suivant a été publié en ligne le 27 novembre 2018 dans le journal PLoS ONE, volume 13(11):e0208021, sous le nom de « Analysis of genetic diversity and structure in a worldwide walnut (*Juglans regia* L.) germplasm using SSR markers » (DOI 10.1371/journal.pone.0208021).

PLOS ONE

RESEARCH ARTICLE

Analysis of genetic diversity and structure in a worldwide walnut (*Juglans regia* L.) germplasm using SSR markers

Anthony Bernard[1,2], Teresa Barreneche[1], Fabrice Lheureux[2], Elisabeth Dirlewanger[1]*

1 UMR 1332 BFP, INRA, Université de Bordeaux, Villenave d'Ornon, France, **2** Ctifl, centre opérationnel de Lanxade, Prigonrieux, France

* elisabeth.dirlewanger@inra.fr

Check for updates

OPEN ACCESS

Citation: Bernard A, Barreneche T, Lheureux F, Dirlewanger E (2018) Analysis of genetic diversity and structure in a worldwide walnut (*Juglans regia* L.) germplasm using SSR markers. PLoS ONE 13 (11): e0208021. https://doi.org/10.1371/journal.pone.0208021

Editor: Zhengfeng Wang, Chinese Academy of Sciences, CHINA

Received: June 1, 2018

Accepted: November 10, 2018

Published: November 27, 2018

Copyright: © 2018 Bernard et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the Supporting Information files ((S10 Table).

Funding: The Ctifl, holder of the project "INNOV'noyer", and in partnership with INRA, want to thank the 'Région Nouvelle-Aquitaine' for funding a part of this project. The Ctifl also thank the 'Cifre' convention by 'ANRT' (Agence Nationale de la Recherche et de la Technologie) for funding half the cost of the doctoral fellowship of Anthony Bernard (2016/1558). To finish, the project is

Abstract

Persian or English walnut (*Juglans regia* L.), the walnut species cultivated for nut production, is one of the oldest food sources known and is grown worldwide in temperate areas. France is the 7th leading producer as of 2016 with 39 kt. Deciphering walnut genetic diversity and structure is important for efficient management and use of genetic resources. In this work, 253 worldwide accessions from the INRA walnut germplasm collection, containing English walnut and several related species, were genotyped using 13 SSR (Single Sequence Repeat) markers selected from the literature to assess diversity and structure. Genetic diversity parameters showed a deficiency of heterozygotes and, for several SSRs, allele-specificities among the accessions tested. Principal Coordinate Analysis (PCoA) showed the 253 accessions clustered in largely in agreement with the existing botanical classification of the genus. Among the 217 *J. regia* accessions, two main clusters, accessions from Eastern Europe and Asia, and accessions from Western Europe and America, were identified using STRUCTURE software. This was confirmed by Principal Coordinate Analysis and supported by Neighbor-Joining tree construction using DARwin software. Moreover, a sub-structure was found within the two clusters, mainly according to geographical origin. A core collection containing 50 accessions was selected using the maximum length sub-tree method and prior knowledge about their phenotype. The present study constitutes a preliminary population genetics overview of INRA walnut genetic resources collection using SSR markers. The resulting estimations of genetic diversity and structure are useful for germplasm management and for future walnut breeding programs.

Introduction

Persian walnut (*Juglans regia* L.) is a monoecious and dichogamous tree species [1] whose center of domestication is thought to be located in Central Asia [2]. Nowadays, Persian walnut is widely disseminated and grown in many temperate regions of Europe, North and South America, South Africa, Asia, Australia and New-Zealand [3]. The genus *Juglans* (order Fagales, family Juglandaceae) includes more than 20 diploid species, with 2n = 2x = 32 chromosomes [4].

labelled by the 'AgriSudOuest Innovation' competitiveness cluster.

Competing interests: The authors declare that they have no competing interests.

Formerly divided into four sections–*Trachycaryon* (butternut, a single species, *J. cinerea*), *Cardiocaryon* (heartnuts), *Rhysocaryon* (black walnuts, notably *J. nigra*) and *Dioscaryon* containing only *J. regia* [5]–it is now suggested that *J. cinerea* should be included in *Cardiocaryon* section [6]. According to Food and Agriculture Organization of the United Nations and International Nut and Dried Fruit Council data (www.fao.org, www.nutfruit.org, 2014 data), world-wide in-shell walnut production exceeds 3,400 kt. The five largest producers are China, USA, Iran, Ukraine and Chile,). France is the 7th largest producer with 39 kt. French walnut orchard area increased by nearly 19% between 2000 and 2010 to reach approximately 21,000 hectares in 2016, making it the most important French fruit crop other than apple. France exports 80% of its production in-shell, mainly to Europe where it is marketed in particular, thanks to the quality of its product. There are Protected Designations of Origin for the two main production areas: 'Noix du Périgord' and 'Noix de Grenoble'.

Knowledge of the genetic diversity of a species is crucial for effective management and use of its germplasm. Several types of molecular markers have been used previously for assessment of diversity and relationships in other walnut germplasm or populations. For example, isozymes were used to compare European and Asian walnut genotypes [7], and to study natural populations from Italy [8]. Restriction Fragment Length Polymorphism (RFLP) markers were used to assess the genetic diversity of the genus *Juglans* [6]. Randomly Amplified Polymorphic DNA (RAPD) markers were used by the breeding program of the University of Davis, CA to analyze parents and releases [9]. RAPDs were also used to characterized Iranian genotypes [10]. Inter-Simple Sequence Repeat (ISSR) were used [11] to show the genetic diversity of Greek natural populations was greater than a comparison set of international cultivars [12]. Amplified Fragment Length Polymorphism (AFLP) analysis of walnut genotypes from Turkey showed genotypes with low chill requirement had a narrow genetic base [13]. AFLPs also were used to assess the genetic diversity of natural populations from Kurdistan [14]. Use of Simple Sequence Repeats (SSRs) has grown considerably since 2002 and more than 20 publications were recently reviewed [15].

One of the uses of genetic resources is the choice of foundation material for an improvement program. In France, the Institut National de la Recherche Agronomique (INRA–French National Institute for Agricultural Research) led a walnut breeding program from 1977 to 2005. As a first step, many crosses were performed between French cultivars (such as 'Franquette', with a late bud break and a good fruit quality) and California cultivars (such as 'Pedro' and 'Chandler' for their lateral bearing habit). This resulted in the release of productive cultivars with good organoleptic qualities such as: 'Fernette','Fernor', 'Ferjean', 'Feradam', 'Ferbel', 'Ferouette' and 'Fertignac'. Following extensive walnut germplasm collecting in the Mediterranean area, Iran, Japan, and Central Asia, additional parents containing a wider genetic base were identified and used for breeding. Although several resulting hybrids appear promising, the current context of climate change and global competition make it crucial to initiate a new improvement program, utilizing the new technologies now available.

Germplasm constitutes an essential reservoir of allelic diversity for traits currently being exploited or not yet used in breeding programs, but its preservation and management is costly [16]. Hence, creation of a core collection representing the highest genetic diversity, while decreasing the number of accessions, is an effective way to reduce costs. Consequently, INRA walnut germplasm collection, which includes 253 accessions of worldwide origin maintained at the *Prunus* and *Juglans* Genetic Resources Center, was evaluated for its genetic diversity. The objectives of this study were; 1) to assess the genetic diversity and structure of INRA walnut collection and; 2) to define a core collection suitable for future walnut breeding programs and that could be more easily managed.

Materials and methods

Plant material

The INRA walnut germplasm collection is a result of important prospecting work performed between 1988 and 2000 in many countries around the world. 253 accessions from this collection were studied. Among them are 217 accessions of *J. regia*, including 194 cultivars and 23 intraspecific hybrids. An additional 36 accessions include 14 related species: *J. ailantifolia* Carr. (syn: *J. sieboldiana* Maxim.), *J. californica* S. Wats., *J. cathayensis* Dode, *J. cinerea* L., *J. hindsii* Jeps. *J. major* Heller, *J. mandshurica* Maxim., *J. microcarpa* Berl., *J. mollis* Engelm., *J. nigra* L., *J. pitteursii* C. Morren (isotype of *J. nigra*), *J. rupestris* Engelm. ex Torr. (isotype of *J. major*), *J. sieboldiana* Maxim., *J. sieboldiana* var. *cordiformis*. These related species accessions were added to the analysis to give a more global idea of the genetic diversity of the collection. Pictures of nuts of 11 *Juglans* species, part of the INRA walnut germplasm collection, are given in Fig 1.

DNA extraction and SSR genotyping

The *Prunus* and *Juglans* Genetic Resources Center gave us the permission to collect the plant material. Leaves were collected and sent to the BioGEVES laboratory in France for DNA extraction using a Macherey-Nagel NucleoSpin 96 Plant II Core kit. 2 μL of DNA extraction solution was used to measure DNA concentration by spectrophotometry (SPECTROstar Omega) and extract concentrations were standardized to 2.5 ng/μL.

All accessions were genotyped using 15 SSR markers chosen from the literature [17, 18, 19], after testing initially using a small sample of accessions. The markers are listed in Table 1 with primer sequences, repeat motif, source species and references. The forward primers were tailed by addition of a 35S 19 bp oligo sequence labeled to the 5′ end, using forward primers labeled with 6-FAM, NED, PET or VIC. The PCR reactions were carried out in 10 μL volumes containing 4 μL of diluted DNA (2.5 ng/μL) and 6 μL of PCR mix. The PCR mix consisted of 3.80 μL of ultra-pure water, 1 μL of 10X buffer, 0.2 μL of dNTP (10 mM), 0.60 μL of $MgCl_2$ (25 mM), 0.10 μL of Ampli Taq Gold polymerase, 0.10 μL of reverse primers (R), 1 μM of forward primers (F) and 10 μM of 35 S oligo sequence. PCR steps were: 10 min– 94˚C, (30 sec– 94˚C, 1 min– 58˚C, 40 sec– 72˚C) for 30 cycles (or 35 depending on the primers) and 10 min– 72˚C. Amplification products were diluted and 5.8 μL of formamide, 0.2 μL of ladder and 4 μL of amplicon were added in each well. The SSR genotyping of all individuals was performed on an ABI 3730 DNA Analyzer (Applied Biosystems), and allele sizing was identified using the GeneMapper™ 3.7 software (Applied Biosystems), using the LIZ 500 (Applied Biosystems) as an internal size standard. To confirm SSR allele size for samples with allele size differing by only one base, a pooling of the amplicons of the samples were analyzed a second time.

Assessment of genetic diversity and structure

Measures of genetic diversity, including the total number of observed alleles (A), the observed heterozygosity (H_o), the expected heterozygosity (H_e) and the inbreeding coefficient (F_{IS}) were estimated for each locus with the "adegenet 2.1.0" R package [20], for the 217 *J. regia* accessions only. The number of effective alleles (A_e) was estimated using GenAlEx 6.5 software [21, 22] and the polymorphism information content (PIC) was calculated using Cervus 3.0 software [23]

To determine the relationships between *Juglans* species, Principal Coordinate Analysis (PCoA) was used. PCoA is a distance-based model using jointly a dissimilarity matrix calculated with a simple-matching index and a factorial analysis. PCoA was performed using

Juglans spp.

1. *Dioscaryon* section : Common walnut

Juglans regia L. – English or Persian Walnut

2. *Rhysocaryon* section : Black walnut

Juglans hindsii Jeps. – Northern California Black Walnut

Juglans major Heller – Arizona Black Walnut

Juglans microcarpa Berl. – Texas or Little Black Walnut

Juglans mollis Engelm. – Mexican Walnut

Juglans nigra L. – Eastern Black Walnut

3. *Cardiocaryon* section

Juglans cathayensis Dode – Chinese Walnut

Juglans cinerea L. – Butternut

Juglans mandshurica Maxim. – Manchurian Walnut

Juglans sieboldiana Maxim. – Japanese Walnut

Juglans sieboldiana var. *cordiformis* – Heartnut

1 cm

Fig 1. Nuts of 11 *Juglans* species available in INRA walnut germplasm collection. These *Juglans* accessions come from 23 countries covering three continents: Europe (France, Spain, Greece, Ukraine, etc.), America (Canada, USA and Chile) and Asia (China, Japan, Iran, etc.). All accessions are 20–30 year-old grafted trees on various rootstocks located in the Fruit Tree Experimental Unit (UEA) of INRA in Toulenne (latitude 44˚34'37.442"N–longitude 0˚16'51.48"W), near Bordeaux (France). The list of accessions studied is shown in S1 Table.

https://doi.org/10.1371/journal.pone.0208021.g001

DARwin 6.0.14 software (Dissimilarity Analysis and Representation for Windows) [24] supplemented by "scatterplot3d" R package for 3D visualization, to determine the relationships between the different *Juglans* accessions.

PCoA was used to investigate the patterns of structure among the 217 *J. regia* accessions and Bayesian model-based analyses was implemented using the software package STRUCTURE 2.3.4 [25]. The following steps were performed to identify the best number of clusters (K): twenty runs of STRUCTURE were done by setting K from 1 to 10. Each run consisted of a length of burn-in period of 5,000 followed by 50,000 Markov Chain Monte Carlo (MCMC) replicates, assuming an admixture model and correlated allele frequencies. When K was estimated, ten runs of STRUCTURE were done by setting the K from 1 to 5. Each run consisted of a length of burn-in period of 100,000 followed by 750,000 MCMC replicates, assuming an

Table 1. List of the 15 SSR markers selected.

SSR marker	SSR type and source species	Primer sequence 5'-3'	Repeat motif	Reference
WGA 001	genomic SSR from *J. nigra*	F ATTGGAAGGGAAGGGAAATG R CGCGCACATACGTAAATCAC	$(GA)_5GCA(GA)_3GCA(GA)_3$	Dangl et al., 2005
WGA 004	genomic SSR from *J. nigra*	F TGTTGCATTGACCCACTTGT R TAAGCCAACATGGTATGCCA	$(GT)_5,(GA)_{15},(GA)_{11}$	Woeste et al., 2002
WGA 009	genomic SSR from *J. nigra*	F CATCAAAGCAAGCAATGGG R CCATTGCTCTGTGATTGGG	$(GA)_{16}$	Dangl et al., 2005
WGA 027	genomic SSR from *J. nigra*	F AACCCTACAACGCCTTGATG R TGCTCAGGCTCCACTTCC	$(GA)_{30}$	Woeste et al., 2002
WGA 069	genomic SSR from *J. nigra*	F TTAGTTAGCAAACCCACCCG R AGATGCACAGACCAACCCTC	$(GA)_4ATATAA(GA)_{16}$	Woeste et al., 2002
WGA 072	genomic SSR from *J. nigra*	F AAACCACCTAAACCCTGCA R ACCCATCCATGATCTTCCAA	$(CT)_{14}$	Woeste et al., 2002
WGA 202	genomic SSR from *J. nigra*	F CCCATCTACCGTTGCACTTT R GCTGGTGGTTCTATCATGGG	$(GA)_{11}$	Dangl et al., 2005
WGA 276	genomic SSR from *J. nigra*	F CTCACTTTCTCGGCTCTTCC R GGTCTTATGTGGGCAGTCGT	$(GA)_{14}$	Dangl et al., 2005
WGA 349	genomic SSR from *J. nigra*	F GTGGCGAAAGTTTATTTTTTGC R ACAAATGCACAGCAGCAAAC	$(CT)_{14}$	Dangl et al., 2005
WGA 376	genomic SSR from *J. nigra*	F GCCCTCAAAGTGATGAACGT R TCATCCATATTTACCCCTTTCG	$(AG)_2AA(AG)_6$	Dangl et al., 2005
JR 0160*	EST-SSR from *J. regia*	F TCTCGGATTTGGGCTGTGAC R TCCGGGACCCTCGTCTAATT	$(TC)_{10}$	Dang et al., 2016
JR 1739*	EST-SSR from *J. regia*	F GGATGTGGAGACGGCAAAGA R CGTCCACCCAAACCAAGAGA	$(GAGCCG)_8$	Dang et al., 2016
JR 1817	EST-SSR from *J. regia*	F CCTCAGAGCCAACCATCCTT R AGAACAGAACCAGCGTCACA	$(AC)_{11}$	Dang et al., 2016
JR 6160	EST-SSR from *J. regia*	F ACTTCAGGTTCCCAACGCAA R TAGAGGGAAGGTCTCCGGTG	$(GA)_{10}$	Dang et al., 2016
JR 6439	EST-SSR from *J. regia*	F TCGATGCGATCATCTCCGTG R CGGCACCAAAACAGAACTCG	$(TGCG)_5$	Dang et al., 2016

F: Forward, R: Reverse

*: not retained for the analysis

https://doi.org/10.1371/journal.pone.0208021.t001

admixture model and correlated allele frequencies. For the choice of the most likely K, the plateau criterion described by Pritchard et al. [25] and the ΔK method described by Evanno et al. [26], implemented in STRUCTURE harvester [27], were used. Then, the run having the highest likelihood estimate to assign cluster proportions to individuals was used. Accessions with an estimated membership below 0.8 were assigned to the "admixed group". Then, STRUCTURE was run in a second step on fragmented datasets to discover if lower levels of structure exist and K was tested from 1 to 10 for each cluster (ten runs, 5,000 burn-in period, 50,000 MCMC replicates) and when K was estimated, ten runs were done by setting the K from 1 to 5 (100,000 burn-in period, 750,000 MCMC replicates). The ΔK method described by Evanno et al. [26], implemented in STRUCTURE harvester [27], was used.

In addition, the pairwise fixation indexes (F_{ST}) between clusters were estimated with the "diveRsity 1.9.90" R package [28]. The genetic relationships between *J. regia* accessions was also assessed by the Neighbor-Joining method [29] using DARwin 6.0.14 software. Dissimilarities were calculated with 10,000 bootstraps, transformed into Euclidean distances using power transformation, and the UnWeighted Neighbor-Joining method was used to build a tree for *J. regia* accessions.

Core collection establishment

Core collections are subsamples of germplasm collections created in order to decrease the number of accessions while still representing the maximum genetic diversity of the larger collection. The function 'maximum length sub tree' of DARwin 6.0.14 software [24] was used to create a *J. regia* core collection based on dissimilarities calculated as described before. This is a stepwise procedure that prunes successively redundant accessions, allowing the choice of sample size that retains the largest genetic diversity. Putative clusters of synonymous accessions were found using the function 'removed edge value' provided by the Neighbor-Joining tree with a threshold value of 0.0005. A core collection must also account for other knowledge available such as the adaptation of an accession to a particular environment or its interest in terms of other phenotypes. In the 'maximum length sub tree' function, the 'excluded' and 'forced' options allow excluding or forcing inclusion of an accession manually. This option was used to keep several reference accessions in the final core collection.

Results

SSR genotyping and genetic diversity analysis

According to the amplification control, the strength of recorded signal and the number and quality of true peaks, 13 SSR markers among the 15 were selected for analysis of the 253 accessions. In addition, because each marker was carefully developed and tested, we assumed that the missing data are in fact null alleles. For several SSRs, PCR amplification was observed only for a portion of the *Juglans* species tested (S2 Table). The WGA 001 and WGA 276 SSRs amplify in *J. regia* and species of section *Rhysocaryon* but not in section *Cardiocaryon* (S2 Table). The JR 1817 SSR does not amplify in *Cardiocaryon* or in *J. californica* but does amplify in the other *Rhysocaryon* and in *J. regia*. The JR 6160 SSR also does not amplify in *J. californica* (S2 Table). Furthermore, for some SSRs, private alleles are observed in some species. For example, for WGA 001 the '184 bp' allele is found only in *J. mollis* and the '197 bp' and '205 bp' alleles are present exclusively in *J. nigra*. For WGA 009, the '247 bp' allele is only detected in *J. regia*, in *Cardiocaryon* species, and in *J. nigra* and *J. pitteursii* of section *Rhysocaryon* and for WGA 069, the '180 bp' allele is found only in *J. regia* and section *Cardiocaryon*. Finally, for WGA 376 the '236 bp' allele is present only in *J. regia* and section *Rhysocaryon* whereas the '244 bp' allele is present only in section *Cardiocaryon*.

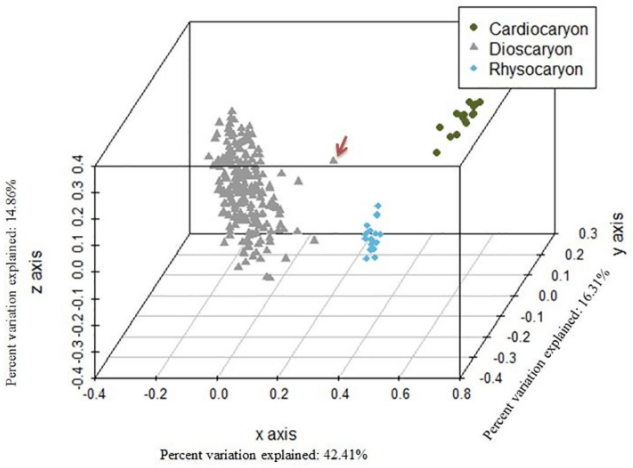


Fig 2. Principal Coordinate Analysis (PCoA) 3D plot of the 253 *Juglans* accessions using 13 SSRs. Blue cluster shows the 18 accessions belonging to the *Rhysocaryon* section while the green cluster shows the 18 accessions belonging to the *Cardiocaryon* section, both part of the *Juglans* genus. In grey, *J. regia* accessions (*Dioscaryon* section). The appearing distant *J. regia* accession marked with a red arrow is the accession 'PI 15 95 68' (ID 207).

https://doi.org/10.1371/journal.pone.0208021.g002

PCoA is a method that explores dissimilarities of data through a dissimilarity matrix and assigns each accession a location in a low-dimensional space. The PCoA results reveal three well-separated clusters among INRA *Juglans* germplasm collection (Fig 2). The largest cluster contains the 217 *J. regia* accessions. A second cluster consists of the *Cardiocaryon* section and a third contains the *Rhysocaryon* species. The *Cardiocaryon* cluster includes 18 accessions of *J. mandshurica*, *J. cathayensis*, *J. sieboldiana*, *J. sieboldiana* var. *cordiformis*, and *J. cinerea*. The *Rhysocaryon* cluster includes 18 accessions of *J. mollis*, *J. nigra*, *J. microcarpa*, *J. californica*, *J. hindsii* and *J. major*.

Among the 217 *J. regia* accessions, the number of alleles per locus (A) ranges from 2 to 17 with an average of 8.92 and the number of effective alleles ranges from 1.18 to 4.61 with an average of 2.84 (Table 2). The observed heterozygosity (H_o) ranges from 0.10 to 0.68 with a mean value of 0.47. The expected heterozygosity H_e is higher than H_o and ranges from 0.15 to 0.78 with a mean value of 0.56 (p-value of *t*-test: $8.24E^{-04}$). The average polymorphism information content (PIC) is close to that of H_e. JR 1817 and JR 6439, both SSRs derived from EST, have the lowest values of H_o, H_e, A_e and PIC. Considering all the accessions including related species, the number of alleles per locus (A) is larger at each locus, ranging from 7 to 32 with an average of 17.31.

Structure analysis

Analysis of structure infers accession ancestry from genotypic information. The most likely number of clusters was evaluated considering the plateau criterion and using the ΔK method. The highest value was for K = 2 (S3 Table; S1 Fig), indicating two clusters: the first containing 63 accessions (referred to as the 'Eastern Europe and Asia' cluster). The second comprises 127 accessions (the 'Western Europe and America' cluster) (Fig 3; S4 Table). An additional 27 accessions clearly showed mixed ancestry (membership values lower than 80% in either of these two clusters). In this admixed cluster, 17/27 accessions are INRA hybrids and USA cultivars resulting from intraspecific crosses ('Feradam', 'Fernette', 'Serr', 'Chico', 'Amigo', 'Gillet', 'Forde' and 'Tulare'). The two accessions with purple leaves are also in this admixed group, as well as other accessions, including 'Lara'.

Table 2. Genetic diversity estimations of the 217 *J. regia* accessions.

SSR markers	A	A_e	H_o	H_e	PIC	F_{IS}
WGA 001	8	3.50	0.65	0.71	0.67	0.08
WGA 004	7	1.99	0.46	0.50	0.42	0.08
WGA 009	7	2.93	0.59	0.66	0.62	0.11
WGA 027	2	1.87	0.43	0.46	0.36	0.07
WGA 069	9	4.25	0.57	0.76	0.73	0.25
WGA 072	6	1.73	0.32	0.42	0.37	0.24
WGA 202	17	4.61	0.68	0.78	0.75	0.13
WGA 276	16	4.15	0.63	0.76	0.74	0.17
WGA 349	11	4.59	0.49	0.78	0.75	0.37
WGA 376	13	2.95	0.61	0.66	0.62	0.08
JR 1817	5	1.18	0.10	0.15	0.15	0.33
JR 6160	10	1.96	0.43	0.49	0.42	0.12
JR 6439	5	1.21	0.14	0.17	0.17	0.18
Min	2	1.18	0.10	0.15	0.15	0.07
Mean	8.92	2.84	0.47	0.56	0.52	0.17
Max	17	4.61	0.68	0.78	0.75	0.37

A: number of different alleles, A_e: number of effective alleles, H_o: observed heterozygosity, H_e: expected heterozygosity, PIC: polymorphism information content, F_{IS}: inbreeding coefficient

https://doi.org/10.1371/journal.pone.0208021.t002

The 'Eastern Europe and Asia' cluster contains accessions from Bulgaria ('Sheinovo' and 'Izvor 10'), Romania ('Sibisel 39', 'Sibisel 44', 'VL4B' and 'Germisara'), Greece ('S 28 A Achille', 'S 4 B Thétis', 'S 34 B Pyrrus' and 'EAA 6'), Hungary ('Milotai n˚10'), Ukraine ('UK-series'), India ('Sopore'), Iran ('IR-series' and 'Z 53'), China ('Jin Long 1' and 'Lu Guang'), Japan ('Shinrei'), and surprisingly, France (some INRA hybrids), Switzerland ('RA 1223') and USA ('Sexton', 'Chase C7', 'Wepster W2', 'Adams 10' and the 'PI series').

The 'Western Europe and America' cluster contains accessions from France (old cultivars such as 'Lub', 'Franquette', 'Saint Jean', 'Marbot', 'Bijou', modern cultivars obtained from intraspecific crosses such as 'Fernor', 'Ferjean', 'Ferbel', 'Ferouette', 'Fertignac', and INRA hybrids), Germany ('Allemagne 139' and 'Geisenheim 286'), England ('Northdown Clawnut 252'), Spain ('MB- and MBT-series', 'Del Carril' and 'Gran Jefe'), Greece ('S 1 A Diane' and 'S 1 B Ariane'), Portugal ('Rego'), Switzerland (such as a laciniate walnut), Hungary ('M 10–37'), Chile ('AS 1') and USA (old cultivar 'Payne' and modern cultivars obtained from intraspecific crosses such as 'Howe', 'Pedro', 'Vina', 'Chandler', etc.).

Moreover, genetic diversity parameters show a lower level of diversity among the 'Western Europe and America' cluster (S7 Table) and pairwise F_{ST} calculation between the two clusters is higher than that between each of the two clusters and the admixed group (S8 Table). Among the 217 *J. regia* accessions, a higher number of private alleles was observed in the 'Eastern Europe and Asia' cluster compared to the number in the 'Western Europe and America' cluster (61 alleles vs 5 alleles) (S9 Table).

As the ΔK method often detects the uppermost level of structure of a collection, each cluster was analyzed independently to explore if a lower structure could be found within each group. The two separated datasets comprised 63 accessions coming from 'Eastern Europe and Asia' cluster, and 127 accessions coming from 'Western Europe and America' cluster. The 27 accessions considered as admixed were excluded from further analyses. Within each cluster, STRUCTURE identified three subclusters (Fig 4). Among the subclusters of 'Eastern Europe and Asia', 1–1 contains 29/31 accessions of the 'UK-series' from Ukraine and one accession

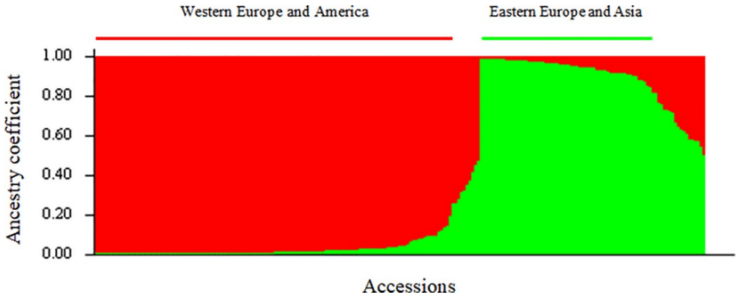


Fig 3. Structure of the *J. regia* collection using STRUCTURE software. Bar plot of individual ancestry proportions for the genetic clusters inferred using K = 2 and the dataset of 217 individuals and 13 SSR loci. Individual ancestry proportions (Q values) are sorted within each cluster. 'Eastern Europe and Asia', and 'Western Europe and America' clusters, are shown in green and red, respectively.

https://doi.org/10.1371/journal.pone.0208021.g003

from Romania ('VL4B'). Subcluster 1–2 contains one Greek accession ('EAA 6'), one Chinese accession ('Jin Long 1'), the Japanese accession ('Shinrei'), and unexpectedly a French INRA's hybrid and three accessions from USA ('Chase C7', 'Wepster W2' and 'Adams 10'). Subcluster 1–3 contains accessions from Asia including Chinese accession 'Lu Guang' and Indian accession 'Sopore', accession 'RA 1223', the two other accessions (2/31) from Ukraine, and several

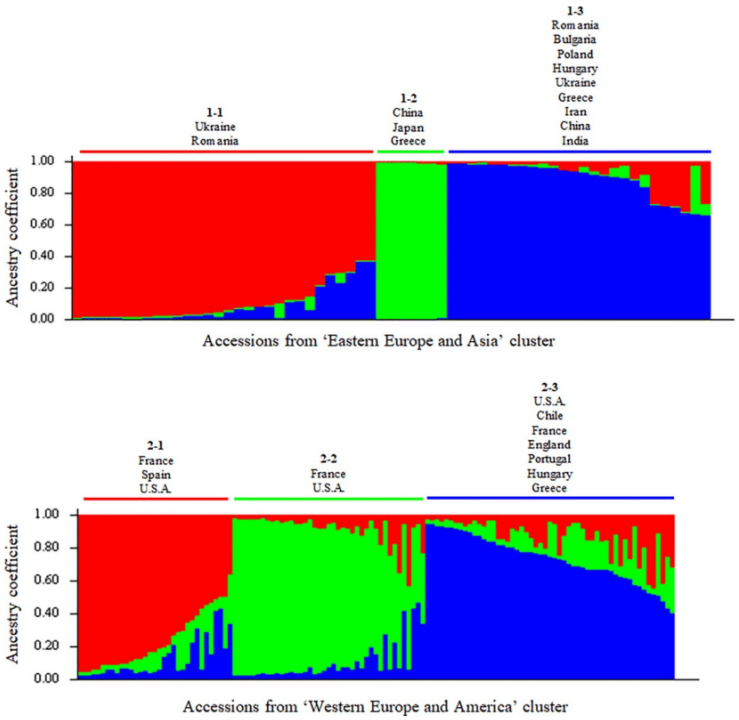


Fig 4. Substructure of the 'Eastern Europe and Asia' and 'Western Europe and America' clusters using STRUCTURE software. Bar plot of individual ancestry proportions for the genetic clusters inferred, using K = 3 and the dataset of 63 individuals and 13 SSR loci, for the 'Eastern Europe and Asia' cluster at the top. Bar plot of individual ancestry proportions for the genetic clusters inferred, using K = 3 and the dataset of 127 individuals and 13 SSR loci, for the 'Western Europe and America' cluster at the bottom. Individual ancestry proportions (Q values) are sorted within each cluster. Substructure within each of the two large groups follows a geographical origin, based on the supposed origin of the accessions.

https://doi.org/10.1371/journal.pone.0208021.g004

accessions from eastern European countries such as Greece, Romania, Bulgaria, Poland and Hungary. (S5 Table; S2 Fig). Subcluster 2–1 of the 'Western Europe and America' cluster includes old French cultivars such as 'St Jean', 'Corne', 'Grosvert', 'Bijou' and 'Verdelet', accession 'Lieb Mayette', all the Spanish accessions, some accessions from USA and several others including the laciniate walnut. Subcluster 2–2 includes other old French cultivars ('Meylannaise', 'Franquette' and 'Mayette') and accessions from USA ('Pedro', 'Midland', 'Vina', 'Cisco', 'Chandler' and 'Carmelo') and subcluster 2–3 includes other old French cultivars ('Lozeronne', 'Parisienne' and 'Marbot'), accessions from USA ('Tehama' and 'Waterloo'), Chilean accession 'AS 1', and accessions from England, Portugal, Hungary and Greece (S6 Table; S3 Fig).

PCoA performed on the set of 217 *J. regia* accessions (Fig 5) showed very similar clustering to those obtained with STRUCTURE. Each cluster again separated distinctly although the 'Eastern Europe and Asia' accessions were more spread, suggesting a greater genetic diversity.

Results obtained with STRUCTURE for *J. regia* accessions generally also aligned with assessment using the Neighbor-Joining method (Fig 6). The main branching groups of the tree agree with STRUCTURE results with K = 2. Several accessions exhibit longer branch length, indicating a greater genetic diversity. These include accessions 'PI 15 95 68' (ID 207), 'Sopore' (ID 181), 'Jin Long 1' (ID 95) and 'UK 239–23' (ID 146), which belongs to 'Eastern Europe and Asia' cluster. This finding is in agreement with the potential greater level of genetic diversity of the accessions of 'Eastern Europe and Asia' cluster found using PCoA.

Core collection establishment

The Neighbor-Joining tree based on the dissimilarity matrix between the 217 *J. regia* accessions (Fig 6) allowed construction of a core collection. The function 'maximum length sub tree' of DARwin 6.0.14 software was used iteratively to eliminate the putatively synonymous and the most redundant accessions. Additionally, some accessions were fixed based on knowledge of the material, using the DARwin 'forced' option. This revision of the Neighbor-Joining tree is shown in Fig 7.

This study led to the choice of 50 accessions for a core collection (S4 Table). Among these, 11 were fixed selections and six were retained for their phenotypic traits or adaptability. This set included 'Chandler', the most important cultivar in California and also cultivated in France,

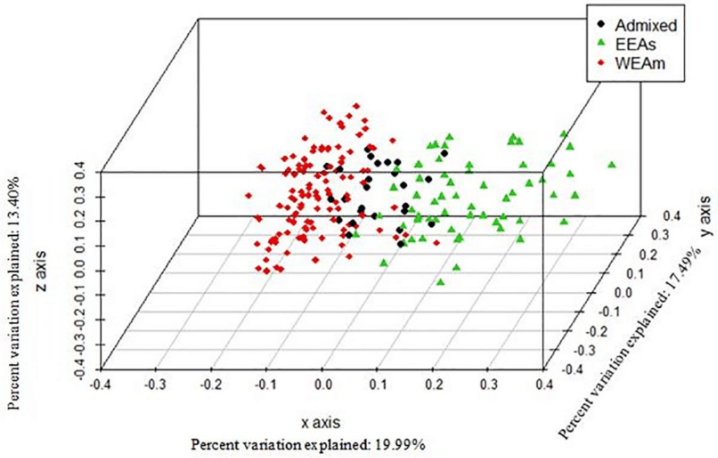


Fig 5. Principal Coordinate Analysis (PCoA) 3D plot of the 217 *J. regia* accessions from the INRA walnut germplasm collection using 13 SSRs. Colors indicate STRUCTURE clusters K = 2: red for 'Western Europe and America' (WEAm) cluster, green for 'Eastern Europe and Asia' (EEAs) cluster, and black for admixed accessions.

https://doi.org/10.1371/journal.pone.0208021.g005

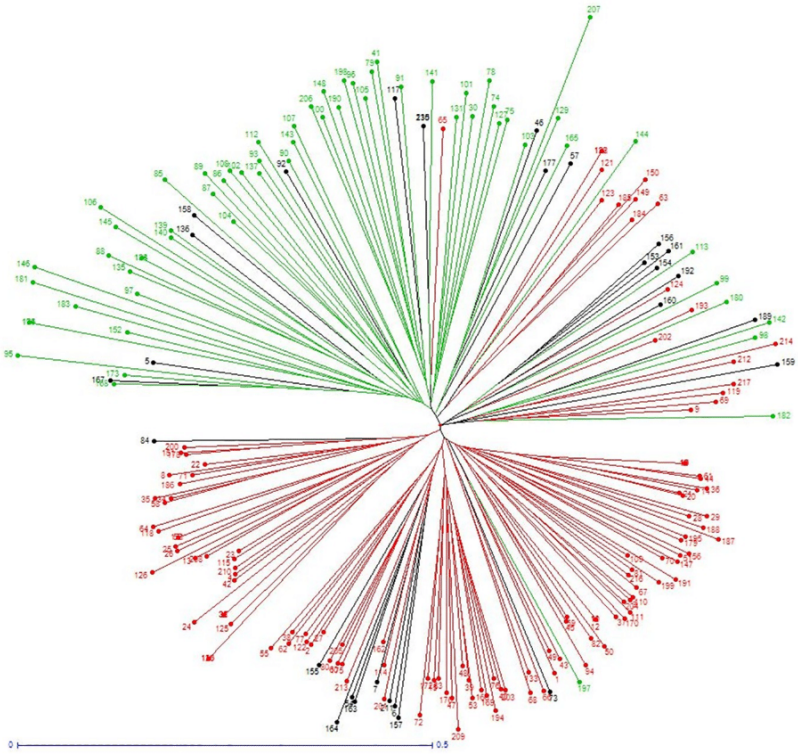


Fig 6. Neighbor-Joining tree based on dissimilarity matrix calculated from the dataset of 13 SSRs across the 217 *J. regia* accessions from the INRA's walnut germplasm collection. Colors reflect STRUCTURE clusters K = 2: red for 'Western Europe and America' cluster, green for 'Eastern Europe and Asia' cluster, and black for admixed accessions.

https://doi.org/10.1371/journal.pone.0208021.g006

'Early Ehrhardt', the earliest to break bud, 'Bijou', with the biggest fruits, 'Franquette' and 'Lara', among the most planted cultivars in French orchards, and 'Rouge de la Réole', for its red kernels. Five additional accessions were selected because they belong to the two French Protected Designations of Origin for the two main production areas: 'Noix du Périgord' cultivars 'Grandjean', 'Corne' and 'Marbot', and 'Noix de Grenoble' cultivars 'Mayette' and 'Parisienne'. With the exception of 'Lara', which is admixed, all these accessions are part of the 'Western Europe and America' cluster. In Fig 7 they form the red cluster from ID 33 to ID 202.

The remaining 39 accessions include only four accessions from the 'Western Europe and America' cluster, in particular a laciniate leafed walnut and the Greek accession 'S 1 B Ariane'. The remaining 35 are part of the 'Eastern Europe and Asia' cluster, and include 17 'UK-series', the three 'PI-series', five Iranian genotypes, Chinese cultivar 'Jin Long 1', the Greek 'EAA 6', the Indian 'Sopore', 'Germisara', 'Sibisel 44' and 'Sibisel 39' from Romania, 'Izvor 10' and 'Cheinovo' from Bulgaria, 'Sexton', and an ornamental tree.

Discussion

Genetic diversity of the INRA walnut germplasm collection

This study provides the first overview of genetic diversity in a large germplasm collection of walnut. Although WGA SSRs are derived from *J. nigra*, they have good transferability to *J. regia*. The two EST-SSRs derived from *J. regia*, JR 1817 and JR 6439, have the lowest H_o and H_e

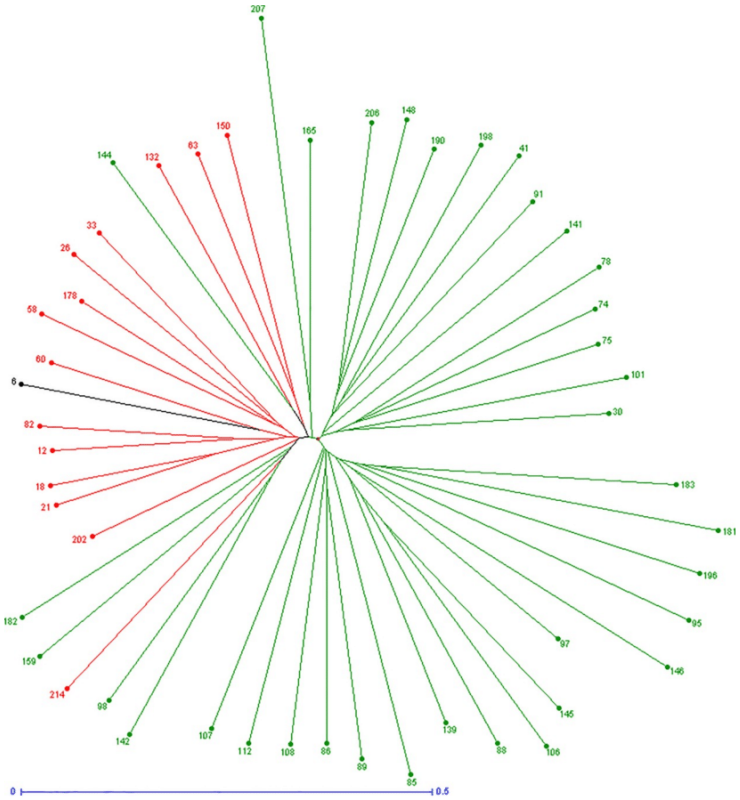


Fig 7. Neighbor-Joining tree of the accessions chosen to construct the INRA walnut germplasm core collection. Colors correspond to STRUCTURE clusters K = 2: red for the 'Western Europe and America' cluster, green for the 'Eastern Europe and Asia' cluster, and black for admixed accessions.

https://doi.org/10.1371/journal.pone.0208021.g007

values. EST-SSR polymorphism reflects the genetic diversity in coding sequences. The low polymorphism observed may be due to selection against alteration in conserved coding sequences. This was also observed in grape [30], cucumber [31] and various other plant species [32].

The three clusters revealed by PCoA (Fig 2) closely follow the botanical classification of the species of genus *Juglans* as previously observed using RFLPs [6] and nuclear DNA sequences [33]. In this study, section *Rhysocaryon* seems to be closer to *Juglans regia* than is the *Cardiocaryon* section. This contrasts with other evidence *Cardiocaryon* is the section most closely related to *J. regia* [3]. This could be a result of the small number of accessions of these two sections included for analysis or to unsuccessful amplification of SSRs WGA 001 and WGA 276 in the *Cardiocaryon* section.

Expected heterozygosity (H_e) of the 217 *J. regia* accessions was higher than observed heterozygosity (H_o), showing a deficiency of heterozygotes that suggests the presence of pedigree inbreeding. Average H_o (0.47) and average H_e (0.56) are similar to those found in other studies on walnut natural populations or germplasm collections, with a lower H_o [34–44] and indicates a potential structure within the INRA walnut germplasm collection.

Structure and genetic relationships of the INRA walnut germplasm collection

The diversity of the INRA *J. regia* germplasm collection, in terms of geographical origin, is a result of international cooperation and collections by Eric Germain, the former INRA walnut

breeder. He travelled extensively and collected diverse plant materials from both natural populations and the collections of foreign breeding programs. Regarding the structure of the collection, the Bayesian model-based STRUCTURE method and the PCoA gave similar results, as was also observed in coconut data [45]. The INRA collection consists of two well-differentiated clusters, as is often reported in other plant germplasm collections, for example, soybean, cherry, maize, apple and coffee [46–50]. The two genetic clusters of the INRA walnut collection fit well with the geographical origin of the accessions. Results were consistent with Pollegioni et al. [51] regarding the history of *J. regia* in Europe. That work showed that the genetic structure of 91 walnut populations in Eurasia separated into two main clusters, Western Europe (Italy, Spain, France, Slovakia and Hungary) and Eastern Europe and Asia (Romania, Greece, Turkey, Iran and the Himalayas). Similar results were found for sweet chestnut, with a strong differentiation between trees from the Northern Iberian Peninsula, and Central and Southern Iberian Peninsula, as well as in natural populations [52], and among European cultivars [53, 54]. All walnut accessions in the 'admixed group' are known hybrids. It includes the French variety 'Lara', the result of a complex cross: ['Hartley' × 'Payne'] × ['PI 18256' × 'Conway Mayette']. The nine 'admixed' INRA hybrids, for example 'Fernette', have 'Lara' in their pedigree. The six other 'admixed' modern cultivars are the American accessions 'Serr' (a cross between 'Payne' and 'PI 15 95 68' from Afghanistan), 'Chico' and 'Amigo' (resulting both from a cross between 'Sharkey', maybe with Chinese pedigree, and 'Marchetti'), 'Tulare' (with 'Serr' parentage), and 'Gillet' and 'Forde' (both with 'Chico' parentage). Therefore, it is not surprising to obtain lower values of pairwise F_{ST} between the 'admixed' and 'Western Europe and America' clusters, and between the 'admixed' and 'Eastern Europe and Asia' clusters. Interestingly, the unique Israeli accession 'Kfar Hanania' shares its genome with both clusters at 50%.

The genetic diversity parameters H_e, H_o, F_{IS} and the number of alleles are lower among the 127 'Western Europe and America' accessions than among the 63 of the 'Eastern Europe and Asia' cluster (S7 Table). This is in agreement with the fact that walnut was domesticated in Central Asia [2], and that regions of origin are expected to contain larger genetic diversity. In addition, these results show that the accessions of the 'Western Europe and America' cluster, in particular those from France and USA, might in fact be more related than expected under a model of random mating. Furthermore, the pairwise fixation index (F_{ST}) calculated between the two clusters shows a clear level of differentiation, with a value of 0.101 (S8 Table). A search of private alleles also shows a greater allelic diversity in the 'Eastern Europe and Asia' cluster even though there are half as many individuals in this cluster as in the other (S9 Table). Investigation of substructure among each of the two main clusters gives a good understanding of the plant material. Within the 'Eastern Europe and Asia' cluster, the subclustering also follows the geographical origins. Accessions of the 'UK-series' (subcluster 1–1, Fig 4), from a likely center of origin, were well separated from those of Eastern Europe, Iran and India (1–3). A small subcluster (1–2) contained both accessions from the Far East and Greek accession 'EAA 6'. According to the plant material register of E. Germain, this accession was collected in Central Greece but it seems to be different from other Greek accessions. In the same way, the subclustering within the 'Western Europe and America' group follows the geographical origin of accessions. Spanish accessions (2–1) are separated from Greek accessions (2–3). French accessions, highly represented in this study, are scattered throughout the three subclusters accordingly to their regional origin. For example, the landraces 'Corne', 'Grosvert', 'Verdelet' and 'Saint Jean' belong to the 2–1 group. These are from the Dordogne or Corrèze departments in Nouvelle-Aquitaine region (western France), one of the main French walnut production regions, whereas the landraces 'Meylannaise', 'Mayette' and 'Franquette', which are part of 2–2 group, come from the Isère department in the Auvergne-Rhône-Alpes region (eastern France), another important region of walnut production. The 2–3 group consists of other French

landraces including 'Candelou' from the Lot department in the Occitanie region. So, even among a relatively small number of accessions, the French landraces show a level of genetic diversity corresponding to their geographical origin.

The structure observed in INRA walnut collection provides interesting insights into source materials and the pedigrees of hybrids and cultivars developed by past breeding programs, and provides information for the better use of germplasm collections, including for the choice of parents, for the new French walnut breeding program in particular [15]. For example, the INRA hybrids H 91–88 ('Franquette' × 'Payne'), H 93–63 ('Franquette' × 'Pedro'), and H 99–104 ('Franquette' × 'Chandler') and their progenitors belong to the 'Western Europe and America' cluster, along with all old French cultivars. Conversely, USA accessions 'Chase C7', 'Wepster W2' and 'Adams 10' belong to the 'Eastern Europe and Asia' cluster, probably because they are selections of 'Manregian-type' walnuts. 'Manregian' walnuts trees originate from seed collected in northeastern China. Accessions 'PI 2 657 12' and 'PI 15 95 68', respectively from Russia and Afghanistan, also belong to the 'Eastern Europe and Asia' cluster. Indeed, in INRA walnut collection, we have the information of the source country which provided us the accessions. But the source country could be different from the real geographical origin of the accessions and this work gives us new information on the plant material. Furthermore, even if it is important to note that the INRA walnut collection includes a lot of French and American selections or hybrids, for some they derive from crosses using parents from various countries of origin.

Some INRA *J. regia* accessions show interesting or unusual traits such as weeping branches or laciniate leaves and have been used for ornamental purposes. There are also four accessions of 'red walnuts'. These have red pellicles. Pellicle coloration in general is an important commercial trait. Lighter kernels command higher prices but at a time when consumers are looking for diversity on their plates, such as the black tomato, red walnuts could become a new trend. The genetic determinant of kernel pellicle coloration was identified recently [55, 56]. In this study, four accessions 'Robert Livermore' (ID 200), 'Rouge de la Donau' (ID 185), 'Rouge de la Réole' (ID 178) and 'Rouge de Laquenexy' (ID 179) were studied. Results found with STRUCTURE K = 2 and DARwin show a high degree of similarity between 'Robert Livermore' and 'Rouge de la Réole', explained by the fact that 'Robert Livermore' is derived from a controlled cross between 'Howard' and 'Rouge de la Réole', the latter having been introduced under the name 'UC86-11' into the California germplasm [57]. Moreover, according to the Neighbor-Joining tree, the two other accessions seem to be genetically distant, probably due to their assumed origin: 'Rouge de la Donau' from Austria and 'Rouge de Laquenexy' that originated close to Metz, near the German border. The nature of available diversity is interesting if future selection for this trait is needed.

Core collection

Results of the Neighbor-Joining method supported those found with STRUCTURE and PCoA. The constructed tree identified 18 putative synonyms. These include the three 'Corne' accessions and three 'Parisienne' accessions. A core collection should conserve the maximum phenotypic and genotypic variation in as few accessions as possible [58], avoiding redundant entries and including all relevant geographical regions [59]. This study proposes the first core collection for INRA walnut germplasm collection based on genotypic variation and additional knowledge of the plant material.

The 'force' functionality in DARwin was used for the initial selection step to include 11 accessions with interesting phenotypic variability such as early budbreak date, red kernel, or good adaptation to French growing conditions such as the cultivars belonging to Protected

Designations of Origin. These accessions, containing a relatively low level of genetic diversity, are part of the 'Western Europe and America' cluster (except 'Lara' which is admixed).

The second step selected 39 accessions with maximum of genetic diversity, including four from the 'Western Europe and America' cluster and 35 from the 'Eastern Europe and Asia' cluster. Notable were 'UK-series' accessions from a probable center of origin, and the three 'PI-series' from Poland, Afghanistan and Russia [60]. Also included were accessions from Iran, China, Greece, India, Romania and Bulgaria. Additional selections for the core collection were 'Sexton', with Chinese parentage, and a tree with large leaves that could be interesting for ornamental use. Seventy percent of the core collection is comprised of accessions from the 'Eastern Europe and Asia' cluster, reflecting the greater diversity available in this cluster.

Conclusions

This study used SSR markers to assess the genetic diversity and structure of the INRA walnut germplasm collection. Resulting knowledge and the core collection constructed will be useful for rational and economically sustainable management of INRA collection. In addition, this information will provide a valuable and fruitful tool for walnut breeders in selecting new cultivars and identifying promising parents for crossbreeding. A highly promising hybrid from the second INRA improvement program, which is currently under evaluation, illustrates this point well. This hybrid is a cross between 'Fernor' and 'Shinrei' from 'Western Europe and America' and 'Eastern Europe and Asia' clusters respectively and demonstrates the efficiency of choosing parents based on their cluster membership. Use of the knowledge developed in this study should increase the efficiency of the new French breeding program in addressing challenges to the walnut industry such as adaptation to climate change and resistance to new pests and diseases, while also increasing yield and kernel quality. This work also highlights the importance of collecting germplasm in, or close to, countries of the crop's center of origin.

In addition, assessment of the genetic diversity of INRA walnut collection could be used to define additional core collections suitable for further analysis, such as a genome-wide association study. These collections would be characterized by unique and great phenotypic variability specifically regarding particular traits of interest such as early budbreak and bloom dates, nut and kernel quality, and tolerance to pests and diseases.

Supporting information

S1 Fig. Detection of the number of populations K using plateau criterion (Pritchard et al., 2000) and ΔK method (Evanno et al., 2005). 217 J. regia accessions, K = 2
(TIFF)

S2 Fig. Graphical method allowing the detection of the number of populations K using ΔK (Evanno et al., 2005). 63 'East. Eur. and Asia' accessions, K = 3.
(PDF)

S3 Fig. Graphical method allowing the detection of the number of populations K using ΔK (Evanno et al., 2005). 127 'West. Eur. and Am.' accessions, K = 3.
(PDF)

S1 Table. List of accessions studied.
(XLSX)

S2 Table. SSR allele-specificity among *Juglans* species.
(XLSX)

S3 Table. Table summarizing the results using the Evanno method (output of Structure Harvester), based on the analyses of the 217 *J. regia* accessions.
(XLSX)

S4 Table. List of *J. regia* accessions with their memberships to clusters (K = 2), subclusters (K = 3) and core collection.
(XLSX)

S5 Table. Table summarizing the results using the Evanno method (output of Structure Harvester), based on the analyses of the 63 Eastern Europe and Asia accessions.
(XLSX)

S6 Table. Table summarizing the results using the Evanno method (output of Structure Harvester), based on the analyses of the 127 Western Europe and America accessions.
(XLSX)

S7 Table. Genetic diversity estimations of the clusters found with STRUCTURE K = 2.
(XLSX)

S8 Table. Pairwise FST among populations identified with STRUCTURE K = 2.
(XLSX)

S9 Table. Private alleles among clusters for *J. regia* (K = 2).
(XLSX)

S10 Table. Raw data set necessary to replicate the study.
(XLSX)

Acknowledgments

First, we would like to thank the late Eric Germain, former head of the breeding program at INRA of Bordeaux from 1977 to 2007. His remarkable work, then continued by Francis Delort, has given us the opportunity to study a rich set of plant material. We thank the Fruit Tree Experimental Unit of INRA in Toulenne and the *Prunus/Juglans* Genetic Resources Center for the maintenance of the collection and for helping us to collect the samples. We acknowledge the BioGEVES laboratory for DNA extraction and SSR genotyping. Also, Ctifl, holder of the project "INNOV'noyer", in partnership with INRA, want to thank the 'Région Nouvelle-Aquitaine' for funding a part of this project. Ctifl also thanks the 'Cifre' convention by 'ANRT' (Agence Nationale de la Recherche et de la Technologie) for funding half the cost of the doctoral fellowship of Anthony Bernard. Then, it is important for us to note that the project is supported by the 'AgriSudOuest Innovation' competitiveness cluster. Finally, our special thanks are addressed to Charles A. Leslie (UC Davis, Department of Plant Sciences, California) for English language revision of the manuscript.

Author Contributions

Conceptualization: Fabrice Lheureux.

Data curation: Teresa Barreneche, Elisabeth Dirlewanger.

Funding acquisition: Fabrice Lheureux.

Methodology: Teresa Barreneche.

Project administration: Fabrice Lheureux.

Supervision: Elisabeth Dirlewanger.

Writing – original draft: Anthony Bernard.

Writing – review & editing: Anthony Bernard, Teresa Barreneche, Fabrice Lheureux, Elisabeth Dirlewanger.

References

1. Germain E, Prunet JP, Garcin A. Le noyer, monographie. CTIFL. 1999.
2. Zeven AC, Zhukovsky PM. Dictionary of cultivated plants and their centres of diversity excluding ornamentals, forest trees, and lower plants. Wageningen: Centre for Agricultural Publishing and Documentation; 1975
3. Aradhya MK, Potter D, Simon CJ. Cladistic Biogeography of Juglans (Juglandaceae) Based on Chloroplast DNA Intergenic Spacer Sequences. In: Motley TJ, Zerega N, Cross H, editors. Darwins Harvest New Approaches Orig Evol Conserv Crops. Columbia University Press; 2006
4. Woodworth RH. Meiosis of micro-sporogenesis in the Juglandaceae. Am J Bot. 1930; 17:863–9.
5. Manning WE. The classification within the Juglandaceae. Ann Mo Bot Gard. 1978; 65:1058–87.
6. Fjellstrom RG, Parfitt DE. Walnut (Juglans spp.) genetic diversity determined by restriction fragment length polymorphism. Genome. 1994; 37:690–700. PMID: 18470111
7. Fornari B, Malvolti ME, Taurchini D, Fineschi S, Beritognolo I, Maccaglia E, et al. Isozyme and organellar DNA analysis of genetic diversity in natural/naturalised European and Asiatic walnut (Juglans regia L.) populations. Acta Hortic. 2001; 544:167–78.
8. Malvolti ME, Fineschi S, Pigliucci M. Morphological Integration and Genetic Variability in Juglans regia L. J Hered. 1994; 85:389–94.
9. Nicese FP, Hormaza JI, McGranahan GH. Molecular characterization and genetic relatedness among walnut (Juglans regia L.) genotypes based on RAPD markers. Euphytica. 1998; 101:199–206.
10. Fatahi R, Ebrahimi A, Zamani Z. Characterization of Some Iranians and Foreign Walnut Genotypes Using Morphological Traits and RAPDs Markers. Hortic Environ Biotechnol. 2010; 51:51–60.
11. Potter D, Gao F, Aiello G, Leslie C, McGranahan G. Intersimple Sequence Repeat Markers for Fingerprinting and Determining Genetic Relationships of Walnut (Juglans regia) Cultivars. J Am Soc Hortic Sci. 2002; 127:75–81.
12. Christopoulos MV, Rouskas D, Tsantili E, Bebeli PJ. Germplasm diversity and genetic relationships among walnut (Juglans regia L.) cultivars and Greek local selections revealed by Inter-Simple Sequence Repeat (ISSR) markers. Sci Hortic. 2010; 125:584–92.
13. Bayazit S, Kazan K, Gülbitti S, Çevik V, Ayanoğlu H, Ergül A. AFLP analysis of genetic diversity in low chill requiring walnut (Juglans regia L.) genotypes from Hatay, Turkey. Sci Hortic. 2007; 111:394–8.
14. Ali AM, Zubair SJ, Abbas AM, Jubrael JM. Genetic diversity among Walnuts (Juglans regia) population in Kurdistan Region–Iraq using AFLP-PCR. ZANCO J Pure Appl Sci. 2016; 28:50–55.
15. Bernard A, Lheureux F, Dirlewanger E. Walnut: past and future of genetic improvement. Tree Genet Genomes. 2018; 14:1.
16. Hintum TJL, Brown AHD, Spillane C, Hodgkin T. Core collections of plant genetic resources. IPGRI Technical Bulletin. 2000; 3:48.
17. Dangl GS, Woeste K, Aradhya MK, Koehmstedt A, Simon C, Potter D, et al. Characterization of 14 Microsatellite Markers for Genetic Analysis and Cultivar Identification of Walnut. J Am Soc Hortic Sci. 2005; 130:348–54.
18. Woeste K, Burns R, Rhodes O, Michler C. Thirty Polymorphic Nuclear Microsatellite Loci from Black Walnut. J Hered. 2002; 93:58–60. PMID: 12011179
19. Dang M, Zhang T, Hu Y, Zhou H, Woeste K, Zhao P. De Novo Assembly and Characterization of Bud, Leaf and Flowers Transcriptome from Juglans Regia L. for the Identification and Characterization of New EST-SSRs. Forests. 2016; 7:247–63.
20. Jombart T. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics. 2008; 24:1403–5. https://doi.org/10.1093/bioinformatics/btn129 PMID: 18397895
21. Peakall R, Smouse PE. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics. 2012; 28:2537–9. https://doi.org/10.1093/bioinformatics/bts460 PMID: 22820204
22. Peakall R, Smouse PE. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes. 2006; 6:288–95.

23. Kalinowski ST, Taper ML, Marshall TC. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Molecular Ecology. 2007; 16:1099–106. https://doi.org/10.1111/j.1365-294X.2007.03089.x PMID: 17305863

24. Perrier X, Jacquemoud-Collet J. DARwin software. 2006. Available from: http://darwin.cirad.fr/

25. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000; 155:945–59. PMID: 10835412

26. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software structure: a simulation study. Mol Ecol. 2005; 14:2611–20. https://doi.org/10.1111/j.1365-294X.2005.02553.x PMID: 15969739

27. Earl DA, VonHoldt BM. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour. 2012; 4:359–61.

28. Keenan K, McGinnity P, Cross TF, Crozier WW, Prodöhl PA. diveRsity: An R package for the estimation of population genetics parameters and their associated errors. Methods in Ecology and Evolution. 2013; 4:782–8.

29. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol. 1987; 4:406–25. https://doi.org/10.1093/oxfordjournals.molbev.a040454 PMID: 3447015

30. Scott KD, Eggler P, Seaton G, Rossetto M, Ablett EM, Lee LS, et al. Analysis of SSRs derived from grape ESTs. Theor Appl Genet. 2000; 100:723–6.

31. Hu J, Wang L, Li J. Comparison of genomic SSR and EST-SSR markers for estimating genetic diversity in cucumber. Biol Plant. 2011; 55:577–80.

32. Ince AG, Karaca M, Uygur Göçer E, Aydın A. Descriptive Statistics and PIC Values of Genomic-and Transcriptomic-Microsatellites in Several Plant Species. J Sci Eng Res. 2017; 2017:236–46.

33. Dong W, Xu C, Li W, Xie X, Lu Y, Liu Y, et al. Phylogenetic Resolution in Juglans Based on Complete Chloroplast Genomes and Nuclear DNA Sequences. Front Plant Sci. 2017; 8:1148. https://doi.org/10.3389/fpls.2017.01148 PMID: 28713409

34. Ebrahimi A, Fatahi R, Zamani Z. Analysis of genetic diversity among some Persian walnut genotypes (Juglans regia L.) using morphological traits and SSRs markers. Sci Hortic. 2011; 130:146–51.

35. Noor Shah U, Mir JI, Ahmed N, Fazili KM. Assessment of germplasm diversity and genetic relationships among walnut (Juglans regia L.) genotypes through microsatellite markers. J Saudi Soc Agric Sci. 2016. Available from: http://linkinghub.elsevier.com/retrieve/pii/S1658077X16300662

36. Han H, Woeste KE, Hu Y, Dang M, Zhang T, Gao XX, et al. Genetic diversity and population structure of common walnut (Juglans regia) in China based on EST-SSRs and the nuclear gene phenylalanine ammonia-lyase (PAL). Tree Genet Genomes. 2016; 12:111.

37. Ebrahimi A, Zarei A, McKenna JR, Bujdoso G, Woeste KE. Genetic diversity of Persian walnut (Juglans regia) in the cold-temperate zone of the United States and Europe. Sci Hortic. 2017; 220:36–41.

38. Vischi M, Chiabà C, Raranciuc S, Poggetti L, Messina R, Ermacora P, et al. Genetic Diversity of Walnut (Juglans Regia L.) in the Eastern Italian Alps. Forests. 2017; 8:81.

39. Mahmoodi R, Rahmani F, Rezaee R. Genetic diversity among Juglans regia L. genotypes assessed by morphological traits and microsatellite markers. Span J Agric Res. 2013; 11:431–7.

40. Wang H, Pei D, Gu R, Wang B. Genetic Diversity and Structure of Walnut Populations in Central and Southwestern China Revealed by Microsatellite Markers. J Am Soc Hortic Sci. 2008; 133:197–203.

41. Aradhya M, Woeste K, Velasco D. Genetic Diversity, Structure and Differentiation in Cultivated Walnut (Juglans regia L.). Acta Hortic. 2009; 861:127–132.

42. Ruiz-Garcia L, Lopez-Ortega G, Denia AF, Tomas DF. Identification of a walnut (Juglans regia L.) germplasm collection and evaluation of their genetic variability by microsatellite markers. Span J Agric Res. 2011; 9:179–192.

43. Karimi R, Ershadi A, Vahdati K, Woeste K. Molecular Characterization of Persian Walnut Populations in Iran with Microsatellite Markers. HortScience. 2010; 45:1403–1406.

44. Mohsenipoor S, Vahdati K, Amiri R, Mozaffari MR. Study of the Genetic Structure and Gene Flow in Persian Walnut (Juglans regia L.) Using SSR Markers. Acta Hortic. 2010; 861:133–42.

45. Odong TL, van Heerwaarden J, Jansen J, van Hintum TJL, van Eeuwijk FA. Determination of genetic structure of germplasm collections: are traditional hierarchical clustering methods appropriate for molecular marker data? Theor Appl Genet. 2011; 123:195–205. https://doi.org/10.1007/s00122-011-1576-x PMID: 21472410

46. Torres AR, Grunvald AK, Martins TB, dos Santos MA, Lemos NG, Silva LAS, et al. Genetic structure and diversity of a soybean germplasm considering biological nitrogen fixation and protein content. Sci Agric. 2015; 72:47–52.

47. Campoy JA, Lerigoleur-Balsemin E, Christmann H, Beauvieux R, Girollet N, Quero-García J, et al. Genetic diversity, linkage disequilibrium, population structure and construction of a core collection of Prunus avium L. landraces and bred cultivars. BMC Plant Biol. 2016; 16:1–15. https://doi.org/10.1186/s12870-015-0700-5

48. Vigouroux Y, Glaubitz JC, Matsuoka Y, Goodman MM, Sánchez G J, Doebley J. Population structure and genetic diversity of New World maize races assessed by DNA microsatellites. Am J Bot. 2008; 95 (10):1240–53. https://doi.org/10.3732/ajb.0800097 PMID: 21632329

49. Liang W, Dondini L, De Franceschi P, Paris R, Sansavini S, Tartarini S. Genetic Diversity, Population Structure and Construction of a Core Collection of Apple Cultivars from Italian Germplasm. Plant Mol Biol Report. 2015; 33(3):458–73.

50. Cubry P, De Bellis F, Pot D, Musoli P, Leroy T. Global analysis of Coffea canephora Pierre ex Froehner (Rubiaceae) from the Guineo-Congolese region reveals impacts from climatic refuges and migration effects. Genet Resour Crop Evol. 2013; 60(2):483–501.

51. Pollegioni P,Woeste K, Chiocchini F, Del Lungo S, Ciolfi M, Olimpieri I, Tortolano V, Clark J, Hemery GE, Mapelli S, Malvolti ME. Rethinking the history of common walnut (Juglans regia L.) in Europe: its origins and human interactions. PLoS One. 2017; 12(3): e0172541 https://doi.org/10.1371/journal.pone.0172541 PMID: 28257470

52. Fernández-Cruz J, Fernández-López J. Genetic structure of wild sweet chestnut (Castanea sativa Mill.) populations in northwest of Spain and their differences with other European stands. Conserv Genet. 2016; 17(4):949–67.

53. Pereira-Lorenzo S, Ramos-Cabrer AM, Barreneche T, Mattioni C, Villani F, Díaz-Hernández MB, et al. Database of European chestnut cultivars and definition of a core collection using simple sequence repeats. Tree Genet Genomes. 2017; 13(5).

54. Pereira-Lorenzo S, Costa RML, Ramos-Cabrer AM, Ciordia-Ara M, Ribeiro CAM, Borges O, et al. Chestnut cultivar diversification process in the Iberian Peninsula, Canary Islands, and Azores. Genome. 2011; 54(4):301–15. https://doi.org/10.1139/g10-122 PMID: 21491973

55. Martínez-García PJ, Crepeau MW, Puiu D, Gonzalez-Ibeas D, Whalen J, Stevens KA, et al. The walnut (Juglans regia) genome sequence reveals diversity in genes coding for the biosynthesis of nonstructural polyphenols. Plant J. 2016; 87:507–32. https://doi.org/10.1111/tpj.13207 PMID: 27145194

56. Butterfield T, Phu M, Chakraborty S, et al. Identifying the genetic determinant(s) of pellicle coloration in walnut. Walnut Res Rep. 2015.

57. McGranahan G, Leslie C. 'Robert Livermore', a Persian walnut cultivar with a red seedcoat. HortScience. 2004; 39:1772.

58. Odong TL, Jansen J, van Eeuwijk FA, van Hintum TJL. Quality of core collections for effective utilisation of genetic resources review, discussion and interpretation. Theor Appl Genet. 2013; 126:289–305. https://doi.org/10.1007/s00122-012-1971-y PMID: 22983567

59. Brown AHD. Core collections: a practical approach to genetic resources management. Genome. 1989; 31:818–824.

60. Tulecke W, McGranahan G. The Walnut Germplasm Collection of the University of California, Davis. A Description of the Collection and a History of the Breeding Program of Eugene F. Serr and Harold I. Forde. Report No. 13. University Of California, Genetic Resources Conservation Program, Davis, CA; 1994.

Choix des accessions pour l'étude d'association pangénomique et la cartographie génétique

Si l'analyse globale de la diversité génétique de la collection via l'utilisation de marqueurs microsatellites a été valorisée par la publication d'un article, ces données de génotypage ont servi également à réaliser le choix d'un groupe d'accessions pour mener une étude d'association pangénomique (GWAS). De plus, ces données ont permis de mettre en évidence des aliens au sein de la descendance F_1 de cartographie.

Une puce de génotypage à haute densité de 700K SNP spécifique de *J. regia* a été mise au point grâce à l'équipe de l'Université de Davis en Californie (Marrano et al. 2019a) et la société Affymetrix®. Trois puces ont été utilisées contenant chacune 96 puits. Avec un puits par plaque dédié au témoin, il reste donc 285 puits disponibles pour le génotypage. Ainsi, les 85 individus de la descendance F_1 de cartographie ont été analysés et un choix de 200 accessions de la collection a donc dû être réalisé parmi les 253 étudiées (Figure 22).

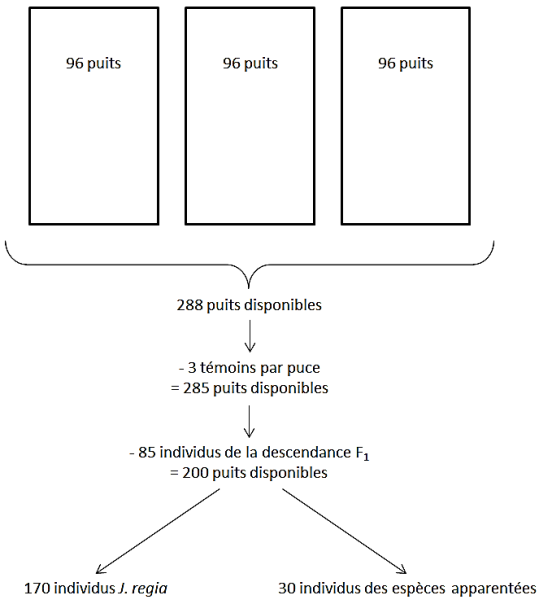


Figure 22 : Utilisation des puits disponibles sur l'ensemble des 3 puces de génotypage.
200 accessions de la collection de ressources génétiques ont pu être génotypées.

Démarche concernant le choix des 200 accessions de la collection pour la GWAS

Suppression des accessions identiques

Au sein de la collection, certaines accessions ont été introduites en plusieurs exemplaires, notamment pour cinq cultivars français : 'Parisienne' (3 exemplaires), 'Franquette' (4 exemplaires), 'Mayette' (2 exemplaires), 'Corne' (3 exemplaires) et 'Marbot' (3 exemplaires). L'analyse du polymorphisme des profils microsatellites montre qu'il s'agit

bien de clones et l'exemplaire combinant le meilleur état sanitaire en verger et la plus grande disponibilité de données phénotypiques issues des archives a été retenu pour la GWAS. C'est le cas également de certaines accessions d'espèces apparentées présentes en deux exemplaires comme 'CT 3-1' (*J. cathayensis*), 'AN 3' (*J. ailantifolia*), 'PS 2-3' (*J. pitteursii*) et 'MO 1' (*J. mollis*).

Suppression des hybrides INRAE dont les parents sont présents

La collection comprend des hybrides créés par E. Germain et son équipe pour le besoin de son programme de création variétale. Les hybrides présentant un avantage pour la sélection et inscrits en niveau 1 ou niveau 2 lors du dernier programme ont été conservés. Pour les autres, si les parents du croisement dont est issu l'hybride sont présents dans la collection, ils n'ont pas été retenus. Par exemple, les accessions H 95-87 (issue d'un croisement de 'Meylannaise' par 'Payne') et H 91-88 (issue d'un croisement de 'Franquette' par 'Payne') n'ont pas été retenues puisqu'elles n'ont pas été inscrites en niveau 1 ou niveau 2 et puisque les parents sont présents dans la collection.

Suppression des accessions ayant peu d'informations

Certaines accessions introduites dans la collection possèdent peu d'informations dans le registre d'introduction et n'ont donc pas été retenues. Il s'agit généralement de semis et d'autofécondations de certains cultivars qui ont dû être testés dans le cadre du programme d'amélioration variétale mais n'ont pas dû présenter de grand intérêt. C'est le cas par exemple de l'accession RA 0248 AF5, une autofécondation de 'Lozeronne' et de l'accession RA 0248-8, un semis de 'Lozeronne'.

Suppression d'accessions d'espèces apparentées pour un équilibre numérique

La collection contient un nombre d'accessions très hétérogène parmi les espèces apparentées. Par exemple, *J. nigra*, beaucoup utilisé en porte-greffe, compte de nombreuses accessions tandis que *J. microcarpa*, rare dans son milieu naturel en Amérique du Nord, ne compte qu'une accession. Afin d'équilibrer le nombre d'accessions par espèce, certaines n'ont pas été retenues sur la base des résultats de polymorphisme des profils microsatellites et de l'état sanitaire en verger. Ainsi, cinq accessions de *J. nigra* ont été retirées comme NG 0261-3 et NG 0073-2, une accession de *J. cinerea* CN 13-3 et une accession de *J. cordiformis* ST 1.

Suppression d'accessions selon leur provenance pour un équilibre numérique

La collection comprend des accessions de nombreux pays mais certains pays sont plus représentés que d'autres. Afin d'équilibrer le nombre d'accessions par pays de provenance,

certaines n'ont pas été retenues toujours sur la base des résultats de diversité des profils microsatellites et de l'état sanitaire en verger. Ainsi, l'accession grecque 'S 14 B Sparte' et les accessions catalanes 'MBT 218' et 'MBT 232' ont été retirées. Aussi, parmi les 33 accessions provenant du jardin botanique de Kiev en Ukraine et originaires du centre d'origine du noyer cultivé, 16 ont été retirées.

Après application de ces cinq filtres, 200 accessions comprenant 170 accessions de *J. regia* et 30 accessions appartenant aux espèces apparentées ont été retenues pour le génotypage avec la puce à haute densité de 700K SNP.

Filtre qualité des données de génotypage Affymetrix®

Le programme « Axiom Analysis Suite 3.1 » permet de réaliser le contrôle qualité des données de génotypage. La société éditrice de la puce préconise de ne retenir que les individus qui possèdent un « dish QC » supérieur ou égal à 0,82 et ceux dont le « QC call rate » est supérieur ou égal à 97%. Malheureusement, ces 30 accessions appartenant aux espèces apparentées ont été exclues automatiquement par ces filtres, avant le génotypage. Il semble donc que les SNP développés à l'heure actuelle sur *J. regia* ne soient pas transférables au sein des autres espèces du genre *Juglans*.

Démarche concernant le tri des 85 individus de la descendance F_1 de cartographie

De la même façon que pour les individus de la collection, un individu n'a pas satisfait le filtre qualité Affymetrix®. De plus, 4 individus ont été retirés sur la base des données microsatellites. Ces 4 individus possédaient plusieurs allèles différents de ceux des parents 'Franquette' et 'UK 6-2' et ne semblaient donc pas issus du croisement souhaité au départ. Enfin, 2 individus supplémentaires ont été écartés sur la base de la qualité des données de génotypage SNP. Au final, il reste donc 78 individus dans cette descendance F_1 utilisés pour la cartographie.

3. Etude du déterminisme génétique du type de fructification et des caractères liés à la phénologie

Utilisation conjointe de la GWAS et de la détection de QTL avec deux années de données phénotypiques et les données anciennes

Un meilleur rendement, des noix de beau calibre, un cerneau clair et une facilité d'énoisage sont des caractères primordiaux dans les programmes d'amélioration de la noix à travers le monde. Mais l'aptitude d'adaptation aux conditions climatiques changeantes est une priorité ; en particulier en France où les gelées printanières tardives sont fréquentes. Le travail précédent a permis d'apprécier la diversité génétique des accessions *J. regia* et d'effectuer le choix des accessions formant le panel de GWAS. L'information de structure du panel a également été obtenue, information essentielle pour la construction des modèles GWAS. En parallèle, le phénotypage complet et précis de la phénologie a été réalisé, ainsi que celui du type de fructification, caractère intervenant directement dans le niveau de production. Tous les éléments sont donc réunis pour répondre à la question suivante :

Quelle est l'architecture génétique des caractères liés à la phénologie et au type de fructification chez le noyer cultivé ?

Une GWAS grâce à la puce à haute densité de 700K SNP a donc été conduite sur le panel de 170 accessions de *J. regia* et conjointement, l'utilisation de la descendence F_1 a permis de réaliser une cartographie génétique avec une détection de QTL pour ces caractères. Via la GWAS, de fortes associations ont été découvertes entre des SNP localisés au début du chromosome 1 et les dates de débourrement et de floraison femelle. Les données de phénotypage de 2018 et 2019 ont été utilisées, ainsi que les données chronologiques issues des archives d'Eric Germain de l'INRAE. Ces résultats ont été confirmés par des QTL détectés dans la même région génomique. D'autres associations ont été observées avec la dichogamie et le type de fructification sur le chromosome 11.

Le premier marqueur de type KASP chez le noyer disponible pour un caractère lié à la phénologie, la date de débourrement, a été développé en utilisant du matériel végétal et des données de phénotypage de l'Université de Davis en Californie. Ce marqueur est donc un outil concret pour le·la sélectionneur·euse. Enfin, des gènes candidats ont été proposés, notamment un gène codant pour une enzyme du catabolisme des sucres pour la dichogamie et un gène lié aux mécanismes de division cellulaire pour la date de floraison femelle.

L’article suivant a été publié en ligne le 04 mars 2020 dans le journal BMC Genomics, volume 21:203, sous le nom de « Association and linkage mapping to unravel genetic architecture of phenological traits and lateral bearing in Persian walnut (*Juglans regia* L.) » (DOI 10.1186/s12864-020-6616-y).

Bernard *et al. BMC Genomics* (2020) 21:203
https://doi.org/10.1186/s12864-020-6616-y

BMC Genomics

RESEARCH ARTICLE Open Access

Association and linkage mapping to unravel genetic architecture of phenological traits and lateral bearing in Persian walnut (*Juglans regia* L.)

Anthony Bernard[1,2], Annarita Marrano[3], Armel Donkpegan[1], Patrick J. Brown[3], Charles A. Leslie[3], David B. Neale[3], Fabrice Lheureux[2] and Elisabeth Dirlewanger[1*]

Abstract

Background: Unravelling the genetic architecture of agronomic traits in walnut such as budbreak date and bearing habit, is crucial for climate change adaptation and yield improvement. A Genome-Wide Association Study (GWAS) using multi-locus models was conducted in a panel of 170 walnut accessions genotyped using the Axiom™ *J. regia* 700 K SNP array, with phenological data from 2018, 2019 and legacy data. These accessions come from the INRAE walnut germplasm collection which is the result of important prospecting work performed in many countries around the world. In parallel, an F_1 progeny of 78 individuals segregating for phenology-related traits, was genotyped with the same array and phenotyped for the same traits, to construct linkage maps and perform Quantitative Trait Loci (QTLs) detection.

Results: Using GWAS, we found strong associations of SNPs located at the beginning of chromosome 1 with both budbreak and female flowering dates. These findings were supported by QTLs detected in the same genomic region. Highly significant associated SNPs were also detected using GWAS for heterodichogamy and lateral bearing habit, both on chromosome 11. We developed a Kompetitive Allele Specific PCR (KASP) marker for budbreak date in walnut, and validated it using plant material from the Walnut Improvement Program of the University of California, Davis, demonstrating its effectiveness for marker-assisted selection in Persian walnut. We found several candidate genes involved in flowering events in walnut, including a gene related to heterodichogamy encoding a sugar catabolism enzyme and a cell division related gene linked to female flowering date.

Conclusions: This study enhances knowledge of the genetic architecture of important agronomic traits related to male and female flowering processes and lateral bearing in walnut. The new marker available for budbreak date, one of the most important traits for good fruiting, will facilitate the selection and development of new walnut cultivars suitable for specific climates.

Keywords: Walnut, *Juglans regia* L., Association genetics, GWAS, Germplasm collection, Linkage map, QTL analysis, Phenology, Bearing habit

* Correspondence: elisabeth.dirlewanger@inrae.fr
[1]INRAE, Univ. Bordeaux, UMR BFP, F-33882 Villenave d'Ornon, France
Full list of author information is available at the end of the article

© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Persian walnut (*Juglans regia* L.) is one of the oldest food sources known [1]. It is a monoecious and dichogamous tree species with 2n = 2x = 32 chromosomes [2], and grows in temperate regions [3]. Worldwide in-shell walnut production, mainly from China, California and Iran, exceeded 3800 kt in 2017, as reported by the Food and Agriculture Organization of the United Nations (www.fao.org). At more than 22,000 ha, Persian walnut is the second leading tree crop in France, after apple. In the last 3 years, France has oscillated between 7th and 9th position for in-shell walnut production (circa 40 kt) [4]. Increased yield, larger nut size, light kernel color, and ease of cracking are among the main goals of walnut breeding worldwide [5]. The ability to adapt to specific climatic conditions is also a breeding priority, especially in France where late spring frosts are prevalent [4]. In that respect, a better understanding of phenology and bearing habit, both key determinants of yield, is of upmost importance for walnut genetic improvement and cultivation [6].

Climate change, particularly global warming, is no longer to be proven within the scientific community [7], and researchers are studying its impact on phenology of temperate trees. In these species, growth is punctuated by an annually repeated phase of rest, called bud dormancy [8]. This dormancy period is influenced by various environmental factors, such as photoperiod and temperature, resulting in fulfilment of chilling and heat requirements [9]. In walnut, chilling and heat requirements were widely estimated in Iran [10] and showed, for instance, a range of chilling requirements from 650 h at + 4 °C for 'Serr' to 1000 h for 'Hartley' cultivars [11]. In France, the frost resistance of walnut were studied [12, 13] and many phenology studies of temperate tree species in Europe report a time shift of phenological events [14–17]. An advancing effect of warm springs on phenological events has been observed for walnut in California, particularly for leafing date [18]. Similar findings have been reported in Slovenia [19] and Romania [20]. Using phenological data recorded by the Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement (INRAE) of Bordeaux from 1989 to 2016, we also observed an average advance in budbreak in France of 5 days over the last 3 decades [21]. In Iran, researchers assessed land suitability for walnut cultivation under present and future climatic conditions, and predict that the currently suitable area will be significantly reduced [22].

Genetic control of phenology-related traits is fundamental for the development of new, resilient cultivars, able to adapt to changing climatic conditions. Many studies have focused on genetic dissection of phenological traits (e.g., chilling requirements and flowering time) in diverse fruit crops, such as peach, apricot and sweet cherry [23, 24]. In walnut, a significant genotype effect has been identified for heat requirements [25]. Moreover, high heritability has been shown for leafing date (71–96%), type of heterodichogamy (90%), and female/male blooming (80%) [26, 27]. Persian walnut has two main types of bearing habit. Fruiting can occur only at the terminal position of new branches or at both terminal and lateral positions [28]. A genetic locus for lateral bearing has been identified based in an F_2 progeny in the United-States [29], but has not been sufficiently robust for wider use in marker-assisted selection.

Release of the first walnut genome sequence [30] facilitated advanced genetic and genomic studies, including development of the first high-density Axiom™ *J. regia* 700 K SNP genotyping array [31]. Application of this powerful genotyping tool allowed genetic dissection of crucial traits in walnut, such as nut-related traits [32] and water use efficiency [33, 34]. A recent study, combining genome-wide association study (GWAS) and classical linkage mapping, found major loci for leafing and harvest dates on chromosome 1 (Chr1), and lateral fruitfulness on Chr11 [35].

Here, we studied for the first time in walnut, the genetic control of budbreak date and female/male flowering dates, using the Axiom™ *J. regia* 700 K SNP array to genotype both a panel of 170 walnut accessions of diverse geographical provenience and an F_1 progeny segregating for these traits. This study sought to identify candidate genes for both female and male flowering dates and to develop the first Kompetitive Allele Specific PCR (KASP) marker for phenology in walnut. This will be useful for walnut breeding programs in selecting of new resilient varieties to climate change.

Results

Phenotypic variations of phenology-related traits and lateral bearing

Two populations were used in this study: a GWAS panel of 170 diverse accessions of worldwide origin and an F_1 mapping progeny of 78 individuals resulting from a biparental controlled cross between 'Franquette' (late flowering), and 'UK 6–2' (intermediate to early flowering). Both populations were maintained at the INRAE of Bordeaux field station and phenotyped during 2018 and 2019. For the GWAS panel, we also used previously collected (legacy) phenotypical data taken between 1989 and 2011.

For the GWAS panel, the 2018–2019 data exhibited high variation in phenology-related traits, particularly for budbreak which ranged in 2019 from 57 Julian days for 'Early Ehrhardt' to 128 for 'Fertignac' (Feb 27th to May 9th) (Figures S1 and S2). The F_1 progeny in 2019 exhibited a smaller range of 76 to 102 Julian days (Figure S3

and S4). Generally, budbreak was earlier in 2019 (87.78 Julian days ±12.65 for the GWAS panel, 90.71 ± 5.48 for the F_1 progeny) than in 2018 (92.47 ± 11.06 for the GWAS panel, 95.55 ± 4.97 for the F_1 progeny).

We found significant positive correlations between budbreak date and female flowering stages for both the GWAS panel (0.83 to 0.84; Fig. 1a), and the F_1 progeny (0.45 to 0.52; Fig. 1b). Similar significant positive correlations were found between budbreak date and male flowering stages for the GWAS panel (0.78 and 0.81; Fig. 1a), and the F_1 progeny (0.61 to 0.84; Fig. 1b). Comparison of the 2 years shows that early accessions in 2018 were also early in 2019, suggesting genetic control of phenology-related traits in walnut. Female flowering was earlier in 2018 than 2019, but the accession order was consistent for both years. In addition, both female and male flowering durations showed low correlations and low statistical significances with other traits. We did not phenotype the F_1 progeny for bearing habit, since this trait did not segregate in that population, but we observed great variability for fruit bearing within the GWAS panel.

High broad-sense heritability values were observed for budbreak date, with H^2 of 0.95 when using legacy data and 0.93 using only two-year data (Table 1). Overall, H^2

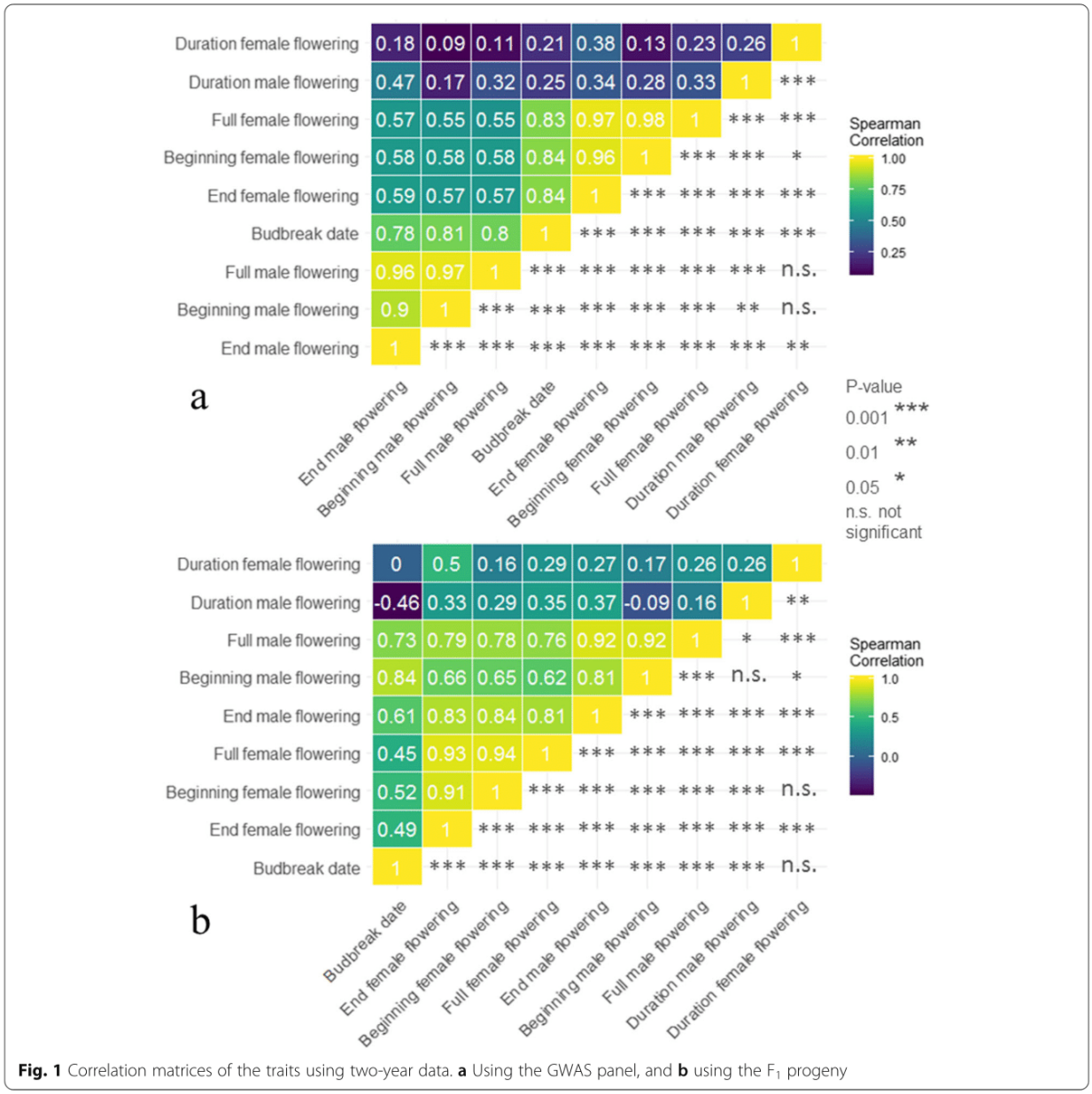


Fig. 1 Correlation matrices of the traits using two-year data. **a** Using the GWAS panel, and **b** using the F_1 progeny

Table 1 Descriptive statistics and broad-sense heritabilities

Trait	Plant material	Year	Mean[a] ± SD[b]	Range[a]	H^2
Bearing habit	GWAS panel	1989–2016	4.01 ± 2.71	1–9	–
		2018	–	–	–
		2019	4.61 ± 2.18	1–9	–
Budbreak date	GWAS panel	1989–2016	99.02 ± 12.87	60–133	0.95
		2018	92.47 ± 11.06	72–115	0.93
		2019	87.78 ± 12.65	57–128	
	F_1 progeny	2018	95.55 ± 4.97	90–105	0.67
		2019	90.71 ± 5.48	76–102	
Beginning female flowering date	GWAS panel	1989–2016	119.11 ± 11.71	69–151	0.91
		2018	111.27 ± 10.19	90–142	0.95
		2019	110.70 ± 13.28	78–141	
	F_1 progeny	2018	112.54 ± 5.12	106–124	0.75
		2019	116.38 ± 5.03	102–128	
Peak female flowering date	GWAS panel	1989–2016	125.11 ± 11.47	78–154	0.93
		2018	115.22 ± 11.42	95–147	0.96
		2019	115.42 ± 13.00	87–144	
	F_1 progeny	2018	116.69 ± 5.63	110–128	0.67
		2019	121.81 ± 5.28	110–132	
End female flowering date	GWAS panel	1989–2016	135.14 ± 12.09	88–167	0.90
		2018	122.32 ± 12.35	103–153	0.96
		2019	122.38 ± 12.86	97–149	
	F_1 progeny	2018	123.35 ± 6.27	112–135	0.64
		2019	128.42 ± 5.44	116–137	
Female bloom duration	GWAS panel	1989–2016	16.47 ± 6.66	1–53	0.37
		2018	11.05 ± 4.17	3–23	0.26
		2019	11.68 ± 2.68	5–19	
	F_1 progeny	2018	10.81 ± 3.11	4–16	0.00
		2019	12.04 ± 2.30	6–17	
Heterodichogamy	GWAS panel	1989–2016	2.80 ± 2.09	1–9	0.95
		2018	3.90 ± 2.15	1–9	0.84
		2019	3.17 ± 2.48	1–9	
Beginning male flowering date	GWAS panel	1989–2016	112.34 ± 10.69	77–149	0.82
		2018	108.17 ± 6.81	99–137	0.86
		2019	105.06 ± 10.68	85–140	
	F_1 progeny	2018	106.17 ± 3.25	102–114	0.75
		2019	104.17 ± 5.25	88–116	
Peak male flowering date	GWAS panel	1989–2016	116.99 ± 10.64	83–154	0.88
		2018	111.13 ± 8.08	103–142	0.92
		2019	109.09 ± 10.94	91–144	
	F_1 progeny	2018	108.38 ± 3.89	104–117	0.86
		2019	108.56 ± 5.69	95–128	
End male flowering date	GWAS panel	1989–2016	122.45 ± 10.58	85–163	0.87
		2018	114.40 ± 9.62	104–145	0.95
		2019	114.33 ± 11.13	97–149	
	F_1 progeny	2018	111.97 ± 4.86	105–123	0.81
		2019	114.74 ± 6.17	102–130	
Male bloom duration	GWAS panel	1989–2016	10.53 ± 4.70	2–35	0.32
		2018	6.23 ± 3.97	2–24	0.22
		2019	9.27 ± 2.45	4–16	
	F_1 progeny	2018	5.81 ± 2.20	2–13	0.00
		2019	10.58 ± 3.16	6–21	

[a] Date and duration traits are in Julian days, bearing habit and heterodichogamy are categorical traits from 1 to 9
[b] SD is the abbreviation for standard deviation

values were lower within the F_1 progeny ($H^2 = 0.67$ for budbreak date). However, we found low values for male flowering duration ($H^2 = 0.22$) and female flowering duration ($H^2 = 0.26$) within GWAS panel (using the recent two phenotyping years), while no genetic effect was found for the F_1 progeny. Therefore, we did not consider both male and female flowering durations in the GWAS and QTL mapping analyses.

Population structure of the GWAS panel

A total of 364,275 SNPs were retained after filtering for high resolution SNPs categories (Poly High Resolution and No Minor Homozygotes), for genotyping rate > 90%, and minor allele frequency > 5% (Table 2). We investigated the population structure of our association panel using the Bayesian clustering approach implemented in fastSTRUCTURE, and Principal Component Analysis

Table 2 SNPs used for the GWAS analyses and the construction of the parental linkage maps 'Franquette' and 'UK6–2'

	Number of markers		Percentage of markers	
Total of SNPs	*609,658*		*100*	
To keep SNPs of high resolution from Axiom® Analysis Suite				
High resolution SNPs				
PolyHighResolution	397,921		65,27	
NoMinorHom	75,564		12,39	
MonoHighResolution	36,684		6,02	
Low resolution SNPs				
CallRateBelowThreshold	27,761		4,55	
OffTargetVariant	4787		0,79	
Other	66,941		10,98	
Total of retained SNPs	*510,169*		*83.68*	
To keep SNPs with mendelian inheritance using F_1 progeny				
SNPs having no mendelian inheritance	661			
Total of retained SNPs	*509,508*		*83.57*	
To keep SNPs having genotyping rate > 90%				
	GWAS		**Linkage maps**	
	Number of markers	**Percentage of markers**	**Number of markers**	**Percentage of markers**
SNPs having genotyping rate < 90%	13,993		31,050	
Total of retained SNPs	*495,515*	*81.28*	*478,458*	*78,48*
To keep SNPs having minor allele frequency > 5%				
SNPs having minor allele frequency < 5%	123,751		–	
Total of retained SNPs	*371,764*	*60.98*	–	–
To delete homozygote markers within parents				
Homozygote markers	–		264,623	
Total of retained SNPs	–	–	*213,835*	*35.07*
To delete same heterozygote markers within parents				
Same heterozygote markers	–		40,860	
Total of retained SNPs	–	–	*172,975*	*28.37*
To delete redundant SNPs in the genome				
Redundant SNPs	7489		10,857	
Total of retained SNPs	*364,275*	*59.75*	*162,118*	*26.59*
To delete distorded and identical markers				
Distorded and identical markers	–		160,181	
Total of retained SNPs	–	–	*1937*	*0.32*
			'Franquette' map: 849	
			'UK 6–2' map: 1088	

(PCA). The fastSTRUCTURE analysis infers accession ancestry from genotypic information and permitted us to determine the best number of clusters (K). The most likely K subpopulations were K = 2 and K = 3 (Figure S5). At K = 2, admixture proportions clustered the accessions according to their geographical origin. In particular, the cluster in purple named "Western Europe and America" includes 86 accessions from Austria, Chile, England, France, Germany, Netherlands, Portugal, Serbia, Slovenia, Spain, Switzerland and USA. The cluster in green named "Eastern Europe and Asia" includes 50 accessions from Afghanistan, Bulgaria, China, Greece, Hungary, India, Iran, Israel, Japan, Poland, Romania, Russia and Central Asia (Fig. 2). At K = 3, a new cluster includes all the hybrids and admixed accessions from France and USA (Fig. 2, Table S1).

PCA shows similar clustering of our germplasm collection as fastSTRUCTURE (Figure S6). PC1, which explains 7.37% of total variance, separated the "Western Europe and America" (WEAm) accessions from the "Eastern Europe and Asia" (EEAs) accessions. PC2 accounted for 5.80% of variance explained and separated the hybrids and admixed accessions from France and USA, observed with K = 3 in fastSTRUCTURE.

Relatedness of the GWAS panel

In addition to population structure, we investigated the familial relatedness within our association panel by estimating kinship coefficient (k) with the KING method. To identify first-degree relationships and differentiate "parent-offspring" from "full sibling" pairs, we used the estimates of k and the proportion of zero identical-by-state (IBS0) observed in the F_1 progeny (Figure S7). In particular, we defined all pairwise relationships in the GWAS panel with $k > 0.17$ and $0 < IBS0 < 0.019$ to be parent-offspring relationships. Results confirmed known pedigrees, particularly for the hybrids accessions and the modern cultivars from France and the USA. We also identified new relationships, such as that between 'Grosvert n°1' and 'Verdelet', French landraces from the departments of Dordogne and Corrèze, which may be

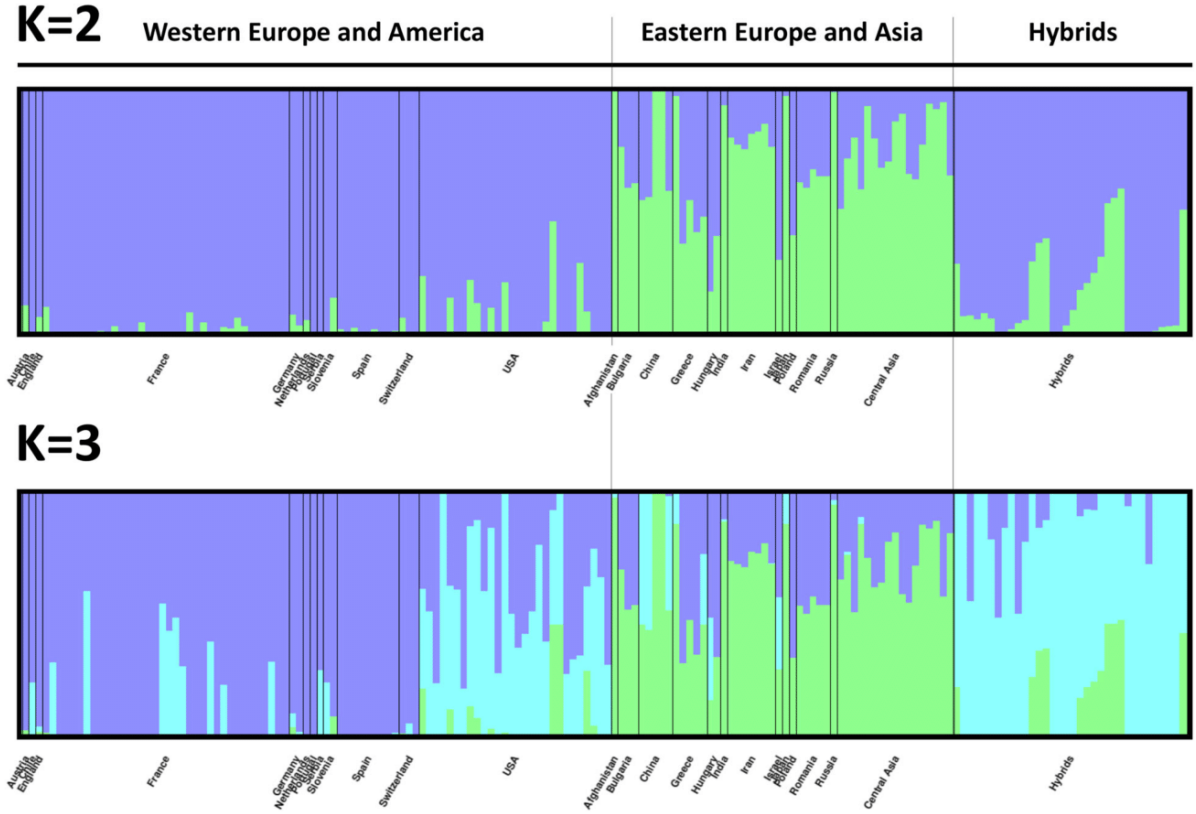


Fig. 2 Structure of the GWAS panel. The fastSTRUCTURE software was used. Bar plot of individual ancestry proportions (Q values) for the genetic cluster inferred using the whole set of 364,275 robust SNPs. For K = 2, accessions are geographically separated in two main groups: the purple group for 'Western Europe and America' accessions, and the green group for 'Eastern Europe and Asia' accessions. For K = 3, the blue group, highlights hybrids

full-sibs (Figure S8). Moreover, 'Ashley' and 'Payne', said to be identical, show the highest kinship coefficient.

Genome-wide analysis for bearing H abit

For bearing habit, we found no influence of population structure (PC = 0 according to the 'model selection' function implemented in GAPIT; Table S2). We used multi-locus mixed model (MLMM), and Fixed and random model Circulating Probability Unification method (FarmCPU). GWAS results using both models showed a significant association on Chr11 with bearing habit, using only the 2019 data (Fig. 3). The most significantly associated marker was the SNP 'AX-171191765' (physical position: 20,831,267 bp; *p*-value: 2.98E-14), and two additional associations are also found on Chr6 (SNP 'AX-171108125'; p-value = 4.08E-09) and Chr8 (SNP 'AX-171083929'; p-value = 1.47E-08), according to the false discovery rate (FDR) threshold (≥ 0.05).

The boxplots show the bearing habit phenotypes of 2019 for the different alleles of the three associated SNPs (Fig. 3). For the most significantly associated SNP 'AX-171191765', the allele G is linked to a terminal bearing habit, whereas the allele C is linked to a lateral bearing habit (R^2 = 34.3%, allelic estimated effect = 2.59), leading to an increased yield.

Association and linkage mapping for Budbreak date and female flowering dates

Using 1937 SNPs (Table 2), the 'Franquette' and 'UK 6–2' parental genetic maps constructed have a length of

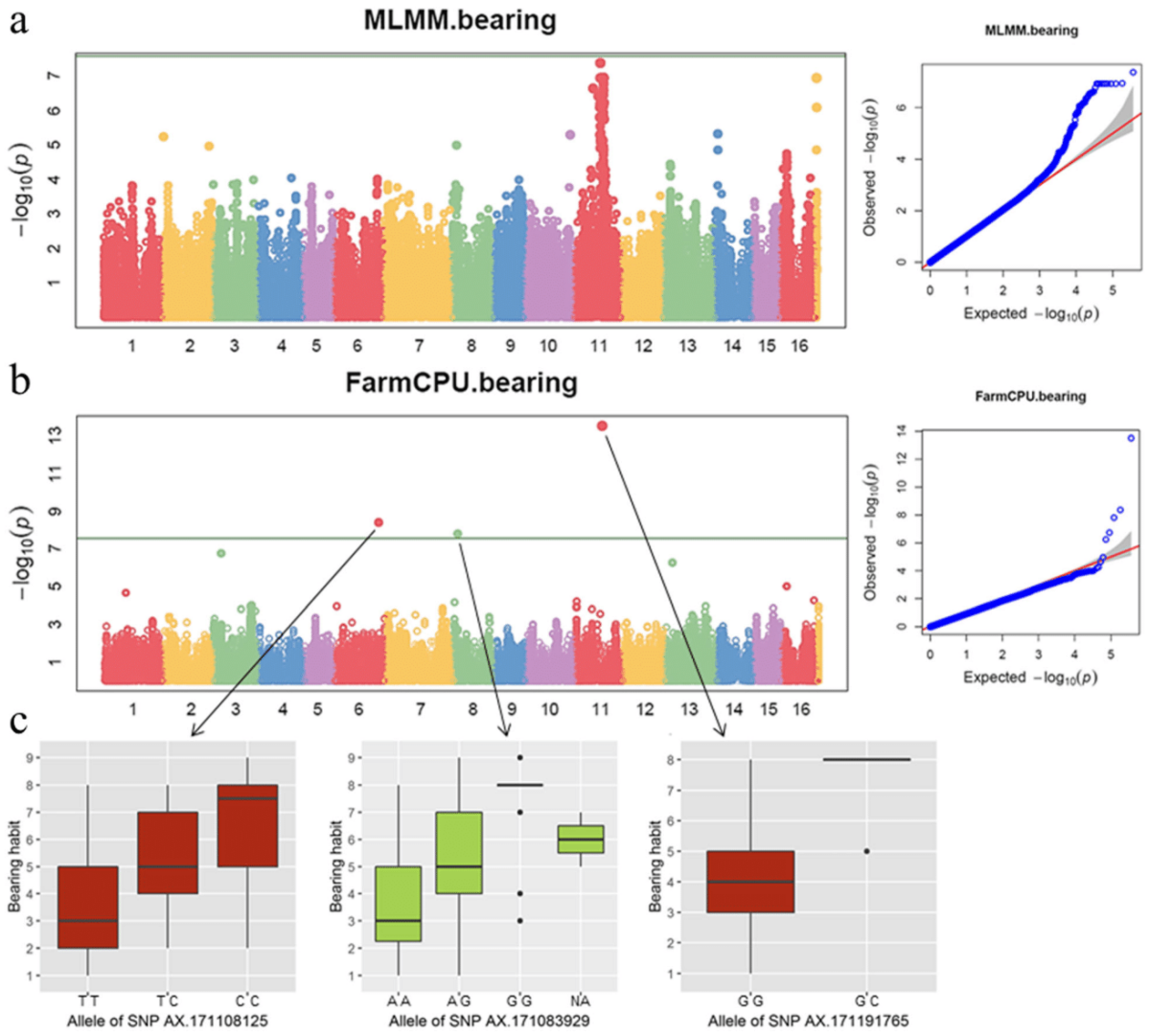


Fig. 3 GWAS results for bearing habit using 2019 data. Manhattan plots followed by Q-Q plots using **a**) MLMM model, **b**) FarmCPU model, and **c**) box plots of the allele effects for the 3 SNPs associated with bearing habit

1015 and 1346 cM, and a number of markers of 849 and 1088 SNPs, respectively (Table S3). The marker names of the genetic maps were changed with the corresponding chromosome number and its physical position for a better visualization (Figure S9). For all the phenology-related traits, we also found that population structure did not influence phenology in our GWAS panel. Both GWAS and classical QTL mapping identified marker-trait associations for budbreak date in the same region on Chr1 (Fig. 4). The most significant associated SNP 'AX-171179714' on the Chr1 (physical position: 6,514, 832 bp) was found using the Best Linear Unbiased

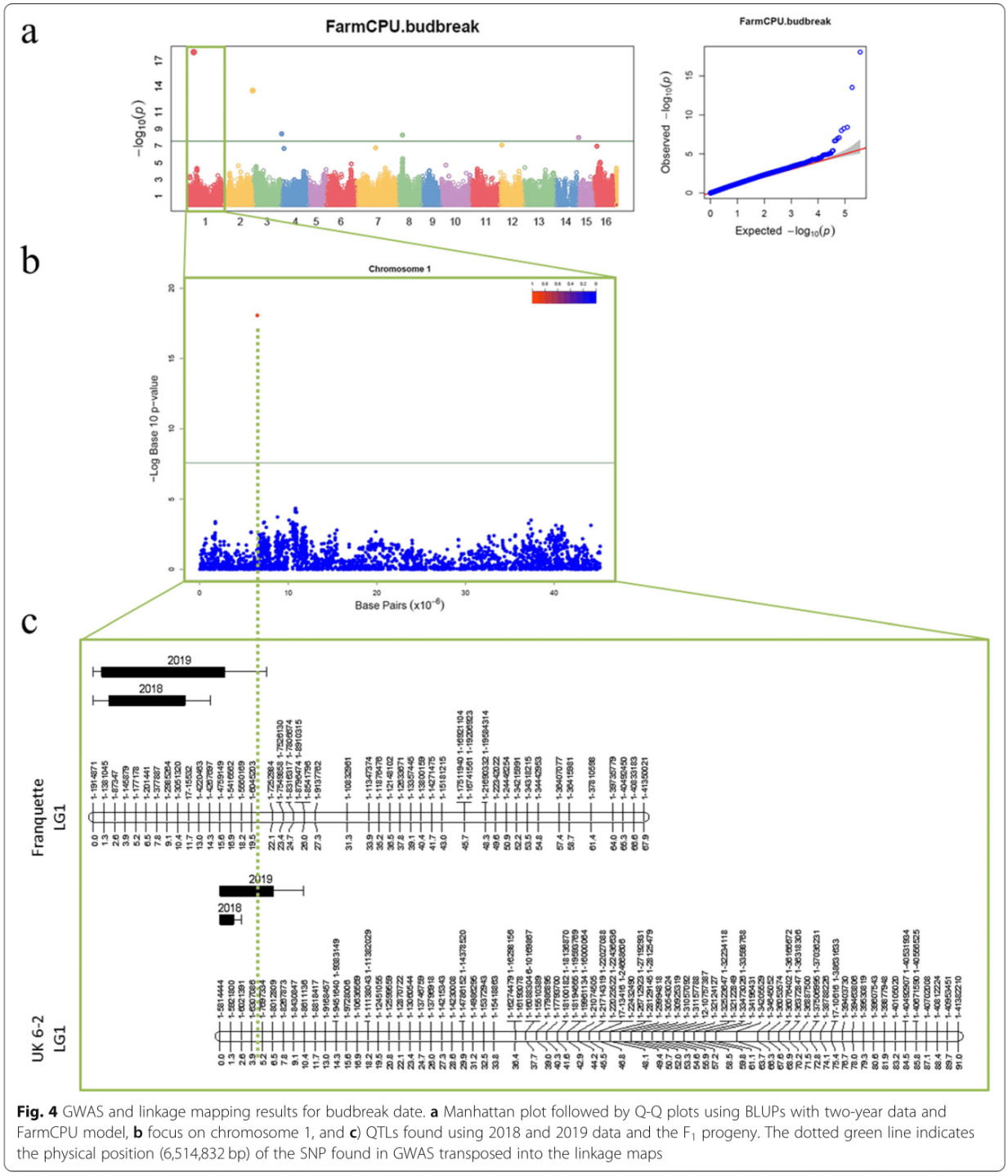


Fig. 4 GWAS and linkage mapping results for budbreak date. **a** Manhattan plot followed by Q-Q plots using BLUPs with two-year data and FarmCPU model, **b** focus on chromosome 1, and **c**) QTLs found using 2018 and 2019 data and the F_1 progeny. The dotted green line indicates the physical position (6,514,832 bp) of the SNP found in GWAS transposed into the linkage maps

Predictions (BLUPs) of two-year data and co-localizes with the major QTLs identified for both parents in 2019 using the F_1 progeny data. The allele T of this SNP is linked to a late budbreak date (R^2 = 30.6%, allelic estimated effect = 5.9) (Table 3). We also ran a Kruskal-Wallis test to find if the phenotypic differences were significant among the three genotypes using the different phenotypic datasets, and this allelic effect remains consistent (*p*-values = 1.84E-13 for two-year data, 6.63E-12 for 2018, 9.76E-13 for 2019, and 2.61E-09 for legacy data; Fig. 5). The co-localizing major QTLs found in 2019 in LG 1 in 'Franquette' and 'UK 6–2' explain 23.9 and 34.8% of the budbreak date variance, respectively. In addition, GWAS with two-year data found four additional associations on chromosomes 2, 4, 8 and 15, while the classical linkage mapping analysis identified minor QTLs on linkage groups 6, 11, 12 and 14.

The high power of our gene tagging approach based on both GWAS and QTL mapping, was also confirmed for beginning, peak, and end, of female flowering dates. The SNP 'AX-170990138' (physical position: 9,298,520 bp) on Chr1 was systematically found associated with all stages of female flowering (Table 3). For beginning female flowering date, we found this SNP associated using two-year data, and using each year separately. We also observed this marker-trait association for peak female flowering date using legacy data, and for end female flowering date with two-year data and with 2019 data. The SNP 'AX-170990138' is 2.8 Mbp apart from the most significant marker-trait association found for the budbreak date on Chr1, and the allele G of this SNP is linked to a delayed female flowering, with a R^2 ranging from 34.8 to 39.6%, and an allelic estimated effect ranging from 3.4 to 4.5, depending on the stage and the dataset. We identified additional QTLs for all three stages of female flowering but the most significant ones, segregating in both parental maps, co-localize with those previously found associated with the budbreak date on Chr1 (Table 3).

Besides the major QTL on Chr1 identified with both GWAS and QTL mapping, we found three significant associations also on Chr7 for all three stages. These three SNPs are located in a region of about 23 to 25 Mb (Table 3). In addition, we found QTLs on LGs 2 and 14 in 'Franquette' map, and on LGs 3, 9, 11 and 12 in 'UK 6–2' map.

Association and linkage mapping for Heterodichogamy and male flowering dates

Results for male flowering are similar to the female flowering results in that a few SNPs in a very close region are associated with all three stages (Table 4). On Chr11, we found four associated SNPs depending on the stage and the dataset, in a region of about 31.8 Mbp and a window of 52 kb. The most significant QTLs for all three stages of male flowering for both parental maps co-localize with those previously identified as associated with budbreak date and the three stages of female flowering. Two additional QTLs on 'UK 6–2' map were found on LG 11 for peak male flowering date, using two-year data (26,141,527 – 35,537,934 bp), and for the end of male flowering using 2019 data (27,685,560 – 35,537,934 bp), supporting the GWAS results.

For heterodichogamy trait (computed by subtracting peak female date from peak male date; Table S5), the significant associations found with GWAS using two-year data and legacy data co-localized with the associations identified for male flowering dates on Chr11 in the region of about 31.8 Mbp.

Candidate genes for bearing habit and phenology-related traits using the walnut genome

By combining GWAS and QTL results and considering their consistency over phenotypic datasets, we decided to focus on a robust subset of eight loci to find candidate genes for bearing habit and phenology-related traits (Table 5). Using HaploView v4.2 software and the new chromosome-scale reference walnut genome v2.0 [36], several interesting coding sequences were found within the defined Linkage Disequilibrium (LD) blocks for different traits. The SNP 'AX-171557178' on Chr2 and associated with budbreak date, falls within a candidate gene encoding for a *putative BPI/LBP family protein At1g04970*. The corresponding LD block of 78 kb also contains a candidate gene coding for a *GrpE-like protein* and one encoding a *65-kDa microtubule-associated protein 1-like*. Only one candidate gene, encoding an *uncharacterized protein LOC108987988*, overlaps with the most significant SNP associated with budbreak date, 'AX-171179714' on Chr1.

The two SNPs on Chr11 associated with all three stages of male flowering date and with heterodichogamy, belong to the same LD block of 19 kb. Within this block is located a candidate gene encoding for a *probable trehalose-phosphate phosphatase D*. The other SNP on Chr4 associated with all three stages of male flowering date, belongs to a LD block of 63 kb comprising a candidate gene encoding for a *trichome birefringence-like 13 protein*. Only one candidate gene was found in LD with the associated SNP on Chr1 for all three stages of female flowering date. The SNP 'AX-170990138' belongs to a small LD block on Chr1 spanning from 9,298,520 to 9,300,288 bp (Fig. 6). The identified candidate gene of 1.75 kb (interval from 9,298,602 to 9,300,352 bp) overlaps with the LD block and encodes a *chromosome transmission fidelity protein 8 homolog*.

Table 3 Summary of association and linkage mapping results related to bearing habit, budbreak date and female flowering dates

Dataset	GWAS using FarmCPU model						QTLs detection using MIM method						
	SNP	Chr [a]	Physical position	Significance [b]	R^2 [c]	Alleles/Effect [d]	Peak SNP [e]	LG [f]	Physical position	Confidence interval [g]	LOD [h]	P.E.V. [i]	d [j]
Bearing habit													
2019	AX-171108125	6	33,909,308	4.08E-09	30.8	C/T (0.7)							
	AX-171083929	8	3,102,647	1.47E-08	32.5	A/G (−0.8)							
	AX-171191765	11	20,831,267	2.98E-14	34.3	C/G (2.6)							
Budbreak date													
1989–2016	AX-171509395	5	21,137,700	1.38E-10	9.0	C/T (4.1)							
	AX-171490851	7	37,022,231	5.76E-11	35.0	C/G (−4.9)							
	AX-170810238	8	23,687,589	7.78E-12	25.6	C/T (−4.1)							
2018–2019	AX-171179714	1	6,514,832	8.95E-19	30.6	G/T (−5.9)	AX-171180038	F1	177,178	2.8–10.8	11.9	23.7	5.2
	AX-171557178	2	36,586,423	2.84E-14	27.3	A/T (−4.7)	AX-170771092	U1	5,814,444	0.0–2.8	29.5	44.3	−6.9
	AX-171040881	4	591,186	3.70E-09	4.8	G/T (−3.2)	AX-171486072	U6	2,467,998	0.0–46.7	6.5	6.7	−2.6
	AX-171522523	8	5,949,933	5.41E-09	3.2	A/T (−2.6)	AX-170627138	U11	26,651,484	0.0–30.2	4.7	4.7	−2.0
	AX-171509630	15	1,111,543	9.92E-09	6.8	A/G (2.6)	AX-170901703	U12	19,779,190	39.0–67.4	6.8	7.7	2.8
							AX-170818106	F14	9,871,222	3.6–50.0	6.0	11.2	3.5
2018	AX-171520811	1	6,537,230	2.54E-10	29.1	A/G (4.3)	AX-171180038	F1	177,178	0.0–14.4	6.3	31.0	5.4
	AX-171557178	2	36,586,423	4.31E-09	25.1	A/T (−3.4)	AX-170771092	U1	5,814,444	0.0–2.6	15.4	50.2	−7.0
	AX-170682386	15	1,162,635	1.05E-08	18.8	A/G (3.6)	AX-170760959	U6	2,438,026	0.0–61.4	3.4	7.9	−2.6
	AX-170691861	16	3,791,224	1.47E-10	25.5	C/T (−4.3)							
2019	AX-171557178	2	36,586,423	2.70E-11	26.2	A/T (−4.6)	AX-171180038	F1	177,178	0.0–21.4	5.6	23.9	5.4
	AX-171489741	4	3,486,412	8.12E-10	26.2	C/G (−4.8)	AX-170771092	U1	5,814,444	0.0–10.3	11.1	34.8	−6.7
	AX-171139350	6	31,522,110	6.17E-09	< 0.1	C/T (3.6)	AX-170901703	U12	19,779,190	27.8–76.4	3.7	9.2	3.3
	AX-170938460	7	22,604,203	8.50E-13	26.8	A/G (5.0)							
	AX-171047138	10	36,604,822	7.89E-09	0.1	A/G (4.5)							
	AX-170691931	16	3,762,646	2.69E-10	24.0	A/G (−4.9)							
Beginning female flowering date													
2018–2019	AX-170990138	1	9,298,520	1.56E-09	35.0	A/G (−3.4)	AX-171521263	F1	201,441	3.4–10.2	12.9	30.6	5.2
	AX-170938895	7	23,018,189	6.62E-13	31.4	A/C (4.0)	AX-170771092	U1	5,814,444	< 0.5	35.1	46.5	−7.1
							AX-170684860	U3	2,653,247	0.0–105.6	5.0	2.6	1.9
							AX-171583903	U6	2,237,730	0.0–82.9	5.9	4.1	−2.1
							AX-170825127	U11	25,463,971	0.0–8.3	9.3	6.5	−2.8
							AX-170901663	U12	19,737,912	44.5–60.8	8.1	6.9	2.7
							AX-170818106	F14	9,871,222	0.0–60.5	4.3	3.8	2.5

Table 3 Summary of association and linkage mapping results related to bearing habit, budbreak date and female flowering dates *(Continued)*

Dataset	GWAS using FarmCPU model						QTLs detection using MIM method						
	SNP	Chr [a]	Physical position	Significance [b]	R^2 [c]	Alleles/Effect [d]	Peak SNP [e]	LG [f]	Physical position	Confidence interval [g]	LOD [h]	P.E.V. [i]	d [j]
							AX-170786852	U16	20,462,383	21.7–77.2	5.1	3.9	2.0
2018	AX-170990138	1	9,298,520	1.79E-11	35.4	A/G (−4.2)	AX-171180038	F1	177,178	0.0–14.2	6.6	32.2	5.7
	AX-171029453	4	4,123,643	7.38E-13	23.6	C/T (−4.9)	AX-170771092	U1	5,814,444	0.0–3.1	14.6	47.6	−7.2
2019	AX-170990138	1	9,298,520	3.33E-13	34.8	A/G (−4.5)	AX-171180038	F1	177,178	0.0–16.4	5.6	21.6	4.7
	AX-171029453	4	4,123,643	1.55E-10	25.6	C/T (−4.2)	AX-170771092	U1	5,814,444	0.0–2.3	23.0	44.4	−7.6
	AX-170938895	7	23,018,189	2.99E-13	35.1	A/C (5.1)	AX-170684860	U3	2,653,247	0.0–23.8	5.3	5.8	2.7
	AX-170586625	16	626,899	2.92E-12	2.6	C/T (−4.2)	AX-171569583	U6	12,329,274	26.3–62.5	7.5	8.5	−3.3
							AX-170825127	U11	25,463,971	39.9–68.1	7.5	8.5	−3.2
							AX-170901663	U12	19,737,912	0.0–3.6	7.5	8.7	3.3
Peak female flowering date													
1989–2016	AX-170990138	1	9,298,520	8.15E-13	39.6	A/G (−3.7)							
	AX-171596225	6	14,925,533	1.39E-09	26.7	A/G (3.0)							
	AX-171029018	7	48,197,343	3.80E-09	0.1	A/C (2.2)							
	AX-171161427	7	37,022,706	9.52E-09	35.9	C/T (3.0)							
	AX-171093716	10	23,540,554	5.39E-09	8.3	A/G (−4.6)							
	AX-171553780	10	6,443,284	1.65E-08	0.4	A/T (2.2)							
	AX-171161986	13	3,292,458	2.11E-08	30.9	A/G (2.8)							
2018–2019	AX-170990138	1	9,298,520	4.75E-11	36.2	A/G (−4.5)	AX-171180038	F1	177,178	2.6–11.2	9.7	22.3	5.0
	AX-171204022	7	25,355,784	6.12E-09	31.2	A/G (−3.6)	AX-170771092	U1	5,814,444	0.0–1.9	42.5	49.0	−8.3
							AX-170869355	F2	35,198,358	18.9–67.9	5.2	7.4	−3.2
							AX-171121327	U3	1,942,419	0.0–33.7	6.4	4.3	2.3
							AX-171204628	U9	21,282,336	32.9–57.3	5.9	3.7	2.3
							AX-170627138	U11	26,651,484	0.0–7.3	13.1	7.7	−3.5
							AX-170901663	U12	19,737,912	37.3–65.2	10.0	6.9	3.0
2018	AX-170990138	1	9,298,520	1.41E-08	35.4	A/G (−3.6)	AX-171521263	F1	201,441	0.0–22.0	4.5	23.4	5.4
	AX-170975750	6	22,755,843	1.39E-08	< 0.1	A/G (−3.8)	AX-170771092	U1	5,814,444	0.0–4.9	13.7	45.5	−7.8
	AX-170676955	9	8,265,700	3.56E-09	6.7	A/G (−3.3)	AX-170697456	U4	5,261,727	0.0–49.2	3.3	8.5	3.0
							AX-170787652	U15	151,849	0.0–33.6	3.8	10.0	−3.5
2019	AX-170990138	1	9,298,520	2.70E-11	35.6	A/G (−4.4)	AX-171180038	F1	177,178	0.0–25.2	4.5	18.2	4.5
	AX-170938895	7	23,018,189	6.82E-15	36.7	A/C (5.1)	AX-170771092	U1	5,814,444	0.0–2.0	24.0	46.4	−7.7
							AX-170684716	U3	2,526,842	0.0–50.7	4.4	4.5	2.3
							AX-171569583	U6	12,329,274	0.0–98.6	3.9	3.8	−2.2

Table 3 Summary of association and linkage mapping results related to bearing habit, budbreak date and female flowering dates *(Continued)*

Dataset	GWAS using FarmCPU model						QTLs detection using MIM method						
	SNP	Chr [a]	Physical position	Significance [b]	R^2 [c]	Alleles/Effect [d]	Peak SNP [e]	LG [f]	Physical position	Confidence interval [g]	LOD [h]	P.E.V. [i]	d [j]
							AX-171222942	U8	5,077,987	0.0–98.8	3.8	3.8	2.1
							AX-171204628	U9	21,282,336	37.2–57.3	4.5	5.1	2.5
							AX-170582801	U11	26,090,506	0.0–7.3	9.4	11.0	−3.7
							AX-171033431	U12	17,150,374	34.4–70.9	5.1	5.3	2.6
							AX-170581901	U16	17,340,045	2.2–66.1	4.0	4.0	2.2
End female flowering date													
2018–2019	AX-170990138	1	9,298,520	8.81E-11	39.3	A/G (−4.1)	AX-171180038	F1	177,178	2.9–10.5	10.8	24.8	5.2
	AX-171103669	1	3,411,327	8.46E-10	28.0	A/C (3.5)	AX-170771092	U1	5,814,444	< 0.5	32.9	49.9	−8.2
	AX-171199539	15	16,085,753	2.06E-10	< 0.1	A/G (−3.4)	AX-170868251	F2	34,498,992	32.8–67.9	7.2	13.7	4.0
							AX-171121327	U3	1,942,419	0.0–41.2	6.3	3.6	2.7
							AX-170620438	U5	18,147,969	0.0–78.7	5.4	2.9	2.3
							AX-171214747	U9	18,177,182	14.3–57.3	5.0	5.0	2.5
							AX-170825127	U11	25,463,971	0.0–13.8	6.6	3.4	−2.7
							AX-170818106	F14	9,871,222	0.0–56.8	5.8	7.7	3.5
2018	AX-171572440	9	20,815,377	3.24E-09	1.3	A/G (−2.9)	AX-171521263	F1	201,441	2.6–9.8	6.6	25.4	6.4
							AX-170771092	U1	5,814,444	0.0–7.1	13.0	48.0	−8.7
							AX-170867988	F2	34,288,275	18.6–67.9	4.0	13.8	−4.4
2019	AX-170990138	1	9,298,520	6.75E-11	38.1	A/G (−4.3)	AX-170771092	U1	5,814,444	0.0–3.6	19.8	40.1	−7.6
	AX-171029453	4	4,123,643	1.12E-10	27.6	C/T (−4.9)	AX-171210974	U3	2,714,686	0.0–14.7	6.5	8.5	3.4
	AX-170938854	7	22,982,368	1.14E-12	37.0	C/T (4.5)	AX-170682524	U5	18,023,534	0.0–14.0	6.7	8.7	3.5
	AX-170766461	8	27,866,613	7.85E-09	< 0.1	A/C (−5.3)	AX-170626929	U11	26,817,955	2.3–12.5	7.0	9.1	−3.6
	AX-170860236	14	9,448,288	1.25E-10	< 0.1	G/T (4.3)	AX-170901663	U12	19,737,912	13.8–84.9	3.9	4.7	2.5
	AX-170586625	16	626,899	1.39E-08	1.8	C/T (−3.1)							

[a] *Chr* abbreviation for Chromosome
[b] For GWAS panel, the significance value indicated is the p-value, which is significant if lower than the False Discovery Rate
[c] For GWAS panel, R^2 is the percentage explained variance corrected for genome-wide background
[d] For GWAS panel, the allelic effect is the difference in mean of measured trait between genotypes with one or other allele The sign is with respect to the allele that is first mentioned
[e] For F_1 progeny using QTLs detection, the indicated SNP is the peak SNP inside the QTL interval
[f] LG, abbreviation for Linkage Group with F and U for 'Franquette' and 'UK 6–2' parental maps respectively
[g] For F_1 progeny using QTLs detection, the position in cM indicated in brackets is the 99.9% confidence interval of the QTL
[h] For F_1 progeny using QTLs detection, the significance value indicated is the LOD score (LOD = logarithm of the odds ratio)
[i] For F_1 progeny using QTLs detection, P.E.V. is for percentage explained variance
[j] For F_1 progeny using QTLs detection, d is the estimated substitution effect of the QTL

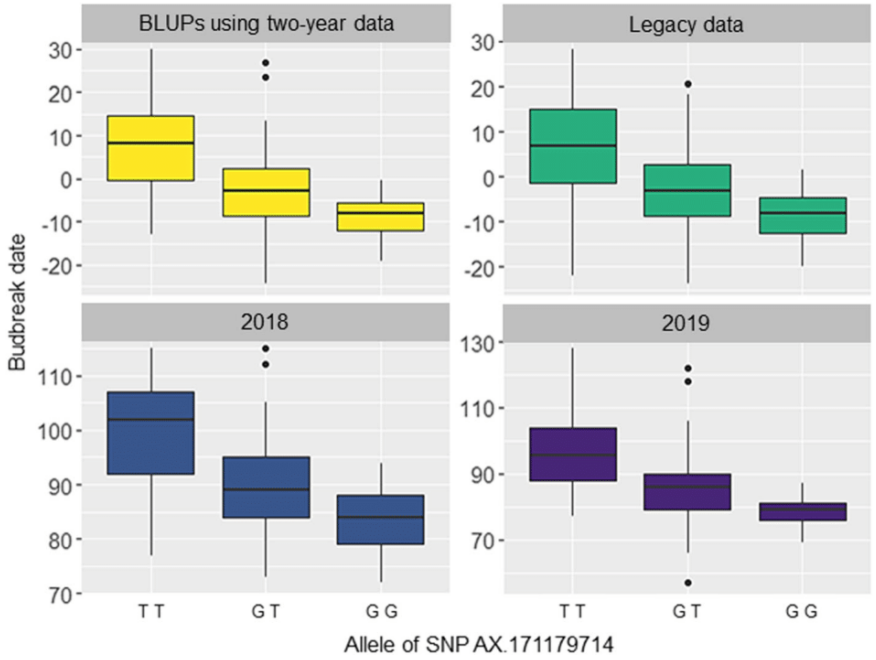


Fig. 5 Box plots of the allele effects for the SNP AX-171179714 associated with budbreak date. FarmCPU model was used and all datasets: BLUPs using two-year data, legacy data, 2018 data and 2019 data

Development of KASP marker for the Budbreak date and validation using 96 accessions of the walnut improvement program of the University of California, Davis

Primers related to the marker-trait association found on Chr1 for budbreak date (SNP 'AX-171179714') were developed. These primers were used to genotype 96 unreleased breeding line accessions of the Walnut Improvement Program of the University of California, Davis. Among them, 48 are early leafing accessions and 48 are late leafing accessions. Leafing date occurs 3 days after budbreak date. Our GWAS results showed that the allele T of the SNP 'AX-171179714' (versus G) has an allelic estimated effect of 5.9 (R^2 = 30.6%), linked with a delayed budbreak date. Since we designed the KASP primers using the complementary strand, we expect to find the allele A (versus C) as involved in late leafing genotypes. The three different genotypes were clearly separated for 95 accessions, whereas one of them could not be determined (Table S4). In order to determine if leafing dates were different among the three genotypes, we ran a Kruskal-Wallis test which showed significance (p-value = 6.88E-13). Then, we ran a pairwise comparison using Dunn's test with Bonferroni p-value adjustment. Results of pairwise comparisons showed that A/A vs. C/A, and A/A vs. C/C have a significant effect on leafing date (p-values = 1.4E-06 and 9.4E-12, respectively), contrary to C/A vs. C/C genotypes (p-value = 1). The homozygous accessions A/A have a median leafing date of almost 110 Julian days, while genotypes C/A and C/C have a median leafing date of only 76 Julian days (Fig. 7).

Discussion

Analysis of population structure and its effect on bearing habit and phenology-related traits

We found that the most likely K subpopulations were K = 2, with a clustering in two major groups according to their geographical origin, and K = 3 that highlights admixed accessions, including hybrids or modern cultivars from France and USA. The EEAs cluster in PCA resulted more scattered, indicating a higher level of SNP-based genetic diversity, and the three most diverse accessions of the EAAs cluster, located at the bottom-right corner of the PCA plot, are 'Shinrei' from Japan, 'JinLong1' from China, and 'EAA6' from Greece. The two last were previously found to be highly diverse by SSRs and included in an SSR-based core collection [37].

For the entire set of studied traits, we observed no influence of population structure on the phenotypes. Most likely, the familial relatedness among accessions accounted for most of the genetic variation affecting the traits studied, leading to redundant PCA information [38]. Even though our collection included accessions from different geographical areas, it consisted of 30% modern cultivars likely brought to the INRAE of Bordeaux as potential genitors for breeding purposes. Therefore, it is not

Table 4 Summary of association and linkage mapping results related to heterodichogamy and male flowering dates

Dataset	GWAS using FarmCPU model						QTLs detection using MIM method						
	SNP	Chr [a]	Physical position	Significance [b]	R^2 [c]	Alleles/Effect [d]	Peak SNP [e]	LG [f]	Physical position	Confidence interval [g]	LOD [h]	P.E.V. [i]	d [j]
Heterodichogamy													
1989–2016	AX-171595404	11	31,887,506	2.22E-42	69.6	C/G (2.7)							
2018–2019	AX-171218000	11	31,885,611	8.97E-49	77.2	A/G (− 2.7)							
Beginning male flowering date													
2018–2019	AX-171121884	2	21,247,915	1.74E-09	7.2	C/T (−2.5)	AX-171521263	F1	201,441	4.1–9.1	14.4	32.6	4.5
	AX-171578684	4	5,966,307	2.41E-10	6.0	C/T (4.5)	AX-170771092	U1	5,814,444	< 0.5	27.6	52.4	−5.3
	AX-170614801	4	3,642,466	2.78E-10	20.2	C/T (3.6)	AX-170983192	U9	22,032,363	23.6–57.3	5.5	4.9	2.2
	AX-171490851	7	37,022,231	3.56E-10	17.8	C/G (−2.3)							
	AX-170625459	10	10,656,433	2.73E-09	20.2	G/T (−2.4)							
	AX-171147588	11	31,874,617	1.58E-10	2.0	A/G (3.1)							
	AX-170794443	14	10,691,227	2.22E-09	< 0.1	C/T (−2.2)							
	AX-170692084	16	3,578,209	1.94E-08	18.2	G/T (−2.5)							
2018	AX-171589348	3	27,241,886	1.24E-10	< 0.1	C/G (−2.9)	AX-171521263	F1	201,441	2.1–10.4	7.6	36.0	3.8
	AX-171559197	6	13,733,426	9.96E-17	20.7	C/T (3.6)	AX-170771092	U1	5,814,444	0.0–2.0	17.5	50.9	−4.6
	AX-170937789	7	30,051,478	2.70E-12	18.0	C/T (−3.7)							
	AX-170625459	10	10,656,433	4.08E-09	20.6	G/T (−2.2)							
2019	AX-171147582	11	31,882,831	6.23E-13	1.7	C/T (6.2)	AX-171521263	F1	201,441	0.7–13.4	6.4	28.1	5.5
	AX-171525706	14	28,381,698	1.38E-10	10.1	C/G (4.9)	AX-170771092	U1	5,814,444	0.0–5.7	9.2	33.5	−5.9
	AX-170692024	16	3,630,320	1.27E-08	22.8	C/T (−4.2)							
Peak male flowering date													
2018–2019	AX-171578684	4	5,966,307	2.17E-09	6.1	C/T (6.8)	AX-171180038	F1	177,178	3.4–10.2	17.1	32.2	5.4
	AX-171147588	11	31,874,617	1.34E-10	1.9	A/G (4.1)	AX-170771092	U1	5,814,444	< 0.5	30.2	47.0	−6.2
							AX-170869355	F2	35,198,358	14.5–67.9	4.6	6.6	−2.3
							AX-170615140	U4	3,291,736	0.0–24.4	6.2	5.0	2.4
							AX-171478663	U5	7,984,929	30.6–64.9	9.2	9.7	3.0
							AX-171548959	U6	12,747,130	0.0–80.5	6.1	5.5	−2.4
							AX-171214747	U9	18,177,182	23.7–57.3	5.1	5.1	2.2
							AX-171030926	U11	33,195,441	2.3–49.4	4.8	5.2	1.8
							AX-170818106	F14	9,871,222	0.0–58.2	5.6	8.3	2.8
2018	AX-170672894	1	1,835,862	4.05E-11	20.6	G/T (3.1)	AX-171180038	F1	177,178	0.0–12.9	7.2	30.1	4.2
	AX-170802286	4	414,952	8.03E-10	< 0.1	A/G (−3.0)	AX-170771092	U1	5,814,444	0.0–5.1	15.8	49.8	−5.5
	AX-171578684	4	5,966,307	1.04E-08	5.5	C/T (4.3)							

Table 4 Summary of association and linkage mapping results related to heterodichogamy and male flowering dates *(Continued)*

Dataset	GWAS using FarmCPU model						QTLs detection using MIM method						
	SNP	Chr [a]	Physical position	Significance [b]	R^2 [c]	Alleles/Effect [d]	Peak SNP [e]	LG [f]	Physical position	Confidence interval [g]	LOD [h]	P.E.V. [i]	d [j]
	AX-171559197	6	13,733,426	3.12E-12	20.3	C/T (3.2)							
	AX-171571158	12	1,647,435	5.48E-09	< 0.1	A/C (−2.4)							
2019	AX-171578684	4	5,966,307	5.63E-09	8.0	C/T (9.1)	AX-171180038	F1	177,178	0.0–15.6	6.7	28.3	6.1
	AX-171490851	7	37,022,231	1.13E-10	19.2	C/G (−4.3)	AX-170771092	U1	5,814,444	< 0.5	11.8	40.1	−7.6
	AX-171147582	11	31,882,831	1.47E-10	1.5	C/T (5.1)							
	AX-170849618	12	14,608,835	1.85E-08	< 0.1	C/T (−2.9)							
	AX-170692024	16	3,630,320	6.46E-15	22.6	C/T (−6.2)							
End male flowering date													
2018–2019	AX-171578684	4	5,966,307	1.01E-09	5.4	C/T (7.9)	AX-171521263	F1	201,441	4.8–8.1	19.0	33.6	6.5
	AX-171560933	5	604,688	1.30E-08	19.2	C/G (−2.8)	AX-170771092	U1	5,814,444	< 0.5	29.8	43.2	−7.1
	AX-171559197	6	13,733,426	6.50E-10	18.2	C/T (3.3)	AX-171539906	F2	34,841,469	29.2–67.9	5.3	6.3	−2.9
							AX-171002599	U5	20,087,966	43.9–78.7	6.3	3.4	2.6
							AX-170677995	U6	1,908,834	0.0–24.0	8.8	7.8	−3.3
							AX-170901663	U12	19,737,912	12.8–85.3	5.6	3.9	2.4
							AX-170818106	F14	9,871,222	0.0–60.6	6.2	9.2	3.2
2018	AX-171559197	6	13,733,426	5.14E-13	20.0	C/T (4.1)	AX-171180038	F1	177,178	1.5–11.1	7.4	30.7	5.4
	AX-171142468	11	31,834,919	9.09E-11	3.4	A/G (5.3)	AX-170771092	U1	5,814,444	0.0–2.2	13.4	44.3	−6.5
	AX-171531310	11	29,400,625	1.53E-08	0.7	C/T (3.3)							
2019	AX-170817347	1	32,640,957	1.07E-09	< 0.1	C/T (4.1)	AX-171180038	F1	177,178	1.2–11.6	9.8	36.1	7.7
	AX-171578684	4	5,966,307	2.05E-09	7.7	C/T (9.9)	AX-170771092	U1	5,814,444	0.0–2.1	18.4	42.6	−8.7
	AX-171217997	11	31,886,671	3.10E-09	0.2	A/G (4.8)	AX-171165000	U5	10,051,966	23.5–78.7	4.1	6.1	3.2
	AX-170692024	16	3,630,320	2.07E-08	23.4	C/T (−4.1)	AX-171583903	U6	2,237,730	0.0–81.0	4.7	6.9	−3.4
							AX-170838105	U11	35,888,157	13.9–49.4	3.7	5.3	2.8
							AX-170901663	U12	19,737,912	25.7–74.8	5.5	8.3	3.8

[a] Chr, abbreviation for Chromosome
[b] For GWAS panel, the significance value indicated is the p-value, which is significant if lower than the False Discovery Rate
[c] For GWAS panel, R^2 is the percentage explained variance corrected for genome-wide background
[d] For GWAS panel, the allelic effect is the difference in mean of measured trait between genotypes with one or other allele The sign is with respect to the allele that is first mentioned
[e] For F_1 progeny using QTLs detection, the indicated SNP is the peak SNP inside the QTL interval
[f] LG, abbreviation for Linkage Group with F and U for 'Franquette' and 'UK 6–2' parental maps respectively
[g] For F_1 progeny using QTLs detection, the position in cM indicated in brackets is the 99.9% confidence interval of the QTL
[h] For F_1 progeny using QTLs detection, the significance value indicated is the LOD score (LOD = logarithm of the odds ratio)
[i] For F_1 progeny using QTLs detection, P.E.V. is for percentage explained variance
[j] For F_1 progeny using QTLs detection, d is the estimated substitution effect of the QTL

Table 5 List of candidate genes for bearing habit and phenology-related traits

Trait	SNP associated using GWAS	Chromo-some	Physical position [a]	LD block interval [a] / Definition block method	GeneID	Physical position interval [a]	Functional annotation
Bearing habit	AX-171191765	11	20,831,267	20,831,267 - 20,865,013 (33 kb) / Solid spine of LD	109,017,990	20,831,874 - 20,834,977	uncharacterized protein LOC109017990
Budbreak date	AX-171179714	1	6,514,832	6,511,521 - 6,528,313 (16 kb) / Solid spine of LD	**108,987,988**	**6,512,784 - 6,520,863**	**uncharacterized protein LOC108987988**
Budbreak date	AX-171557178	2	36,586,423	36,535,050 - 36,613,482 (78 kb) / Solid spine of LD	108,983,408	36,536,110 - 36,536,482	GrpE-like protein
					108,984,642	36,549,629 - 36,555,410	65-kDa microtubule-associated protein 1-like
					108,984,651	36,557,793 - 36,562,849	homolog of mammalian lyst-interacting protein 5
					108,983,414	36,564,185 - 36,568,214	indole-3-acetate O-methyltransferase 1-like
					108,984,666	36,576,304 - 36,582,108	uncharacterized protein LOC108984666
					108,984,671	**36,582,611 - 36,587,786**	**putative BPI/LBP family protein At1g04970**
					108,984,682	36,601,277 - 36,607,484	transcription initiation factor TFIID subunit-6-like
					108,984,689	36,610,257 - 36,612,478	cell division protein FtsZ homolog 1, chloroplastic-like
Male flowering date	AX-171147588	11	31,874,617	31,868,402 - 31,887,998 (19 kb) / Solid spine of LD	108,984,910	31,868,783 - 31,870,048	uncharacterized protein LOC108984910
					108,984,907	31,884,256 - 31,887,060	probable trehalose-phosphate phosphatase D
Male flowering date	AX-171578684	4	5,966,307	5,946,348 - 6,009,859 (63 kb) / Solid spine of LD	109,020,754	5,953,569 - 5,954,197	uncharacterized protein LOC109020754
					109,022,114	5,954,514 - 5,956,355	trichome birefringence-like 13 protein
					109,022,113	5,960,058 - 5,960,714	uncharacterized protein LOC109022113
					109,022,028	5,984,567 - 5,985,798	uncharacterized protein LOC109022028, transcript variant X2
Male flowering date	AX-171559197	6	13,733,426	13,733,274 - 13,733,426 (< 1 kb) / Confidence interval		–	–
Hetero-dichogamy	AX-171218000	11	31,885,611	31,868,402 - 31,887,998 (19 kb) / Solid spine of LD	108,984,910	31,868,783 - 31,870,048	uncharacterized protein LOC108984910
					108,984,907	**31,884,256 - 31,887,060**	**probable trehalose-phosphate phosphatase D**
Female flowering date	AX-170990138	1	9,298,520	9,298,520 - 9,300,288 (1 kb) / Confidence interval	108,998,539	9,298,602 - 9,300,352	chromosome transmission fidelity protein 8 homolog, mRNA

[a] Physical position given in bp
The candidate genes in bold overlap the physical position of the associated SNP

surprising to observe accessions of the same geographical origin that are both early and late in phenology.

Strong correlations between traits Lead to close associations

Results found using the F_1 progeny show that strong QTLs for all the stages of female flowering date segregating in both parents on LG 1, co-localize with strong QTLs for budbreak date. In walnut trees, budbreak phenotyping relies on buds that will produce both leaves and female flowers. Those traits are, therefore, correlated and likely controlled by common genetic and metabolic pathways, as suggested by the identification of QTLs within the same genomic region on Chr1. Similar results were observed in walnut for

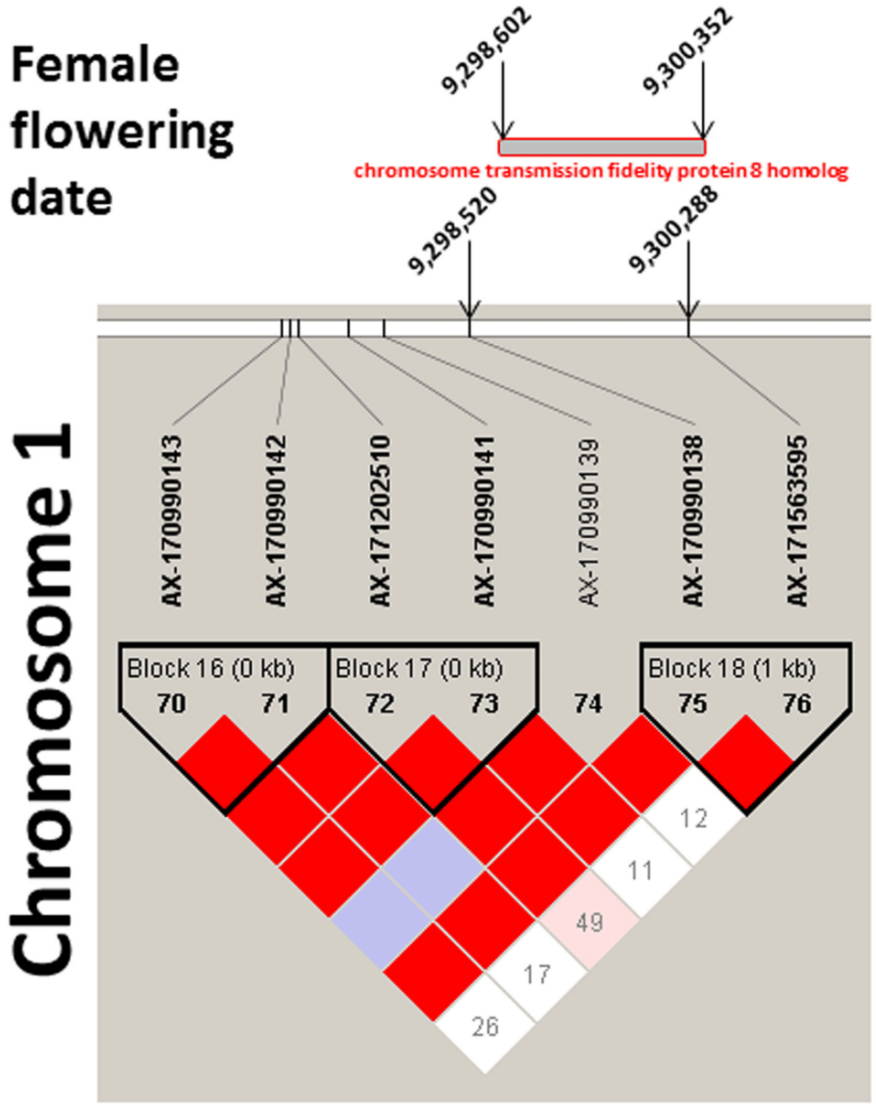


Fig. 6 LD block representation for the SNP AX-170990138 associated with female flowering dates. This SNP overlaps with a candidate gene, coding for a *chromosome transmission fidelity protein 8 homolog*

leafing and harvest dates [35]. On the other hand, GWAS enabled further dissection of the genetic control of these traits by identifying a location for the most significant marker-trait associations for budbreak date at about 2.8 Mbp apart from the SNPs associated to all stages of the female flowering date. This is one of the major advantages of GWAS, whose higher resolution defines more precisely the underlying regions of traits of interest compared to classical QTL mapping.

While the major loci controlling the traits studied were identified with both GWAS and linkage mapping, this did not happen for the minor associations. For instance, for budbreak date, GWAS found a second trait-locus association on Chr2, while one minor QTL was detected on LG 6 of the 'UK 6–2' map. This finding can be explained in two ways: (*i*) the loci identified in the GWAS likely are not segregating in the F_1 progeny, and (*ii*) the QTLs found in the linkage mapping analysis have rare alleles in the association panel. For both GWAS results and QTLs mapping, the minor trait-loci associations were not stable across years, suggesting they are more affected by environmental conditions. For this reason, we decided to focus mostly on the GWAS results based on the two-year BLUPs. Nevertheless, results based on the legacy data allowed confirmation of the major SNP associated with budbreak date.

We observed a different scenario for all the stages of male flowering date: while the major QTL found on LG

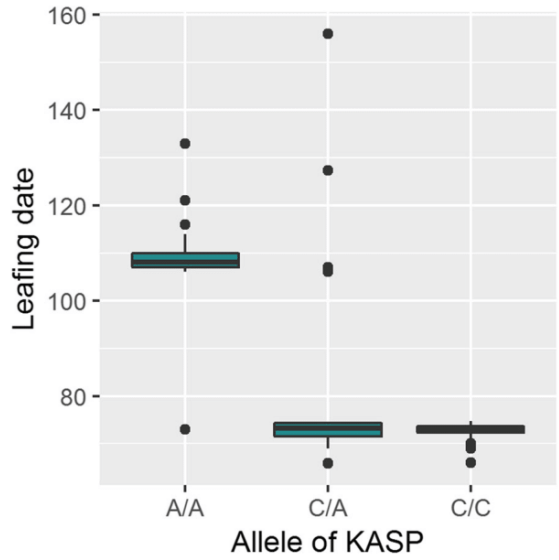


Fig. 7 Allelic effect of the KASP marker associated to the leafing date. 96 unreleased breeding line accessions of the Walnut Improvement Program of the University of California, Davis, were used

1 co-localizes with those for female flowering dates and budbreak date, these results were not confirmed in the GWAS panel. Here, instead, we identified the most significant associations on chromosomes 11 and 4. Then, since heterodichogamy is the measurement of how the male and female flowering dates overlap, it is not surprising to find a marker-trait association on Chr11 with more consistency across years. Since the F_1 progeny used for QTL mapping included only 78 individuals, likely leading to an overestimation of the QTL effects (Beavis effect) [39], we decided to analyze the GWAS results in more depth. In addition, a few regions were not segregating in our F_1 progeny, due to monomorphism in the parents, likely preventing accurate QTL detection. This is the case, for example, of the QTLs detected on LG 2 for both parental maps. By checking the physical position of the associated markers, we observed that the beginning of the Chr2 lacks regions from 0 to 5 Mbp for the 'Franquette' map, and from 0 to 7.7 Mbp for the 'UK 6–2' map.

Combining multi-locus models for GWAS has become widespread [40], and the associations found by multiple methods are usually reliable [41]. Similar work on walnut, also using MLMM and FarmCPU models [35], found three associations within a physical genomic region spanning from 3,187,214 to 4,805,396 bp for the leafing date. This is very close to our association for budbreak date, detected at 6,514,832 bp, and our QTL for the same trait using the 'Franquette' parental map, located at 87,347 to 3,051,320 bp. These results indicate that most likely on Chr1 there is a genomic region controlling budbreak, leafing, and female flowering dates in walnut. Since strong positive correlations exist among these traits, with high allelic effects of the marker-trait associations, this would have an impact for selection. For instance, breeding for budbreak and female flowering (and more broadly for all phenological-related traits) would be a difficult job, and would complicate the creation of shorter-cycle cultivars.

In this work, classical QTL mapping mostly confirmed the major marker-trait associations identified. We found a major locus on Chr1 for budbreak date and female flowering dates, and a major locus on Chr11 for male flowering dates. Separating the flowering stages (beginning, peak and end) allowed "repeat" phenotyping, generating robust data for identification of the major loci required for marker development purposes. Also, the high complexity of phenology-related traits indicates that many loci may be involved in their expression [23], making the detection of minor loci difficult.

Genetic bases for the expression of lateral bearing

Lateral bearing is a crucial trait in walnut breeding since it contributes to increased yield [42]. Fruiting at lateral positions on shoots was originally introduced into California from the cultivar 'Payne', which is in the background of all current varieties released by the Walnut Improvement Program of the University of California, Davis [4]. Using a large F_2 progeny from 'Chandler', a lateral bearing cultivar heterozygous for this locus, the genetic basis of lateral fruitfulness was found to be located in the centromeric region of Chr11 [29]. The same location was obtained using F_1 progeny from 'Chandler' by 'Idaho' cross [43]. Additional associations for lateral bearing were found using GWAS in a region covering 25 Mb on Chr11, and a major QTL was detected in 'Chandler' in proximity to the centromeric region (peak at 17,341,634 bp) of Chr11 [35].

In the present study, we reconfirmed that walnut bearing habit is controlled by a major QTL in the centromeric region of Chr11. In line with previous studies, we observed that an allelic substitution at the most associated locus, 'AX-171191765', causes lateral bud fruiting. We confirm previous results regarding genetic control of lateral bearing by performing GWAS in genetically different plant materials, further supporting the power and high resolution of this method in walnut. In our F_1 progeny, this trait did not segregate and all hybrids are lateral bearing. 'Franquette' is terminal bearing and homozygous G/G for this locus, whereas 'UK 6–2' is lateral bearing and heterozygous G/C for this locus. Maybe because our progeny is too small or because of segregation distorsion, we did not observe any terminal bearing individuals.

Gene annotations may give clues to dissect genetic control of flowering process

Availability of the new chromosome-scale reference genome Chandler v2.0 [36] allowed us to explore the gene space surrounding the most significant marker-trait associations identified for bearing habit and phenology. For instance, we found the *chromosome transmission fidelity (ctf) protein 8 homolog* gene as candidate gene for all female flowering dates. Previous studies in *Arabidopsis thaliana* suggest the involvement of the *ctf* genes family in the cell division processes [44]. In particular, the inactivation of the *ctf7–1* and *ctf7–2* genes resulted in developmental defects, including aberrations in flower morphology and male and female gametophytes. Interestingly, a TPX2 encoding gene required for spindle assembly during cell division process was also found likely to be involved in harvest date determination in walnut [35]. These findings suggest that cell division is a crucial process for correct flower and fruit development in walnut, and suggest the *ctf8 protein homolog* locus is an excellent candidate gene for female flowering.

Trichomes consist of only one cell type, appearing as a long, slender appendage [45]. In *J. regia*, during staminate flower development, the tepals differentiate into anthers and filaments shortly before anthesis, enveloping the stamens [46]. Since these tepals become hirsute, it is not surprising to find a *trichome birefringence-like 13 protein* encoding gene within the LD block surrounding the SNP on Chr4 which is associated with male flowering dates. Also, the SNP on Chr11 associated with heterodichogamy and male flowering dates is in LD with a candidate gene encoding a *probable trehalose-phosphate phosphatase D*. This dephosphorylating enzyme and other members of the gene family are highly expressed in male flowers of *Jatropha*, a perennial woody plant, and in transgenic plants of *A. thaliana* with delayed flowering [47], suggesting the putative involvement of this class of genes in the flower development process. Other authors reviewed the roles of sugars in the control of flowering time, and apart from its involvement in carbohydrate metabolism, trehalose-6-phosphate seems to serve as a signal for inducing flowering transition in plants [48, 49]. In this regard, we found a *probable trehalose-phosphate phosphatase D* encoding gene for heterodichogamy.

Surprisingly, we did not find widely known genes to be involved in flowering process or dormancy in our candidate genes. In *A. thaliana*, sweet cherry or other species, these mechanisms are regulated by *flowering locus*, *early flowering*, and *MADS-box* homologous genes, among others [50–52]. For this reason, we decided to list several homologous genes within the RefSeq walnut gene annotation using three keywords: 'flower', 'dormancy', 'MADS', and we checked the physical positions of these genes. We then compared them with our marker-trait associations, regardless of the LD blocks, and we found interesting results. We found a *agamous-like MADS-box protein AGL11* encoding gene (approx. 2.0 Mbp) and a *MADS-box transcription factor 6-like* encoding gene (8.8 Mbp), both at the beginning of the Chr1. The first one co-localizes with the major QTL found for budbreak date using the F_1 progeny and the second is very near to the major marker-trait association found for female flowering dates (9.3 Mbp). Then, our results showed several marker-trait associations for both female and male flowering dates on Chr7 between 23.0 and 48.0 Mbp. Interestingly, we found the following encoding genes within this window: a *flowering time control protein FCA-like* (33 Mbp), a *flowering-promoting factor 1-like protein 3* (34.4 Mbp), a *dormancy-associated protein homolog 4* (42.9 Mbp), a *protein early flowering 3-like* (44.3 Mbp), a *MADS-box protein SOC1-like* (48.0 Mbp), and a *protein flowering locus D-like* (48.6 Mbp). Finally, we noticed that two genes are near to our marker-trait associations for the dichogamy (31.9 Mbp) and the male flowering dates (from 29.4 to 31.9 Mbp) on Chr11: a *MADS-box transcription factor 27-like* (29.9 Mbp) and a *protein embryonic flower 1* (34.3 Mbp).

A KASP marker related to phenology is released for the first time in walnut

For the first time in walnut, a KASP marker related to phenology is released. By designing specific primers for the SNP AX-171179714, we found that the homozygous accessions A/A are significantly later leafing. Since this marker is dominant, it will greatly help breeders to accurately select individuals with delayed budbreak. However, due to the complex genetic basis of phenology-related traits in perennials [23], additional markers, especially for minor loci, will be needed to improve the selection. Finally, this SNP is located in a gene encoding an uncharacterized protein, and it would be interesting to know more about the functional role of this gene. The marker can predict a significant portion of the phenotype but we still do not know the regulatory networks involved in this complex trait.

Conclusions

Due to the significant influence of environment on phenology-related traits in walnut, unravelling their genetic architecture is of upmost importance for the development of markers that could assist the selection of superior genotypes and, therefore, the release of new walnut cultivars adapted to different climatic conditions. Using GWAS with two different multi-locus models, we identified significant associations for budbreak date, and male and female flowering dates, confirmed by classical QTL mapping. In addition, we provide a list of candidate

genes for these traits, that will be fundamental in future studies of functional genomics and understanding the metabolic pathways underlying phenology in walnut. We also developed and validated the first KASP marker for budbreak date in walnut, which will allow accurate selection of individuals with a delayed budbreak and, therefore, suitable for cultivation in France and other regions where late spring frosts are challenging. In parallel, the genetic bases for the expression of lateral bearing were confirmed. Since the future French walnut breeding program needs cultivars with high quality kernels, efforts are underway phenotype the entire collection regarding chemical content (e.g. fatty acids and tocopherols) and physical characteristics (nut length, filling ratio, and ease of removal). Our F_1 progeny is too small to pursue investigations but this study confirms that the INRAE walnut germplasm repository contains an array of plant material suitable for this type of work. New genome-wide analyses now are being initiated to further increase our knowledge concerning the genetic architecture of the main traits of interest in walnut.

Methods

Plant material

The panel for GWAS consists of 170 unique *J. regia* grafted trees accessions originating worldwide and maintained at the *Prunus* and *Juglans* Genetic Resources Center located in the Fruit Experimental Unit of the INRAE, Toulenne, France (44°34′37.442″N – 0°16′51.48″W), near Bordeaux. This INRAE walnut germplasm collection is publicly available and is a result of important collecting work performed between 1988 and 2000 by Eric Germain (retired and former head of the INRAE walnut breeding program) in 23 countries including the European, American, and Asian continents. Eric Germain initiated an international cooperation during his activity concerning the exchange of walnut plant materials (especially in Europe) and obtained all permissions to bring them. Experimental research on this plant material complies with our INRAE institutional guidelines. The original source of each accession (institute/lab source or collecting source) is given (Table S1). The panel choice was based on previous genetic diversity work using 13 SSR markers and evaluation of phenotypic variability [37].

The intraspecific F_1 mapping progeny results from biparental controlled crosses made from 1997 to 2004 by Eric Germain between two cultivars with contrasting phenology-related traits, for his needs under his breeding program. The female parent was the 'Franquette' cultivar, a French landrace from Isère valley and having a late budbreak date and terminal bearing. The male parent was the accession 'UK 6–2', obtained from the Kiev Botanical Garden in Ukraine and thought to originate from the center of origin of *J. regia* in Central Asia (Uzbekistan, Tajikistan, or Kyrgyzstan). 'UK 6–2' has an early to intermediate budbreak date and lateral bearing. The progeny consists of 78 individuals, also located at the Fruit Experimental Unit of the INRAE in Toulenne, which were previously genotyped using the same SSR markers as those of the GWAS panel [37] to confirm that they were true progeny. This plant material also complies with our INRAE institutional guidelines.

Phenotypic evaluation and data analysis

Phenotypic evaluation for the following 10 traits was conducted in 2 years (2018 and 2019) for both the GWAS panel and the F_1 progeny, at a rate of two to three visits in orchards per week during the months of March, April and May: budbreak date, beginning, peak, end, and duration of female flowering, heterodichogamy, and beginning, peak, end, and duration of male flowering (Table S5). Heterodichogamy, degree of male and female flowering overlap, was computed by subtracting peak female flowering date from peak male flowering date. A negative value (categories '1' and '3', Table S5) means the accession is protandrous since peak male flowering in Julian days occurs before the peak female date. Categories '7' and '9' indicate protogyny.

For the GWAS panel, we also used phenotypical data based on the same scoring scales used for many years previously (mainly between 1989 and 2011) on three sites. These legacy data are available from the INRAE walnut germplasm collection database [21]. Moreover, as bearing habit is not affected by environment conditions, this trait was recorded only in 2019. This trait did not segregate within the F_1 progeny, so only the GWAS panel was evaluated. Lateral bearing was found in recently released cultivars from France ('Ferbel', 'Fertignac') and the USA ('Pedro', 'Chico', 'Cisco', 'Chandler'), and in accessions coming from the Central Botanical Garden of Kiev, Ukraine (Table S1), supposedly originating from the walnut center of domestication, according to the introduction book of the INRAE. Terminal bearing was observed mainly in French landraces such as 'Franquette', 'Grosvert' and 'Verdelet'. Data management and visualisation, and Spearman correlation matrices (more appropriate for discrete variables) were performed using "R" software [53], with the packages "tidyverse" [54] and "corrplot" [55], respectively.

For both the GWAS panel and the F_1 progeny, the means of genotypic effects were obtained for each accession by adjusting for known environmental factors using the BLUPs. When using two-year data, the means were predicted using a mixed linear model considering the year effect (a). When using phenological data from many years and three sites only available for the GWAS panel [21], the means were predicted using a mixed linear

model considering the effects of year, site and combined of year and site (b):

$$P_{ik} = \mu + Y_i + g_k + e_{ik} \quad \text{(a)}$$

$$P_{ijk} = \mu + Y_i + S_j + \left(Y_i \times S_j\right) + g_k + e_{ijk} \quad \text{(b)}$$

where $P_{i(j)k}$ refers to the observed phenotype of the k^{th} accession in the i^{th} year in the j^{th} site; μ is the mean value of the trait; Y_i is the fixed effect of the i^{th} year, S_j is the fixed effect of the j^{th} site, g_k is the random effect of the k genotype; and $e_{i(j)k}$ is the residuals of the model. The BLUPs were performed using "R" package "lme4" [56].

Based on the previous mixed linear models, broad-sense heritability of each trait was estimated using the variance components. When using two-year data, we accounted for the variance component of genotype × year interaction (a). When using phenological data from many years and three sites, we accounted for the variance components of genotype × year and genotype × site interactions (b):

$$H^2 = \sigma^2_G / [(\sigma^2_G + \left(\sigma^2_{GxY}/n_y\right) + \left(\sigma^2_\varepsilon/n_y\right)] \quad \text{(a)}$$

$$H^2 = \sigma^2_G / [(\sigma^2_G + \left(\sigma^2_{GxY}/n_y\right) + \left(\sigma^2_{GxS}/n_s\right) + \left(\sigma^2_\varepsilon/n_y n_s\right)] \quad \text{(b)}$$

where σ^2_G is the genotypic effect variance; σ^2_{GxY} is the genotype × year interaction variance; σ^2_{GxS} is the genotype × site interaction variance; σ^2_ε is the variance of residuals; n_y is the number of years; and n_s is the number of sites. We have always only one replication by genotype. For two-year data, $n_y = 2$ (2018 and 2019). Nevertheless, the legacy data are very unbalanced for the years and sites available considering each genotype. We consider $n_y n_s = 5$ because, in average, we have 5 observations by genotype.

SNP genotyping and quality control

The genomic DNAs of both the GWAS panel (*J. regia* accessions) and F_1 progeny (78 individuals) were extracted from young leaves as described in Bernard et al. [37]. The accessions were genotyped using the Axiom™ *J. regia* 700 K SNP array containing 609,658 SNPs uniformly distributed over the 16 *J. regia* chromosomes [31]. These SNPs were then filtered through several criteria (Table 2). First, the filtering metrics were performed by ThermoFisher considering default thresholds: dish quality control greater than or equal to 0.82, and quality control call rate of 97%. The quality control steps were performed using "PLINK 1.9" software [57]. SNPs with Mendelian errors in the F_1 progeny were removed. After, Poly High Resolution (PHR) and No Minor Homozygotes (NMH) SNPs were filtered using stringent thresholds: SNP call rate (> 90%), minor allele frequency (MAF > 5%), and redundancy in the genome (SNP probes aligning in duplicated regions). Finally, 364,275 robust SNPs (59.8% of the total number of SNPs) were retained for the following genome-wide analysis.

The same steps were performed for the F_1 progeny, except filtering for MAF, gaining a final panel of 478,458 SNPs. In addition, individuals were checked for outlying heterozygosity rate.

SNP linkage map construction and quantitative trait locus analyses

Based on the pseudo-testcross strategy [58], 849 SNPs were retained for the 'Franquette' female parent, and 1088 for the 'UK 6–2' male parent. The two parental genetic linkage maps were constructed using JoinMap® 4.0 software [59] and distorted markers were omitted. A minimum LOD value of 16.0 was chosen for mapping. Kosambi's mapping function was used to convert recombination frequencies into map distances [60], and graphical display of linkage maps was performed using MapChart 2.32 software [61].

QTLs were determined using MultiQTL 2.6 software (http://www.multiqtl.com/; MultiQTL Ltd., Institute of Evolution, Haifa University, Haifa, Israel). Single trait analysis was performed using a Multiple Interval Mapping (MIM) method [62], in which the highest effect QTL is taken as a cofactor to control the genetic background, whereas another QTL is searched in a different position. All LGs for each parent were scanned using the one QTL model, with 1000 runs of permutation tests, to compare H1 hypothesis (one QTL is present in the LG) and H0 hypothesis (no QTL in the LG). The threshold for MIM was 0.05 and computation of Type I error rate for each QTL (α_{chr}) was performed as follows:

$$\alpha_{chr} = 1 - \{1 - [1 - (1 - 0.05)^{1/M}\}^m$$

where M is the total number of markers used for the QTL detection on each parental map, and m is the number of markers in the LG carrying the QTL [63].

Population structure and kinship analyses

For the GWAS panel, the R packages "SNPRelate" and "gdsfmt" [64] were used to perform PCA and relatedness estimations based on a LD pruned set of 29,825 SNPs to avoid capturing too much variance of high LD regions. For the PCA, the ten largest eigenvalues were obtained to check for structure. The KING method of moment

was used to estimate identity-by-descent (IBD) proportions between all pairwise comparisons [65]. Population structure was also investigated using the "fastSTRUCTURE" software [66], and the most likely K was determined using the ΔK method [67]. Then, "CLUMPP" [68] and "distruct" [69] softwares were used for clustering accessions and graphical presentation respectively.

Genome-wide association analysis

GWAS was carried out using the R package "GAPIT" [70], accounting for the familial relationships in the form of a kinship matrix. The 'model selection' function implemented in GAPIT was used to select the best number of PCs by running a mixed linear model (MLM) with a maximum of ten PCs tested. Then, the best number of PCs to include in the GWAS analysis was selected according to the Bayesian Information Criterion (BIC). Using BLUPs previously calculated as phenotypic data and each year separately, two multi-locus models were applied: multi-locus mixed model (MLMM) [71] and Fixed and random model Circulating Probability Unification method (FarmCPU) [72]. Thresholding with FDR [73] classically implemented in "GAPIT" was applied to define the significant associations.

Search of annotations within LD blocks of associated loci

LD levels around the most associated loci were estimated using HaploView v4.2 software [74]. Each physical position of these trait-SNPs associations was investigated to explore the extension of the surrounding LD blocks and determine the genomic regions where to search for candidate genes. The LD blocks were investigated using the method of confidence interval [75], and "solid spine of LD" method, in which the first and last SNPs in a block are in strong LD with all intermediate markers but the intermediate ones are not necessarily in LD with each other. The identified LD blocks were then searched for candidate genes using RefSeq walnut gene annotation (https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Juglans_regia/100/) and mapped onto the new chromosome-scale reference genome v2.0 [36] (https://www.hardwoodgenomics.org/Genomeassembly/2539069).

Kompetitive allele specific PCR marker development for Budbreak date and validation

Kompetitive allele specific PCR (KASP) markers are based on the dual Fluorescence Resonance Energy Transfer (FRET) method, in which the sample DNA is amplified with allele specific primers conjugated to fluorometric dyes HEX and FAM at their 5′ end. Based on GWAS results regarding the budbreak date, the following KASP primers were developed (LGC Genomics, Hoddesdon, UK) to target the most significantly associated SNP with this trait:

Allele A: 5′-AGGACAGCAATAAACTCAATCACACA-3′.

Allele C: 5′-GGACAGCAATAAACTCAATCACACC-3′.

Allele A tail (FAM tail): 5′-GAAGGTGACCAAGTTCATGCT-3′.

Allele C tail (HEX tail): 5′-GAAGGTCGGAGTCAACGGATT-3′.

Common reverse: 5′-AGGTTCTGCCAAGCTACAGGGTATA-3′.

These primers were developed by the BioGEVES laboratory (Beaucouzé, France) on the complementary strand. The KASP reaction and its components are explained at https://www.biosearchtech.com/support/education/kasp-genotyping-reagents/how-does-kasp-work. The KASP reaction was prepared as follows: 1.95 μL of PCR mix PACE (3CR Biosciences Limited, Harlow, UK), 2 μL of genomic DNA (2 ng/μL), and 0.053 μL of the three primers (Integrated DNA Technologies, Leuven, Belgium), for a total volume reaction of 4.003 μL. Amplification was performed in a hydrocycler, starting with 15 min at 94 °C, a touchdown phase of 10 cycles at 94 °C for 20 s and at 61 °C for 60 s with a 0.6 °C decrease in temperature per cycle, followed by 35 cycles of 94 °C for 20 s and 55 °C for 60 s. Once the cycle was complete, Fluostar Omega reader (BMG Labtech, Paris, France) was used to read the fluorescence signal. The KASP method was tested using DNA from a set of 96 unreleased breeding line accessions from the Walnut Improvement Program of the University of California, Davis.

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1186/s12864-020-6616-y.

Additional file 1: Table S1. List of association panel accessions with their origin, breeding level, pedigree when known, and their membership to clusters based on fastSTRUCTURE results for K = 3.

Additional file 2: Table S2. Number of principal components required for bearing habit trait using Bayesian Information Criterion.

Additional file 3: Table S3. 'Franquette' and 'UK 6–2' parental maps for QTL detection.

Additional file 4: Table S4. Genotyping results of SNP AX-171179714 by KASP.

Additional file 5: Table S5. Walnut trait ontology.

Additional file 6: Table S6. Phenotypic raw dataset of the GWAS panel.

Additional file 7: Table S7. Phenotypic raw dataset of the F_1 progeny.

Additional file 8: Figure S1. Scatter plots showing the two-year data related to phenological traits in Julian days for the 170 accessions of the GWAS panel.

Additional file 9: Figure S2. Density plots showing the two-year data related to phenological traits in Julian days for the 170 accessions of the GWAS panel.

Additional file 10: Figure S3. Scatter plots showing the two-year data related to phenological traits in Julian days for the 78 accessions of the F_1 progeny.

Additional file 11: Figure S4. Density plots showing the two-year data related to phenological traits in Julian days for the 78 accessions of the F_1 progeny.

Additional file 12: Figure S5. Detection of the number of clusters using a) Prob(K), and b) deltaK method (Evanno et al., 2005) in the GWAS panel.

Additional file 13: Figure S6. Principal Component Analysis performed on the GWAS panel. The two first principal components show accessions colored according to fastSTRUCTURE results with EEAs for Eastern Europe and Asia, and WEAm for Western Europe and America.

Additional file 14: Figure S7. Scatter plot showing the estimated kinship coefficients by the proportion of zero Identical-By-State (IBS0) in the F_1 progeny.

Additional file 15: Figure S8. Level plot showing the Identical-By-State (IBS) values for the 170 accessions of the GWAS panel.

Additional file 16: Figure S9. Genetic maps and QTLs detected using two-year data. F(X) and U(X) are the linkage groups of 'Franquette' and 'UK 6–2' parental maps respectively. Legend of the QTLs: black for bud-break date, red for beginning female flowering date, deep green for full female flowering date, blue for end female flowering date, yellow for beginning male flowering date, pink for full female flowering date, and light green for end female flowering date. Solid bars indicate the 95% confidence interval of the QTL, and terminal bars indicate the 99.9% confidence interval of the QTL. The marker names were changed with the corresponding chromosome number and its physical position for a better visualization.

Abbreviations

ANRT: Agence nationale de la recherche et de la technologie; BIC: Bayesian information criterion; BLAST: Basic local alignment search tool; BLUPs: Best linear unbiased predictions; Chr: Chromosome; cM: centiMorgan; CTIFL: Centre technique interprofessionnel des fruits et légumes; FAO: Food and agriculture organisation; FarmCPU: Fixed And random model circulating probability unification; FDR: False discovery rate; FRET: Fluorescence resonance energy transfer; GEMMA: Genome-wide efficient mixed model association; GEVES: Groupe d'etude et de contrôle des variétés et des semences; GWAS: Genome-wide association study; IBD: Identity by descent; IBS: Identity by state; INRAE: Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement; KASP: Kompetitive Allele Specific PCR; LD: Linkage disequilibrium; LG: Linkage group; LOD: Logarithm of odds; MAF: Minor allele frequency; MIM: Multiple interval mapping; MLM: Mixed linear model; MLMM: Multi-locus mixed model; NCBI: National center for biotechnology information; NMH: No minor homozygote; PCA: Principal component analysis; PHR: Poly high resolution; QTL: Quantitative trait locus; SNP: Single nucleotide polymorphism; SSR: Simple sequence repeat

Acknowledgments

First, we would like to thank the late Eric Germain, former head of the breeding program at the INRAE of Bordeaux from 1977 to 2007. His remarkable work, then continued by Francis Delort, has given us the opportunity to study a rich set of plant material. The latter helped us to phenotype the bearing habit of the walnut trees. We thank the Fruit Tree Experimental Unit of the INRAE in Toulenne and the *Prunus/Juglans* Genetic Resources Center for the maintenance of the collection and for helping us to collect the samples. We acknowledge the BioGEVES laboratory for DNA extraction and KASP marker development. Then, the CTIFL, holder of the project "INNOV'noyer", in partnership with the INRAE of Bordeaux, want to thank the "Région Nouvelle-Aquitaine", and "Cifre" convention of "ANRT" (Agence Nationale de la Recherche et de la Technologie). It is also important to note that the project is supported by the "Agri Sud-Ouest Innovation" competitiveness cluster. We also want to thank Bénédicte Wenden (INRAE, UMR 1332 BFP, Villenave d'Ornon, France) for her expertise in data visualisation, and Hao Cheng (Department of Animal Sciences, University of California, Davis) for his help with linear model construction. Finally, our special thanks are addressed to the staff of the University of California, Davis, working on walnut for their involvement and support since the beginning of the project, and for welcoming Anthony Bernard.

Authors' contributions

AB performed the phenotypic statistical analyses under the supervision of AM. Linkage mapping and QTL detection were carried out by AB and ED. GWAS methodology was carried out by AB and AM. AM and DN supervised the entire GWAS analyses and brought expertise about quantitative genetics. AD and AB performed the candidate genes analyses. KASP marker development was made possible thanks to PB and CL who have shared DNA and phenotypic data. FL, ED and AB conceived and coordinated the study. AB wrote the manuscript, assisted by AM. All authors critically reviewed the manuscript and they all approved the final version.

Funding

This work has been mainly funded by the "Région Nouvelle-Aquitaine" in the project "INNOV'noyer", coordinated by the CTIFL, and in partnership with the INRAE of Bordeaux. This work has been also funded by the "Cifre" convention number 2016/1558 of ANRT (Agence Nationale de la Recherche et de la Technologie). To finish, Anthony Bernard thanks the "Initiative d'Excellence" (IdEx) program of the University of Bordeaux and the "Dufrenoy-Crédit Agricole d'Ile-de-France Mécénat" grant for contributing to the funding of his travel to the University of California, Davis.

Availability of data and materials

The phenotypic raw datasets generated and analyzed during the current study for the GWAS panel and for the F_1 progeny are available respectively in Table S6, and Table S7.

The SNP genotyping raw datasets in "hapmap" format are freely and openly accessed on the "Portail Data INRA" repository, via the identifier "INRA's Walnut Hapmap" and the following Digital Object Identifier (DOI): https://doi.org/10.15454/XPKII8. The dataset called "GWAS_hapmap.txt" is related to the GWAS panel, and the one called "Progeny_hapmap.txt" is related to the F_1 progeny. The additional file called "List of ID.tab" allows to link the array identifier name of the accessions with the identifier name used in this study.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

ED is a member of the editorial board (Associate Editor) of BMC Genomics.

Author details

[1]INRAE, Univ. Bordeaux, UMR BFP, F-33882 Villenave d'Ornon, France. [2]CTIFL, centre opérationnel de Lanxade, 24130 Prigonrieux, France. [3]Department of Plant Sciences, University of California, Davis, CA 95616, USA.

Received: 29 November 2019 Accepted: 24 February 2020
Published online: 04 March 2020

References

1. Aradhya MK, Potter D, Simon CJ. Cladistic Biogeography of *Juglans* (*Juglandaceae*) Based on Chloroplast DNA Intergenic Spacer Sequences. In: Motley TJ, Zerega N, Cross H, editors. Darwins Harvest New Approaches Orig Evol Conserv Crops: New York: Columbia University Press; 2006. p. 143–70.
2. Woodworth RH. Meiosis of micro-sporogenesis within the *Juglandaceae*. Am J Bot. 1930;17:863–9.
3. McGranahan G, Leslie C. Breeding walnuts (*Juglans regia*). In: Jain SM, Priyadarshan PM, editors. Breeding Plantation Tree Crops: Temperate Species. Springer Science; 2009. p. 249–73.
4. Bernard A, Lheureux F, Dirlewanger E. Walnut: past and future of genetic improvement. Tree Genet Genomes. 2018;14:1.
5. Vahdati K, Arab MM, Sarikhani S, Sadat-Hosseini M, Leslie CA, Brown PJ. Advances in Persian Walnut (*Juglans regia* L.) Breeding Strategies. In: Al-

Khayri J, Jain SM, Johnson DV, editors. Advances in Plant Breeding Strategies: Nut and Beverage Crops: Springer International Publishing Switzerland; 2019. p. 401–72.
6. McGranahan G, Leslie C. Walnuts (*Juglans*). In: Moore JN, Ballington Jr JR, editors. Genetic resources of temperate fruit and nut crops, part 2. Wageningen: International Society of Horticultural Sciences; 1991. p. 907–51.
7. Cook J, Oreskes N, Doran PT, et al. Consensus on consensus: a synthesis of consensus estimates on human-caused global warming. Environ Res Lett. 2016;11:1–8.
8. Cooke JEK, Eriksson ME, Juntilla O. The dynamic nature of bud dormancy in trees: environmental control and molecular mechanisms. Plant Cell Environ. 2012;35:1707–28.
9. Campoy JA, Ruiz D, Egea J. Dormancy in temperate fruit trees in a global warming context: a review. Sci Hort. 2011;130:357–72.
10. Hassankhah A, Vahdati K, Rahemi M, Sarikhani S. Persian walnut phenology: effect of chilling and heat requirements on Budbreak and flowering date. Int J Hort Sci Tech. 2017;4:259–71.
11. Aslamarz AA, Vahdati K, Rahemi M, Hasani D. Estimation of chilling and heat requirements of some Persian walnut cultivars and genotypes. HortScience. 2009;44(3):697–701.
12. Charrier G, Ngao J, Saudreau M, Améglio T. Effects of environmental factors and management practices on microclimate, winter physiology and frost resistance in trees. Front Plant Sci. 2015;6:259.
13. Charrier G, Chuine I, Bonhomme M, Améglio T. Assessing frost damages using dynamic models in walnut trees: exposure rather than vulnerability controls frost risks. Plant Cell Environ. 2018;41:1008–21.
14. Chmielewski FM, Rötzer T. Response of tree phenology to climate change across Europe. Agric For Meteorol. 2001;108:101–12.
15. Menzel A, Sparks TH, Estrella N, et al. European phenological response to climate change matches the warming pattern. Glob Change Biol. 2006;12: 1969–76.
16. Khanduri VP, Sharma CM, Singh SP. The effects of climate change on plant phenology. Environmentalist. 2008;28:143–7.
17. Vitasse Y, François C, Delpierre N, et al. Assessing the effects of climate change on the phenology of European temperate trees. Agric For Meteorol. 2011;151:969–80.
18. Luedeling E, Gassner A. Partial Least Squares Regression for analyzing walnut phenology in California. Agric For Meteorol. 2012:158, 43–9, 52.
19. Črepinšek Z, Solar M, Štampar F, Solar A. Shifts in walnut (*Juglans regia* L.) phenology due to increasing temperatures in Slovenia. J Hortic Sci Biotechnol. 2009;84:59–64.
20. Cosmulescu S, Ionescu MB. Phenological calendar in some walnut genotypes grown in Romania and its correlation with air temperature. Int J Biometeorol. 2018;62:2007–13.
21. Bernard A, Barreneche T, Delmas M, Durand S, Pommier C, et al. The walnut genetic resources of INRA: chronological phenotypic data and ontology. BMC Res Notes. 2019:12–662. https://doi.org/10.1186/s13104-019-4678-1.
22. Vahdati K, Massah Bavani AR, Khosh-Khui M, Fakour P, Sarikhani S. Applying the AOGCM-AR5 models to the assessments of land suitability for walnut cultivation in response to climate change: a case study of Iran. PLoS One. 2019;14(6):e0218725.
23. Dirlewanger E, Quero-García J, Le Dantec L, et al. Comparison of the genetic determinism of two key phenological traits, flowering and maturity dates, in three *Prunus* species: peach, apricot and sweet cherry. Heredity. 2012;109: 280–92.
24. Olukolu BA, Trainin T, Fan S, et al. Genetic linkage mapping for molecular dissection of chilling requirement and budbreak in apricot (*Prunus armeniaca* L.). Genome. 2009;52:819–28.
25. Charrier G, Bonhomme M, Lacointe A, Améglio T. Are budburst dates, dormancy and cold acclimation in walnut trees (*Juglans regia* L.) under mainly genotypic or environmental control? Int J Biometeorol. 2011;55:763–74.
26. Eskandari S, Hassani D, Abdi A. Investigation on genetic diversity of Persian walnut and evaluation of promising genotypes. Acta Hortic. 2005;705:159–66.
27. Hansche PE, Beres V, Forde HI. Estimates of quantitative genetic properties of walnut and their implications for cultivar improvement. J Am Soc Hortic Sci. 1972;97:279–85.
28. Germain E. Inheritance of late leafing and lateral bud fruitfulness in walnut (*Juglans regia* L.). phenotypic correlations among some traits of the trees. Acta Hortic. 1990;284:125–34.
29. Dvorak J, Aradhya M, Leslie C, Luo MC. Discovery of the causative mutation of the lateral bearing phenotype in walnut. California Walnut Board: Walnut Research Reports; 2015. https://ucanr.edu/repository/a/?get=160313.
30. Martínez-García PJ, Crepeau MW, Puiu D, et al. The walnut (*Juglans regia*) genome sequence reveals diversity in genes coding for the biosynthesis of non-structural polyphenols. Plant J. 2016;87:507–32.
31. Marrano A, Martínez-García PJ, Bianco L, Sideli GM, Di Pierro EA, et al. A new genomic tool for walnut (*Juglans regia* L.): development and validation of the high-density axiom™ *J. regia* 700K SNP genotyping array. Plant Biotechnol J. 2019;17:1027–36.
32. Arab MM, Marrano A, Abdollahi-Arpanahi R, et al. Genome-wide patterns of population structure and association mapping of nut-related traits in Persian walnut populations from Iran using the Axiom *J. regia* 700K SNP array. Sci Rep. 2019;9:6376.
33. Famula RA, Richards JH, Famula TR, Neale DB. Association genetics of carbon isotope discrimination in the founding individuals of a breeding population of *Juglans regia* L. Tree Genet Genomes. 2019;15:6.
34. Arab MM, Marrano A, Abdollahi-Arpanahi R, Leslie CA, Cheng H, et al. Combining phenotype, genotype and environment to uncover genetic components underlying water use efficiency in Persian walnut. J Exp Bot. 2020;71:1107–1127.
35. Marrano A, Sideli GM, Leslie CA, Cheng H, Neale DB. Deciphering of the genetic control of phenology, yield and pellicle color in Persian walnut (*Juglans regia* L.). front. Plant Sci. 2019;10:1140.
36. Marrano A, Britton M, Zaini PA, Zimin AV, Workman RE, et al. High-quality chromosome-scale assembly of the walnut (*Juglans regia* L) reference genome. bioRxiv. 809798. https://doi.org/10.1101/809798.
37. Bernard A, Barreneche T, Lheureux F, Dirlewanger E. Analysis of genetic diversity and structure in a worldwide walnut (*Juglans regia* L.) germplasm using SSR markers. PLoS ONE. 2018;13(11):e0208021.
38. Astle W, Balding DJ. Population structure and cryptic relatedness in genetic association studies. Stat Sci. 2009;24:451–71.
39. Paterson AH, Beavis WD. QTL analyses: power, precision, and accuracy. In: Paterson AH, editor. Molecular dissection of complex traits. CRC Press: New York; 1998. p. 145–62.
40. Wang SB, Feng JY, Ren WL, Huang B, Zhou L, et al. Improving power and accuracy of genome-wide association studies via multi-locus mixed linear model methodology. Sci Rep. 2016;6:19444.
41. Zhang YM, Jia Z, Dunwell JM. Editorial: the applications of new multi-locus GWAS methodologies in the genetic dissection of complex traits. Front Plant Sci. 2019;10:100.
42. Amiri R, Vahdati K, Mohsenipoor S, et al. Correlations between some horticultural traits in walnut. HortScience. 2010;45:1690–4.
43. Aradhya MK, Velasco D, Wang JR, Ramasamy R, You FM, et al. A fine-scale genetic linkage map reveals genomic regions associated with economic traits in walnut (*Juglans regia*). Plant Breed. 2019;00:1–12.
44. Bolaños-Villegas P, Yang X, Wang HJ, Juan CT, Chuang MH, et al. (2013). Arabidopsis *CHROMOSOME TRANSMISSION FIDELITY 7 (AtCTF7/ECO1)* is required for DNA repair, mitosis and meiosis. Plant J. 2013;75:927–40.
45. Behnke HD. Plant trichomes – structure and ultrastructure: general terminology, taxonomic applications, and aspects of trichome-bacteria interaction in leaf tips of *Dioscorea*. In: Rodriguez E, Healey PL, Mehta I, editors. Biology and chemistry of plant trichomes. Plenum: New York; 1984. p. 1–21.
46. Lin RZ, Li RQ, Lu AM, Zhu JY, Chen ZD. Comparative flower development of *Juglans regia*, *Cyclocarya paliurus* and *Engelhardia spicata*: homology of floral envelopes in Juglandaceae. Bot J Linn Soc. 2016;181:279–93.
47. Zhao ML, Ni J, Chen MS, Xu ZF. Ectopic expression of *Jatropha curcas TREHALOSE-6-PHOSPHATE PHOSPHATASE J* causes late-flowering and Heterostylous phenotypes in *Arabidopsis* but not in *Jatropha*. Int J Mol Sci. 2019;20:2165.
48. Cho LH, Pasriga R, Yoon J, Jeon JS, An G. Roles of sugars in controlling flowering time. J Plant Biol. 2018;61:121–30.
49. Ponnu J, Wahl V, Schmid M. Trehalose-6-phosphate: connecting plant metabolism and development. Front Plant Sci. 2011;2:70.
50. Horvath D. Common mechanisms regulate flowering and dormancy. Plant Sci. 2009;177:523–31.
51. Michaels SD, Amasino RM. FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering. Plant Cell. 1999;11: 949–56.

52. Castède S, Campoy JA, Le Dantec L, Quero-García J, Barreneche T, Wenden B, et al. Mapping of candidate genes involved in bud dormancy and flowering time in sweet cherry (*Prunus avium*). PLoS One. 2015;10(11): e0143250.
53. R Development Core Team. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2008. Available online at: http://www.R-project.org/.
54. Wickham H. Tidyverse: Easily Install and Load the 'Tidyverse'. R package version 1.2.1. 2017. https://CRAN.R-project.org/package=tidyverse.
55. Wei T, Simko VR. Package "corrplot": Visualization of a Correlation Matrix (Version 0.84). 2017. https://github.com/taiyun/corrplot.
56. Bates D, Maechler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Softw. 2015;67:1–48.
57. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, et al. PLINK: a toolset for whole-genome association and population-based linkage analysis. Am J hum genet. 2008;81 Available online at: http://pngu.mgh.harvard.edu/purcell/plink/.
58. Grattapaglia D, Sederoff R. Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. Genetics. 1994;137:1121–37.
59. Van Ooijen JW. JoinMap® 4, Software for the calculation of genetic linkage maps in experimental populations. Wageningen: Kyazma B.V; 2006.
60. Kosambi D. The estimation of map distances from recombination values. Ann Eugenics. 1944;12:172–5.
61. Voorrips RE. MapChart: software for the graphical presentation of linkage maps and QTLs. J Hered. 2002;93:77–8.
62. Kao CH, Zeng ZB, Teasdale RD. Multiple interval mapping for quantitative trait loci. Genetics. 1999;152:1203–16.
63. Saintagne C, Bodénès C, Barreneche T, Pot D, Plomion C, Kremer A. Distribution of genomic regions differentiating oak species assessed by QTL detection. Heredity. 2004;92:20–30.
64. Zheng X, Levine D, Shen J, Gogarten S, Laurie C, Weir B. A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics. 2012;28:3326–8.
65. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship inference in genome-wide association studies. Bioinformatics. 2010;26:2867–73.
66. Raj A, Stephens M, Pritchard JK. fastSTRUCTURE: Variational inference of population structure in large SNP data sets. Genetics. 2014;197:573–98.
67. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software structure: a simulation study. Mol Ecol. 2005; 14:2611–20.
68. Jakobsson M, Rosenberg NA. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics. 2007;23:1801–6.
69. Rosenberg NA. Distruct: a program for the graphical display of population structure. Mol Ecol Notes. 2004;4:137–8.
70. Lipka AE, Tian F, Wang Q, Peiffer J, Li M, et al. GAPIT: genome association and prediction integrated tool. Bioinformatics. 2012;28:2397–9.
71. Segura V, Vilhjálmsson BJ, Platt A, Korte A, Seren Ü, Long Q, et al. An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. Nat Genet. 2012;44:825–30.
72. Liu X, Huang M, Fan B, Buckler ES, Zhang Z. Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. PLoS Genet. 2016;12:e1005767.
73. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Statist Soc B. 1995;57:289–300.
74. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 2005;21:263–5.
75. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, et al. The structure of haplotype blocks in the human genome. Science. 2002;296:2225–9.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

BMC

Complément à l'étude du déterminisme génétique des caractères liés à la phénologie

L'évaluation de la phénologie réalisée en 2018 et 2019 a permis d'identifier des SNP associés à sa variation puis de définir une liste de gènes candidats. Le même phénotypage avait été envisagé en 2020 et avait pour but de venir consolider et/ou préciser les zones du génome en lien avec les dates de débourrement et de floraisons mâle et femelle. Malheureusement, les mesures de confinement liées à la pandémie de Covid-19 ont empêché cette troisième campagne de phénotypage dans les vergers. Néanmoins, le phénotypage initié avant la période de confinement peut apporter quelques informations complémentaires concernant les accessions les plus précoces de la collection de ressources génétiques.

Le phénotypage de la phénologie est réalisé deux fois par semaine, de fin février jusqu'à fin mai, comme indiqué dans le chapitre 2. Durant l'année 2020, ce phénotypage a pu être réalisé à six reprises, du 24 février au 12 mars. Cela a permis de noter ou prévoir la date de débourrement de 10 accessions parmi les plus précoces. Sur les 170 accessions du panel dédié à la GWAS, 2 avaient déjà atteint le stade Cf de débourrement durant cette période et 8 autres étaient sur le point de l'atteindre (stade Bf-Cf). La figure 23 ci-dessous reprend les données de date de débourrement de ces 10 accessions pour les 3 années 2018, 2019 et 2020. Elles apportent un complément aux nuages de points des données phénologiques présentées dans le chapitre 3, paragraphe 3.

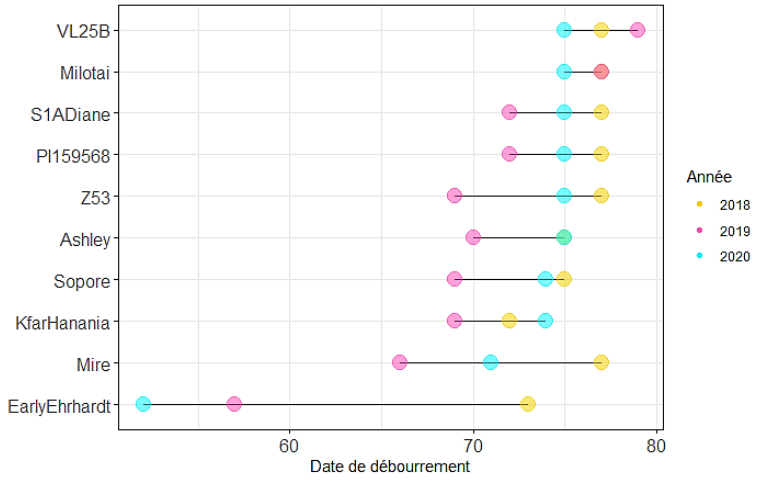


Figure 23 : Nuage de points de la date de débourrement des 10 accessions phénotypées durant les 3 années de thèse.

Au printemps 2019 (points roses), les différentes accessions avaient une date de débourrement majoritairement plus précoce qu'au printemps 2018 (points jaunes). Il est donc intéressant d'observer les données météorologiques des hivers qui correspondent à ces deux campagnes de phénotypage. La figure 24 présente l'évolution de la température moyenne de

Toulenne par rapport à la normale, ou anomalies de températures, selon un diagnostic établi par la moyenne des températures quotidiennes de 1950 à 2017.

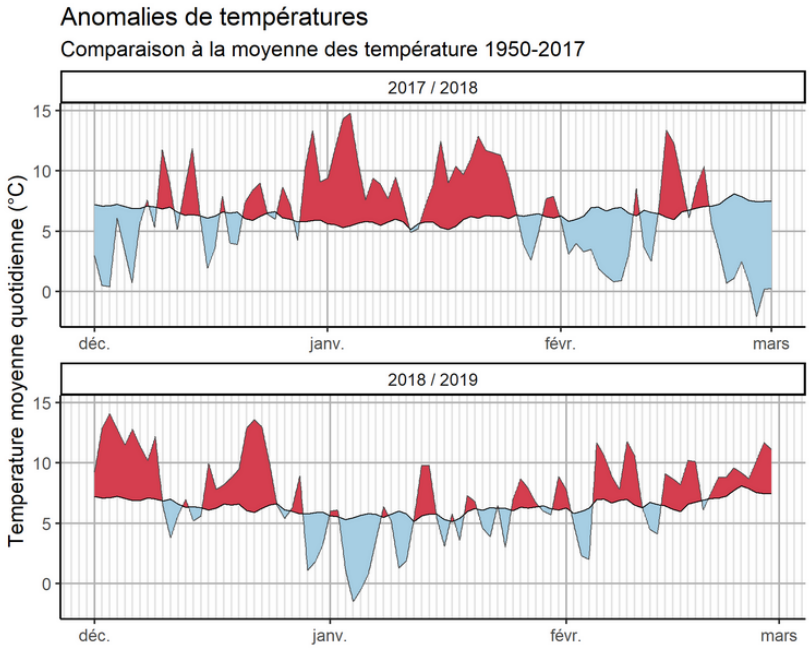


Figure 24 : Données météorologiques des hivers 2017-2018 et 2018-2019.

De par ces graphiques, il est possible de constater que le mois de janvier 2018 a été particulièrement chaud. Néanmoins, février 2018 a subi une chute des températures par rapport à la normale. A l'inverse, janvier 2019 semble se comporter de façon classique tandis que février 2019 a été en moyenne plus chaud. Le mois de février est crucial puisque la période de débourrement débute en général à la fin de ce même mois pour les premières accessions. Les données météorologiques vont donc dans le sens de la phénologie plus précoce constatée en 2019.

L'hiver 2020 quant à lui, a été l'hiver le plus chaud en France depuis le début du XX^{è} siècle. Cependant, les dernières données acquises concernant la date de débourrement des accessions précoces montrent un comportement relativement moyen. La date de débourrement est un caractère dépendant des températures hivernales et une plus grande précocité par rapport à 2019 aurait pu être attendue. Mais il est probable que les accessions précoces n'aient pas eu leurs besoins en froid satisfaits, même s'ils sont plus faibles pour ces accessions, empêchant le débourrement. Ces résultats préliminaires montrent que la phénologie est donc un phénomène complexe, dont la part environnementale et la part génétique sont indéniables.

4. Etude du déterminisme génétique des caractères liés au fruit

La collection de ressources génétiques de l'INRAE de Nouvelle-Aquitaine-Bordeaux a montré qu'elle possédait des accessions adaptées pour réaliser avec succès une GWAS sur les caractères liés à la phénologie. Complémentaires, les caractères liés au fruit sont tout aussi importants pour développer les variétés résilientes de demain. En effet, si le travail sur la phénologie permet de répondre, à un certain dégré, à l'adaptation face au changement climatique, les producteurs·trices devront à l'avenir faire face à une concurrence mondiale toujours grandissante. Le fruit, que l'on consomme, doit évidemment répondre à beaucoup de critères : beau calibre, taux de remplissage par le cerneau élevé, facilité d'énoisage, teneur en acides gras unsaturés élevée, etc. La question suivante s'est donc posée :

Quelle est l'architecture génétique des caractères liés à la qualité du fruit chez le noyer cultivé ?

Par la mise en place du même dispositif (170 accessions de *J. regia* et puce à haute densité de 700K SNP), des associations ont été trouvées pour de nombreux caractères. Un locus majeur a été ainsi détecté sur le chromosome 14 pour la taille globale de la noix, puisqu'il est associé au diamètre de face, au volume de la noix et au volume du cerneau. Un total de 60 associations ont été détectées pour l'ensemble des 25 caractères étudiés et observés durant une à trois années. De nombreux gènes candidats connus pour être impliqués dans la paroi cellaire ont été proposés et les associations déjà identifiées pour la force de cassage de la suture par l'Université de Davis en Californie ont été confirmées. Enfin, pour la première fois, des gènes candidats relatifs au métabolisme des acides gras ont été trouvés. Le phénotypage lié à ce travail a fait appel à des techniques de pointe comme la chromatographie en phase liquide à haute performance et la tomographie à rayons X.

L'article suivant en préparation, porte le nom de « Genome-wide association study reveals candidate genes involved in fruit traits variation in Persian walnut (*Juglans regia* L.) ». Voici donc la version soumise le 16 septembre 2020 dans le journal Frontiers in Plant Science.

Genome-Wide Association Study Reveals Candidate Genes Involved in Fruit Traits Variation in Persian Walnut (*Juglans regia* L.)

Anthony Bernard[1,2], Julie Crabier[1], Armel Donkpegan[1], Annarita Marrano[3], Fabrice Lheureux[2], Elisabeth Dirlewanger[1*]

[1] INRAE, Univ. Bordeaux, UMR BFP, 33882 Villenave d'Ornon, France

[2] CTIFL, centre opérationnel de Lanxade, 24130 Prigonrieux, France

[3] Department of Plant Sciences, University of California, Davis, CA 95616, USA

*Correspondence:

Elisabeth Dirlewanger, e-mail: elisabeth.dirlewanger@inrae.fr

Abstract

Eluciding the genetic determinants of quantitative traits related to weight and ease of cracking in walnut is essential to breed new cultivars meeting the producers' needs. We conducted a Genome-Wide Association Study using multi-locus models in a panel of 170 accessions from the INRAE walnut germplasm collection and genotyped using the Axiom™ *J. regia* 700K SNP array. We phenotyped the panel for 25 fruit traits, for 1 to 3 years depending on the trait, using particularly X-ray computed tomography, texture analyzer, high-performance liquid chromatography and gas chromatography. These traits are related to morphometrics (nut length, nut face/profile diameters, nut surface area), shape (nut Feret shape, nut sphericity, etc.), volume (shell/kernel/empty space volumes, kernel filling ratio), weight (weight of 100 walnuts/100 kernels, etc.), ease of cracking (shell thickness and nut face/suture strength), and molecules of interest contents (poly/monounsaturated fatty acids contents, etc.). We found a highly significant SNP associated with nut face diameter, nut volume and kernel volume on chromosome 14, and 5 significant associations were detected for weight of 100 walnuts. We ultimately found more than 60 marker-trait associations and we proposed several candidate genes involved in fruit characteristics in walnut, including a gene coding for a beta-galactosidase linked to several size-related traits and known to be involved in fruit development in other species. We also confirmed associations on chromosomes 5 and 11 with nut suture strength, recently found by the University of Davis, California. Our study aims to enhance knowledge of the genetic control of important agronomic traits related to the fruit in walnut, and results will pave the way for the development of markers for selection.

Keywords: walnut, *Juglans regia* L., GWAS, fruit traits, X-ray CT, HPLC, gas chromatography, texture analyzer.

Introduction

Persian walnut (*Juglans regia* L.) is one of the oldest food sources known (Aradhya et al., 2006) and is subject to myths and legends since ancient times (Rondeau, 1997). It is a tree species with 2n = 2x = 32 chromosomes (Woodworth, 1930), growing in temperate region (McGranahan and Leslie, 2009). Walnut is the second leading tree crop in France, after apple (more than 22,000 ha), with a worldwide in-shell walnut production mainly provided by China, California and Iran (more than 3,600 kt in 2018, www.fao.org/faostat/). If the ability to adapt to changing climatic conditions is also a breeding priority, larger fruit size, larger filling ratio, ease of cracking are among the main goals of any walnut breeding program (Bernard et al., 2018; Vahdati et al., 2019).

Intelligent management of diversity can be of valuable assistance to breeders and germplasm is, at first, usually evaluated using morphological descriptors (Bernard et al., 2018). Basically, the measurements of nut-related traits (shell thickness, nut length, nut diameter, etc.) in walnut were broadly conducted In Iran and Turkey using a caliper or a micrometer (Eskandari et al., 2005; Khadivi-Khub et al., 2005; Ghasemi et al., 2012; Ahandani et al., 2014; Mahmoodi et al., 2019). Similar studies were released in Europe (Zeneli et al., 2005; Poggetti et al., 2017), but the task is painstaking and of low resolution. Careful consideration of the phenotyping method is of great importance since the heritability of a given trait may depends on the accuracy of the data (Burghardt et al., 2017; Furbank and Tester, 2011). Nowadays, new phenotyping techniques have been able to evolve in line with genotyping technologies. X-ray computed tomography (CT) is an example of imaging technique applied in food sciences to evaluate internal quality (Kotwaliwale et al., 2014) but few works are reported of its use in nut species, except in almond, hazelnut and pecan (Kim and Schatzki, 2001; Khosa and Pasero, 2014; Harrison et al., 1993).

Genetic architecture determination of such traits has been the focus of several recent studies. A new era in walnut is made possible thanks to the high-density Axiom™ *J. regia* 700K SNP genotyping array (Marrano et al., 2019a). By using this genomic tool, association mapping was performed on nut-related traits such as nut length, nut weight, shell thickness, shell texture, kernel percentage to name but a few (Arab et al., 2019). Then, based on this array combined with over 30 years of phenotypic data, GWAS was conducted to decipher the genetic control of yield and kernel pellicle color (Marrano et al., 2019b). Recently, the shell suture strength was also investigated using a texture analyzer and candidate genes were proposed (Sideli et al., 2020). By combining acquisition of accurate phenotypic data using innovative techniques with a unique walnut germplasm collection from INRAE of Bordeaux, we studied the genetic control of 25 traits related to the nut and we sought to propose several candidate genes for those traits. This will be the starting point for walnut breeding programs whose aim is to select new interesting cultivars focused on fruit quality.

Materials and Methods

Plant Material and Phenotypic Data Acquisition

The INRAE walnut germplasm collection is publicly available and is a result of important collecting work performed between 1988 and 2000 by Eric Germain for the need of his breeding activity, in 23 European, American, and Asian countries. The GWAS panel is made of 170 unique *J. regia* grafted trees maintained at the *Prunus* and *Juglans* Genetic Resources Center located in the Fruit Experimental Unit of the INRAE, Toulenne, France (44°34'37.442''N – 0°16'51.48''W), near Bordeaux. The original source of each accession (institute/lab source or collecting source) is given in a previous paper (Bernard et al., 2020).

In-shell walnut sampling (n=100) was performed during harvest seasons in September 2017, 2018, 2019 for each accession and walnuts were dried for 2 days at 25°C using a food dryer. They were stored until analyses in a cold room set to 2°C. Phenotypic evaluation for 25 traits was conducted (S1 Table) and we separated the traits into 6 groups:

(*i*) 4 nut morphometrics related traits: nut length, nut face diameter, nut profile diameter and nut surface area. For 2017 harvest, nut length, nut face diameter and nut profile diameter were measured using an electronic caliper (accuracy 10^{-2} mm) on the sample of 100 walnuts. For 2018 harvest, all traits were measured using an X-ray CT method on a randomly selected subsample of 50 walnuts from the 100 walnuts. X-ray CT scans were performed at the Phenotic imaging platform (Beaucouzé, France) using the 3D X-ray imaging system NSI X-50 model from North Star Imaging©, Inc. (Minnesota, USA). Scans were obtained at constant electron acceleration energy of 120 kV, an electric current of 300 µA and a rotation speed of 4.99 degrees/s resulting in scan duration of 14m34s. Radiographs were used for reconstructing each 3D image using North Star Imaging© reconstruction software EFX-CT (version 1.9.5.12) with a resolution of 992×992×2991 voxels. After 3D reconstruction, a

multi-stage workflow was applied to all CT images consisting in preprocessing steps (denoising), walnuts individualization and traits extraction and quantification using a fully automated in-house image processing pipeline with the Thermo Scientific Avizo© software V9.0.0 built-in functions, the MATLAB© version 7.7.0. R2008b image processing toolbox from The MathWorks©, Inc. (Massachusetts, USA), the TCL scripting language and Spyder Python IDE.

(*ii*) 4 nut shape and texture related traits: nut shape VA3D, nut Feret shape VA3D, nut sphericity and shell rugosity. For 2018 harvest, all traits were measured using the aforementioned X-ray CT method. The nut shape VA3D is defined by $\frac{A^3}{36\pi {V_n}^2}$ and the nut Feret shape 3D is defined by D/d where d is the minimum Feret diameter and D is the maximum Feret diameter in the orthogonal direction (90° from the minimum Feret diameter). The nut sphericity is defined by $\frac{\pi^{\frac{1}{3}}(6V_n)^{\frac{2}{3}}}{A}$ and the shell rugosity is defined by $\frac{A}{\sqrt[3]{36\pi {V_n}^2}}$.

(*iii*) 5 volume related traits: nut volume, shell volume, kernel volume, empty space volume and kernel filling ratio. For 2018 harvest, all traits were measured using the aforementioned X-ray CT method.

(*iv*) 4 weight related traits: weight of 100 walnuts, x3 extreme groups, weight of 50 kernels and breaking yield. All traits were measured for 2017, 2018 and 2019 harvests using an electronic scale (accuracy 10^{-1} mm). A subsample of 50 walnuts from the 100 walnuts was randomly selected to obtain the weight of 50 kernels and the breaking yield. In walnut, the French industry considers 7 sizes depending on the diameter: <28mm, 28-30mm, 30-32mm, 32-34mm, 34-36mm, 36-38mm, >38mm. The trait "x3 extreme groups" is the ratio of the number of walnuts in the 3 extreme groups (34-36mm, 36-38mm and >38mm) to the total groups.

(*v*) 3 cracking related traits: nut face strength, nut suture strength and shell thickness. The subsample of 50 randomly selected walnuts was halved to determine the force needed to

crack the walnut (initial rupture) on the face (25 walnuts) and on the suture (25 walnuts) for 2017, 2018 and 2019 harvests using a texture analyzer (TA-PLUS model, TA1 Texture Analyzer series, from Lloyd Materials Testing™, Ametek®). The following parameters were retained: compression speed of 75mm/min and detection of the initial break when the force drops abruptly by 80%. For 2018 harvest, the shell thickness was obtained using the X-ray CT method.

(*vi*) 5 molecules of interest contents: saturated fatty acids, polyunsaturated fatty acids, monounsaturated fatty acids, tocopherols and vitamin E activity. All traits were measured for 2018 harvest using the following methods: fatty acids extraction by 3/2 (v/v) hexane/isopropanol mixture, miscella washed with 0.8% potassium chloride, vacuum evaporation with a Buchi® rotary evaporator, fatty acid methyl esters analyzed by gas chromatography according to NF EN ISO 12966-2 and 12966-4 standards, and tocopherols/tocotriénols analyzed by high-performance liquid chromatography according to NF EN ISO 9936 standard. Quantifications were performed at ITERG laboratory (Canéjan, France).

Data Analysis, SNP Genotyping, Population Structure and Kinship Analyses

Data management and visualisation were performed using "R" software with the package "tidyverse" (Wickham, 2017). Pearson correlation matrices were performed using the package "corrplot" (Wei and Simko, 2017). The means of genotypic effects were obtained for each accession by adjusting for year effect using the Best Linear Unbiased Predictions (BLUPs), considering the following mixed linear model:

$$P_{ik} = \mu + Y_i + g_k + e_{ik}$$

where P_{ik} refers to the observed phenotype of the k^{th} accession in the i^{th} year; μ is the mean value of the trait; Y_i is the fixed effect of the i^{th} year, g_k is the random effect of the k genotype; and e_{ik} is the residuals of the model. The BLUPs were performed using the package "lme4" (Bates et al., 2015). The broad-sense heritability of the traits phenotyped during at least 2 years (9 among 25 traits) was estimated using the variance components obtained by the previous mixed linear model:

$$H^2 = \sigma^2_G / \sigma^2_P$$

where σ^2_G is the genotypic effect variance and σ^2_P is the phenotypic variance.

The accessions were genotyped using the Axiom™ *J. regia* 700K SNP array containing 609,658 SNPs uniformly distributed over the 16 *J. regia* chromosomes (Marrano et al., 2019a). These SNPs were then filtered through several criteria described in a previous paper (Bernard et al., 2020). Finally, 364,275 robust SNPs were retained for the GWAS.

The population structure was investigated using a sparse non-negative matrix factorization algorithm was implemented using the sNMF 2.0 software, available as a function of the package "LEA" (Frichot and François, 2015). The KING method of moment was used to estimate identity-by-descent (IBD) proportions between all pairwise comparisons (Manichaikul et al., 2010). Both structure and kinship results are available in the previous paper (Bernard et al., 2020).

Genome-Wide Association Analysis, LD Blocks and Search of Annotations

GWAS was carried out using the package "GAPIT" (Lipka et al., 2012). "GAPIT" accounts for relatedness in the form of a kinship matrix and the best number of principal components to include, for accounting for structure, was selected according to the Bayesian Information Criterion (BIC). Two multi-locus models were tested using the BLUPs previously calculated: Multi-Locus Mixed Model (MLMM) (Segura et al., 2012) and Fixed and random model Circulating Probability Unification method (FarmCPU) (Liu et al., 2016). Significant marker-trait associations were determined using both 1% and 5% Bonferroni correction, and previous knowledge based on literature for cracking related traits. The percentage explained variance R^2 was corrected for genome-wide background.

Each physical position of the marker-trait associations was investigated to explore the extension of the surrounding LD blocks and determine the genomic regions where to search for candidate genes, using HaploView v4.2 software (Barrett et al., 2005). The LD blocks were investigated using the method of "solid spine of LD" method, in which the first and last SNPs in a block are in strong LD with all intermediate markers but the intermediate ones are not necessarily in LD with each other. Candidate genes were then searched using RefSeq walnut gene annotation (https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Juglans_regia/100/) and mapped into the new chromosome-scale reference genome v2.0 (Marrano et al., 2020 Preprint) (https://hardwoodgenomics.org/bio_data/2539069?tripal_pane=group_downloads).

Results

Phenotypic Data Visualization

By summarizing the phenotypic data for each trait by year, we observed a large variability within our germplasm collection (Table 1). Considering the range of the nut length in 2017, the values vary by up to 100% across the accessions (from approx. 26 to 53 mm). We even observed variation by up to 400% for the nut volume in 2018 (from approx. 10,400 to 43,000 mm^3) and the weight of 100 walnuts in 2017 (from approx. 522 to 2,278 g), and by up to 700% for the nut suture strength in 2019 (from approx. 101 to 777 N). On the contrary, the variation of the nut sphericity is low (from 0.84 to 0.93) as well as that of proportion of saturated fatty acids (from 8.60 to 11.29 %). The frequency histograms of the 25 traits studied show that they all look normally distributed, except the "x3 extreme groups" trait (S1 Figure), having a bimodal distribution.

Table 1. Descriptive statistics and broad-sense heritability values of the 25 traits.

Trait	Unit [a]	Year	Mean ± SD [b]	Range	H^2
Nut morphometrics related traits					
Nut Length	mm	2017	37.53 ± 4.20	25.99 - 52.69	0.91
		2018	38.39 ± 3.91	28.57 - 51.43	
Nut Face Diameter	mm	2017	31.54 ± 2.99	23.01 - 40.55	0.79
		2018	32.27 ± 2.55	25.99 - 40.75	
Nut Profile Diameter	mm	2017	32.25 ± 3.41	24.10 - 42.87	0.83
		2018	33.29 ± 2.88	27.06 - 42.84	
Nut Surface Area	mm^2	2018	4,019.54 ± 701.70	2,622.59 - 7,093.53	-
Nut shape and texture related traits					
Nut Shape VA3D	-	2018	1.47 ± 0.10	1.24 - 1.69	-
Nut Feret Shape 3D	-	2018	1.25 ± 0.08	1.12 - 1.48	-
Nut Sphericity	-	2018	0.88 ± 0.02	0.84 - 0.93	-
Shell Rugosity	-	2018	1.14 ± 0.03	1.08 - 1.19	-
Volume related traits					
Nut Volume	mm^3	2018	19,400.02 ± 4,889.76	10,382.05 - 42,813.08	-
Shell Volume	mm^3	2018	4,076.78 ± 877.94	2,390.66 - 9,051.88	-
Kernel Volume	mm^3	2018	5,723.89 ± 1,152.14	3,408.85 - 9,548.93	-

Empty Space Volume	mm^3	2018	9,599.35 ± 3,226.42	4,536.51 - 24,212.21	-
Kernel Filling Ratio	%	2018	30.00 ± 3.38	20.70 - 37.40	-
Weight related traits					
Weight of 100 Walnuts	g	2017	1,109.64 ± 262.52	521.74 - 2,278.20	0.86
		2018	1,192.80 ± 255.37	624.40 - 2,251.40	
		2019	1,152.86 ± 263.43	539.98 - 2,288.86	
X3 Extreme Groups	%	2017	30.97 ± 35.88	0.00 - 100.00	0.82
		2018	46.51 ± 35.24	0.00 - 100.00	
		2019	33.65 ± 35.57	0.00 - 100.00	
Weight of 50 Kernels	g	2017	249.27 ± 61.40	113.72 - 416.40	0.77
		2018	276.12 ± 56.51	138.50 - 440.70	
		2019	260.18 ± 56.84	146.73 - 428.03	
Breaking Yield	%	2017	44.86 ± 5.23	30.03 - 65.85	0.80
		2018	45.45 ± 5.11	30.80 - 59.40	
		2019	46.03 ± 5.09	30.47 - 60.54	
Cracking related traits					
Nut Face Strength	N	2017	435.18 ± 106.69	205.72 - 763.34	0.71
		2018	424.24 ± 119.58	182.80 - 893.70	
		2019	409.53 ± 105.89	172.41 - 863.61	
Nut Suture Strength	N	2017	281.06 ± 102.16	87.49 - 614.37	0.72
		2018	250.31 ± 88.59	74.40 - 657.20	
		2019	308.62 ± 104.20	101.35 - 776.97	
Shell Thickness	mm	2018	1.03 ± 0.11	0.73 - 1.49	-
Molecules of interest contents					
Saturated Fatty Acids	%	2018	9.90 ± 0.57	8.60 - 11.29	-
Polyunsaturated Fatty Acids	%	2018	72.36 ± 2.70	63.91 - 78.13	-
Monounsaturated Fatty Acids	%	2018	17.60 ± 2.73	11.77 - 25.33	-
Tocopherols	mg/kg	2018	371.07 ± 110.19	40.00 - 646.00	-
Vitamin E Activity	mg alpha-TE/100g	2018	4.27 ± 1.62	0.20 - 8.30	-

[a] Units abbreviations: mm for millimetre, mm^2 for square millimetre, mm^3 for cubic millimetre, % for percentage, g for gram, N for Newton, alpha-TE for alpha-tocopherol equivalent

[b] SD is the abbreviation for standard deviation

The Pearson correlation matrix (Figure 1) indicates strong positive correlations between all nut morphometrics related traits and volume related traits. For instance, the nut length is positively correlated with the nut face diameter (0.62), the nut volume (0.74) and the kernel volume (0.64). The weight of 100 walnuts is also positively correlated with the weight of 50 kernels (0.81) and also the nut suture strength with the nut face strength (0.73). However, the

nut sphericity and the kernel filling ratio are both traits negatively correlated with nut morphometrics related traits and volume related traits. This means that the bigger a walnut is, the less spheric and the less filled by kernel in proportion it will be. Then, both nut suture and face strengths are negatively correlated with the breaking yield. Finally, the breaking yield, which is the ratio between the weight of 50 kernels to the total weight of 50 nuts, is not correlated with the kernel filling ratio, which is the same ratio but based on volumes, proving an independency between weight and volume in general in walnut.

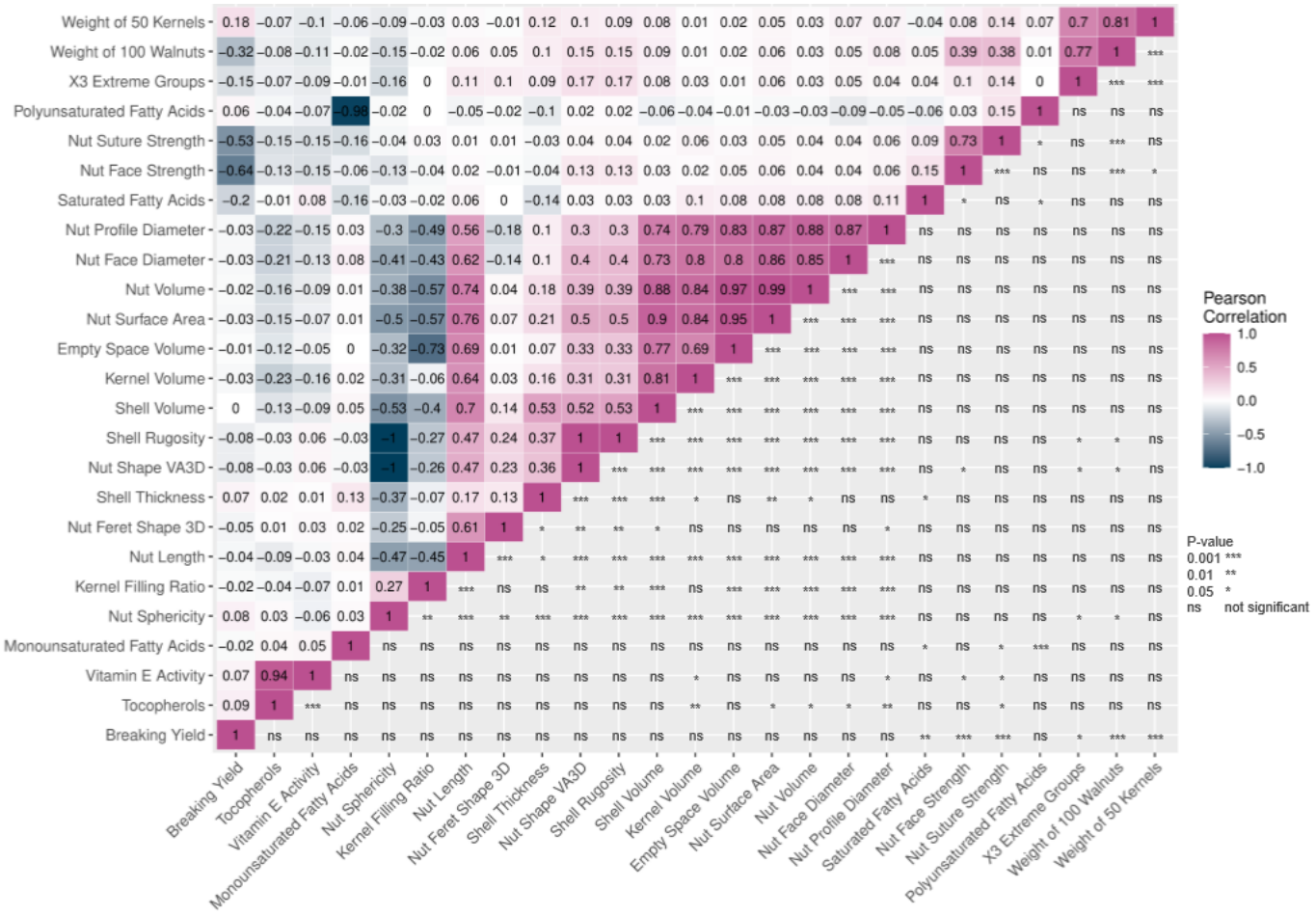


Figure 1. Correlation matrix of the 25 traits.

Pearson correlation coefficient was used and p-value is indicated as follows: 0.001 (***), 0.01 (**), 0.05 (*), not significant (ns).

The broad-sense heritability values estimated for the 9 traits observed during at least 2 years show differences (Table 1). The nut morphometrics related traits have a high H^2 value, from 0.79 to 0.91, as well as those related to the weight, from 0.77 to 0.86. However, the nut face strength and suture strength have a smaller H^2 value, 0.71 and 0.72 respectively, showing that they are both more complex quantitative traits.

BLUPs and Cofounding Effects

The BLUPs were calculated for the 9 traits observed during at least 2 years using a linear mixed model, considering the genotypic mean as a random effect and the year as a fixed effect. The density plots (Figure 2) show a normal distribution for all traits, except for "x3 extreme groups" trait.

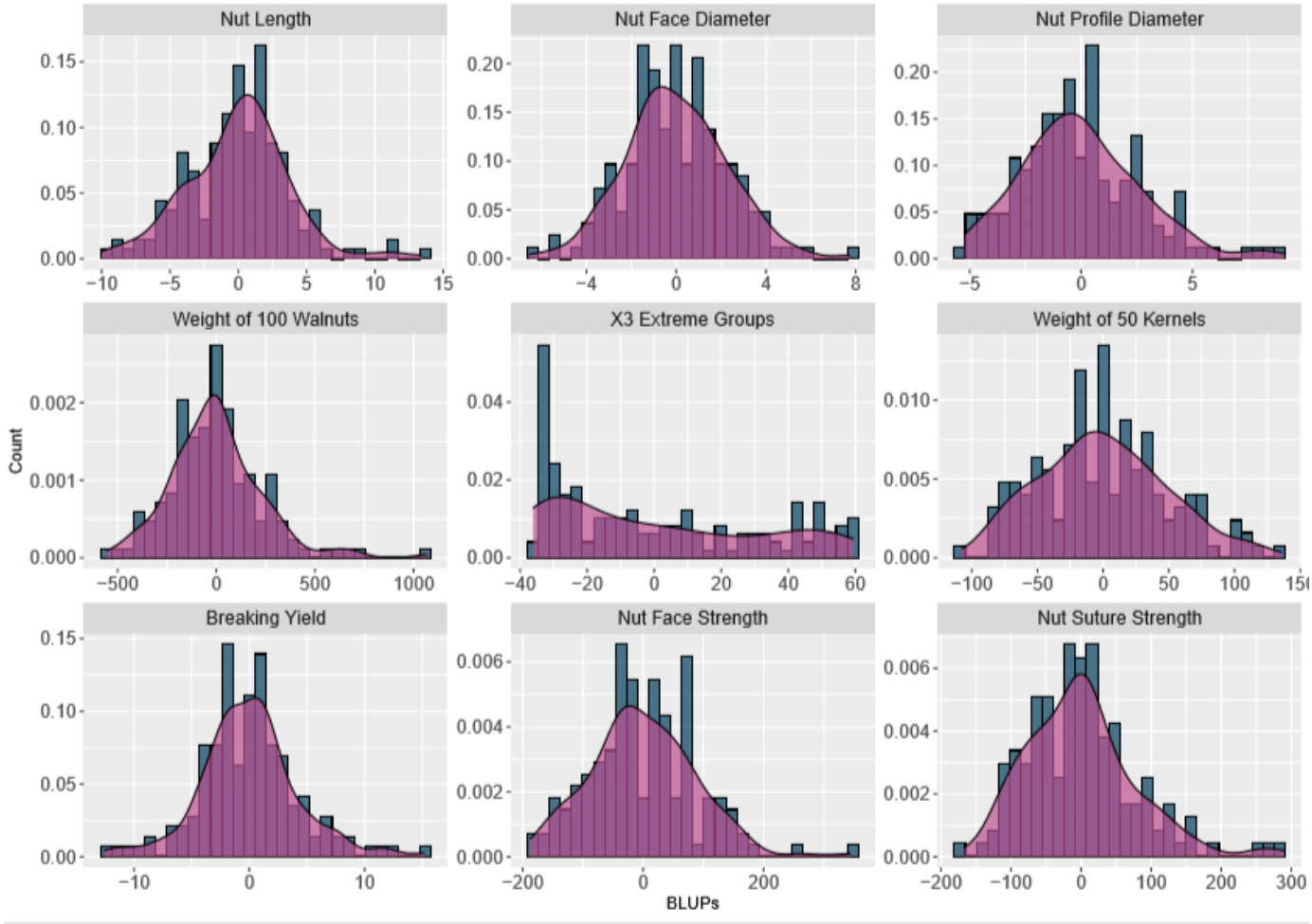


Figure 2. Histograms and density plots of the BLUPs.

The BLUPs were calculated for the 9 traits observed during at least 2 years using a linear mixed model, considering the genotypic mean as a random effect and the year as a fixed effect.

In previous studies, we found a structure in our germplasm linked to the geographical origin of the accessions. We determined the best K as K=2 with a group consisting of the accessions

coming from Western Europe and America and the second with the accessions from Eastern Europe and Asia. GAPIT permits to define the best number of Principal Components (PCs) for accounting for structure using a Bayesian Information Criterion (BIC). Surprisingly, we found that zero is the best number of PCs to include in our model. As a result, we decided to perform a Principal Component Analysis (PCA) using the 25 traits and our previous results of structure to color the accessions. The first two dimensions of the PCA explain 44.6% of the total variance (S2a Figure). The dimension 1 corresponds to the morphometrics, the volume and the weight related traits, whereas the dimension 2 is linked to the shape and texture, and cracking related traits (S2b Figure). The molecules of interest contents are traits badly represented by the PCA. By visualizing the scatterplot of the individuals, we can notice that both groups (Western Europe and America "W", Eastern Europe and Asia "E", and admixed "A") contain accessions with globally large-sized or small-sized walnuts (S2c Figure), since the 95% confidence ellipses intersect and blend together. This means that the structure of our germplasm does not influence the traits, in agreement with the best number of PCs to include of zero.

GWAS Using Multi-Locus Models

Using the automatically implemented more stringent 1% Bonferroni correction, we found 37 marker-trait associations, but we found a total of 58 marker-trait associations using a threshold of 5% (Table 2). We found no SNP associated with 6/25 traits: nut shape VA3D, nut sphericity, nut rugosity, shell volume, saturated and polyunsaturated fatty acids content. Except for these traits, we found 3.2 SNPs associated per trait on average. For instance, we found associations on Chr 5, 8, 11 and 15 with the nut length and on Chr 1, 3, 6, 8 and 12 for the weight of 100 walnuts. Most of the significant marker-trait associations were obtained using FarmCPU model whereas MLMM permitted to highlight two SNPs associated with the tocopherols content, the same associated with the vitamin E activity. However, the Manhattan

plots indicate strong similarities between both MLMM and FarmCPU models for all traits, even if signals obtained with MLMM are not significant (S3 Figure).

Interestingly, we found several marker-trait associations which are very close or even at the same physical position between some of the correlated traits (Table 2). We found the SNP AX-171125096 (Chr 4, 26,041,265 bp) associated with both nut face diameter (R^2 = 19.80) and nut profile diameter (R^2 = 16.50), but both traits have specific associations (for instance, on Chr 1 for the nut face diameter and on Chr 7 for the nut profile diameter). Moreover, the SNP AX-170690867 (Chr 14, 1,248,953 bp) is linked with the nut face diameter, the nut volume and the kernel volume, and very close, we found the SNP AX-171170293, associated with "x3 extreme groups" trait, 119 bp apart (Chr 14, 1,248,839 bp). Similarly, the association on Chr 9 for the nut surface area is also found for the nut volume and the empty space volume.

In our study, we had a different approach for the cracking related traits. Unfortunately, we found no significant association using 5% Bonferroni correction but interesting signals are present on Chr 11 and 2 for the nut face strength, on Chr 5 and 11 for the nut suture strength, and on Chr 2 for the shell thickness, which is the same association as for the nut face strength (Table 2). Based on previous published works, we decided to continue the investigations, since marker-trait associations were already found on Chr 5 and 11 for the suture strength.

LD Blocks and Candidate Genes

Using the "solid spine of LD" method, we found differences in LD level across the loci associated with traits. If we go back to the previous example concerning the two SNPs AX-170690867 (Chr 14, 1,248,953 bp) linked with the nut face diameter, the nut volume and the kernel volume, and SNP AX-171170293 (Chr 14, 1,248,839 bp) linked with "x3 extreme

groups", they belong to a very short LD block of 814 bp (Table 3). On the contrary, the LD block of the SNP AX-171514816 on Chr 7, associated with the nut volume, is longer with 139 342 bp. We also found several loci in complete linkage equilibrium, such as the SNP AX-171001086 (Chr 6, 30,535,497 bp), associated with the nut surface area, which does not belong to any block.

Table 2. Results of GWAS of the 25 traits.

SNP	Chr [a]	Physical position	Significance/ Model [b]	R^2 [c]	Alleles/ Effect [d]	Close position [e]
Nut morphometrics related traits						
Nut Length						
AX-170748449	5	13,134,800	5.02E-10/FarmCPU**	14.45	T,C/2.87	1
AX-170808025	15	4,444,194	3.23E-09/FarmCPU**	15.40	T,G/1.47	
AX-171077810	8	25,778,592	6.88E-09/FarmCPU**	<0.10	A,G/1.23	
AX-171191763	11	20,831,595	3.66E-08/FarmCPU*	14.50	C,T/2.44	
Nut Face Diameter						
AX-171110506	1	44,683,085	5.17E-13/FarmCPU**	5.88	A,G/0.90	
AX-171125096	4	26,041,265	7.98E-09/FarmCPU**	19.80	T,C/-1.41	2
AX-170690867	14	1,248,953	1.11E-08/FarmCPU**	26.36	G,T/0.74	3
AX-170799076	16	19,169,530	1.47E-07/FarmCPU*	3.22	T,G/0.67	
Nut Profile Diameter						
AX-171125096	4	26,041,265	7.10E-08/FarmCPU*	16.50	T,C/-1.48	2
AX-171059229	7	7,823,815	1.33E-07/FarmCPU*	<0.10	T,C/0.72	
Nut Surface Area						
AX-170984383	9	22,804,956	3.58E-08/FarmCPU*	16.51	A,C/250.39	4
AX-171001086	6	30,535,497	1.50E-07/FarmCPU*	16.11	T,G/-327.60	5
Nut shape and texture related traits						
Nut Shape VA3D	No association	-	-	-	-	
Nut Feret Shape 3D						
AX-170668694	13	4,932,826	2.20E-11/FarmCPU**	13.16	C,T/-0.07	
Nut Sphericity	No association	-	-	-	-	
Nut Rugosity	No association	-	-	-	-	
Volume related traits						
Nut Volume						
AX-170690867	14	1,248,953	4.66E-10/FarmCPU**	17.33	G,T/2,101.08	3
AX-171514816	7	50,526,399	1.00E-09/FarmCPU**	16.39	A,T/2,018.71	6
AX-170984383	9	22,804,956	3.95E-09/FarmCPU**	16.38	A,C/1,783.58	4
AX-170810656	7	38,570,409	2.76E-08/FarmCPU*	<0.10	T,C/-1,574.13	
AX-171001086	6	30,535,497	2.76E-08/FarmCPU*	17.09	T,G/-2,228.97	5
AX-170719786	2	31,237,408	2.35E-07/FarmCPU*	<0.10	G,A/-1,760.19	
Shell Volume	No association	-	-	-	-	

Kernel Volume						
AX-170690867	14	1,248,953	2.13E-12/FarmCPU**	24.97	G,T/521.01	3
AX-170640054	5	10,793,840	3.24E-09/FarmCPU**	21.02	G,A/-586.65	
AX-170940799	10	6,432,676	1.41E-08/FarmCPU**	19.54	C,A/-594.42	
Empty Space Volume						
AX-170593629	12	24,730,815	3.45E-16/FarmCPU**	15.03	A,C/-1,717.71	
AX-170984383	9	22,804,956	2.01E-12/FarmCPU**	17.01	A,C/1,462.60	4
AX-171514726	7	50,189,503	4.52E-09/FarmCPU**	17.06	C,G/1,324.30	6
AX-171017658	1	10,234,077	3.27E-08/FarmCPU**	16.07	C,A/-2,014.48	
Kernel Filling Ratio						
AX-171048523	10	19,688,911	2.75E-12/FarmCPU**	14.98	C,T/0.02	
AX-171583500	3	3,603,319	2.12E-11/FarmCPU**	20.44	T,A/0.01	
AX-171540343	13	6,539,102	2.16E-09/FarmCPU**	0.13	A,G/-0.01	
AX-171117250	9	12,230,952	8.33E-08/FarmCPU*	7.17	T,C/0.01	
Weight related traits						
Weight of 100 Walnuts						
AX-171175345	8	26,618,252	1.72E-08/FarmCPU**	0.59	C,T/-176.41	
AX-170573680	3	7,355,283	1.20E-07/FarmCPU*	2.10	G,T/-64.13	
AX-171083810	6	31,463,066	1.54E-07/FarmCPU*	4.23	T,G/-104.12	
AX-170973887	12	3,304,218	1.62E-07/FarmCPU*	17.13	A,G/67.16	
AX-170605550	1	15,926,539	2.02E-07/FarmCPU*	3.47	A,G/50.24	
X3 Extreme Groups						
AX-171170293	14	1,248,839	2.68E-08/FarmCPU**	21.7	A,G/-9.83	3
AX-170834489	1	19,207,870	5.05E-08/FarmCPU*	7.02	T,C/8.82	
AX-170723157	10	4,802,938	7.00E-08/FarmCPU*	5.01	T,C/-12.56	
Weight of 50 Kernels						
AX-171207844	1	39,963,556	5.17E-09/FarmCPU**	11.50	G,T/15.27	
AX-171547969	7	3,283,684	8.25E-09/FarmCPU**	4.66	T,A/-11.89	7
AX-170806411	4	23,487,069	4.85E-08/FarmCPU*	10.54	C,A/28.31	
Breaking Yield						
AX-171005810	14	26,821,528	6.80E-10/FarmCPU**	0.72	C,A/2.55	8
AX-170746651	16	17,458,649	2.11E-08/FarmCPU**	27.23	A,G/-2,09	
Cracking related traits						
Nut Face Strength						
AX-170722428	11	13,511,848	9.10E-06/FarmCPU	16.13	G,A/-53.53	9
AX-170865366	2	32,096,896	2.49E-06/MLMM	11.24	G,A/NA	10
Nut Suture Strength						
AX-170748526	5	13,024,522	2.82E-07/MLMM	12.26	C,A/NA	1
AX-170908256	11	12,909,083	6,72E-05/FarmCPU	14.75	G,A/-49.90	9
Shell Thickness						
AX-170865411	2	32,128,087	1.46E-07/MLMM	13.69	A,C/NA	10
Molecules of interest contents						
Saturated Fatty Acids	No association	-	-	-	-	
Polyunsaturated Fatty Acids	No association	-	-	-	-	
Monounsaturated Fatty Acids						
AX-170701328	14	26,730,713	2.73E-12/FarmCPU**	17.73	C,T/-1.43	8
AX-171494826	9	21,100,618	3.81E-10/FarmCPU**	13.58	A,T/-2.19	
AX-170973656	12	3,155,205	1.42E-08/FarmCPU**	13.35	T,C/0.90	

AX-170996807	14	21,159,561	5.63E-08/FarmCPU*	<0.10	A,G/0.86	
AX-171536319	7	47,712,322	7.00E-08/FarmCPU*	14.27	G,C/-1.72	
Tocopherols						
AX-171134134	10	34,010,861	7.21E-13/MLMM**	14.03	C,T/NA	11
AX-170893393	9	15,280,292	6.22E-09/MLMM**	7.13	C,T/NA	12
AX-170775823	13	23,062,305	5.98E-15/FarmCPU**	12.60	G,T/95.00	
AX-171016396	17	-	8.10E-11/FarmCPU**	4.55	C,A/-40.48	
AX-170615496	8	29,149,403	1.02E-09/FarmCPU**	18.16	A,G/81.85	
AX-171592220	5	16,978,410	9.09E-09/FarmCPU**	<0.10	C,G/-41.90	
AX-170909834	7	3,266,820	3.89E-08/FarmCPU*	11.70	T,C/73.88	7
AX-171595297	14	7,508,539	4.40E-08/FarmCPU*	11.80	A,T/65.54	
Vitamin E Activity						
AX-171134134	10	34,010,861	7.93E-13/MLMM**	13.78	C,T/NA	11
AX-170893393	9	15,280,292	2.16E-08/MLMM**	5.32	C,T/NA	12

[a] Chr, abbreviation for Chromosome

[b] The significance value indicated is the unadjusted p-value

The double asterisk (**) indicates that the association is significant according to Bonferroni correction at 0.01 using the mentioned model

The simple asterisk (*) indicates that the association is significant according to Bonferroni correction at 0.05 using the mentioned model

No asterisk indicates that the association is not significant but of interest regarding Manhattan plot of the mentioned model and previous knowledge

[c] R^2 is the percentage explained variance corrected for genome-wide background

[d] The allelic effect is the difference in mean of measured trait between genotypes with one or other allele

The sign (+/-) is with respect to the major allele that is second-mentioned

[e] The same number indicates that marker-trait associations are very close or at the same physical position

We listed a total of 99 different candidate genes within all the LD blocks (Table 3). Among the genes, 10 characterized genes overlap the physical position of the associated SNP. We focused first on these genes. Regarding the nut morphometrics, shape, volume and weight related traits, a gene coding for a *beta-galactosidase* was found to be associated with the nut face diameter, the nut volume, the kernel volume and "x3 extreme groups". We found a *receptor-like cytosolic serine/threonine-protein kinase RBK1* encoding gene associated with the nut length and *L-Ala-D/L-amino acid epimerase-like* encoding gene associated with the nut surface area and the nut volume. We also found a *probable E3 ubiquitin-protein ligase ARI1* encoding gene linked with the nut Feret shape 3D and a *protein TIFY 4B-like* encoding gene linked with the empty space volume. Then, a *BEL1-like homeodomain protein 4* encoding gene was found to be associated with the weight of 50 kernels. In addition, for the

cracking related traits, we determined a gene coding for a *protein TPX2-like* as involved in the nut face strength. Finally, we found candidate genes linked with several molecules of interest, such as a *calmodulin-7* encoding gene for the monounsaturated fatty acids content, and both *synaptotagmin-5-like* and *E3 ubiquitin-protein ligase HOS1-like* encoding genes for the tocopherols content.

Table 3. List of candidate genes found for each marker-trait associations.

Trait	SNP associated	Chr [a]	Physical position [b]	LD block interval [b,c]	Gene ID	Gene interval [b]	Functional annotation	Identical
Nut morphometrics related traits								
Nut Length	AX-170748449	5	13,134,800	13,128,354 - 13,151,244	108986890	13,104,794 - 13,128,642	uncharacterized LOC108986890	
Nut Length	AX-170808025	15	4,444,194	4,439,011 - 4,444,527	108990373	4,439,082 - 4,440,789	uncharacterized LOC108990373	
Nut Length	AX-171077810	8	25,778,592	25,778,592 - 25,778,723	108979693	25,774,934 - 25,778,731	**receptor-like cytosolic serine/threonine-protein kinase RBK1**	
Nut Length	AX-171191763	11	20,831,595	20,831,267 - 20,865,013	109017990	20,831,874 - 20,834,977	uncharacterized LOC109017990	
Nut Face Diameter	AX-171110506	1	44,683,085	44,682,326 - 44,689,338	108982082	44,678,366 - 44,683,047	transcriptional regulator STERILE APETALA	
Nut Face Diameter	AX-171125096	4	26,041,265	26,031,692 - 26,041,265	No gene	-	-	
Nut Face Diameter	AX-170690867	14	1,248,953	1,248,139 - 1,248,953	109012316	1,248,505 - 1,262,472	**beta-galactosidase**	1
Nut Face Diameter	AX-170799076	16	19,169,530	19,167,864 - 19,170,797	No gene	-	-	
Nut Profile Diameter	AX-171125096	4	26,041,265	26,031,692 - 26,041,265	No gene	-	-	
Nut Profile Diameter	AX-171059229	7	7,823,815	7,819,426 - 7,829,301	No gene	-	-	
Nut Surface Area	AX-170984383	9	22,804,956	22,785,041 - 22,811,656	108992222	22,786,344 - 22,787,141	calcium-binding protein PBP1-like	
					108993282	22,793,390 - 22,797,398	G-type lectin S-receptor-like serine/threonine-protein kinase At4g27290	2
					108993306	22,807,408 - 22,811,134	NAD(P)H dehydrogenase (quinone) FQR1-like	
Nut Surface Area	AX-171001086	6	30,535,497	No block	108998609	30,535,124 - 30,535,841	**L-Ala-D/L-amino acid epimerase-like**	3
Nut shape and texture related traits								
Nut Feret Shape 3D	AX-170668694	13	4,932,826	4,927,503 - 4,935,968	109010710	4,924,947 - 4,928,453	uncharacterized LOC109010710	
					109010712	4,929,324 - 4,935,481	**probable E3 ubiquitin-protein ligase ARI1 isoform X1**	
Volume related traits								
Nut Volume	AX-170690867	14	1,248,953	1,248,139 - 1,248,953	109012316	1,248,505 - 1,262,472	**beta-galactosidase**	1
Nut Volume	AX-171514816	7	50,526,399	50,423,525 - 50,562,867	108986372	50,425,049 - 50,428,790	protein kinase APK1A, chloroplastic-like	
					108986373	50,429,292 - 50,433,281	probable dual-specificity RNA methyltransferase RlmN	
					108986346	50,434,186 - 50,439,306	serine/arginine-rich SC35-like splicing factor SCL30	
					108986347	50,439,787 - 50,442,481	elicitor-responsive protein 1-like	

					108986345	50,443,421 - 50,446,168	uncharacterized LOC108986345	
					108986342	50,446,119 - 50,471,394	AP3-complex subunit beta-A	
					108986340	50,473,399 - 50,474,663	uncharacterized LOC108986340	
					108986361	50,487,892 - 50,498,282	protein INVOLVED IN DE NOVO 2-like	
					108986296	50,500,739 - 50,502,952	expansin-A4-like	
					108986359	50,522,027 - 50,526,294	telomere repeat-binding protein 4	
					108986374	50,539,589 - 50,543,305	scarecrow-like protein 14	
					108986306	50,544,056 - 50,546,156	ribosomal RNA large subunit methyltransferase E	
					108986354	50,547,444 - 50,550,514	protein FIZZY-RELATED 3	
					108986353	50,551,091 - 50,557,917	actin-related protein 5-like	
Nut Volume	AX-170984383	9	22,804,956	22,785,041 - 22,811,656	108992222	22,786,344 - 22,787,141	calcium-binding protein PBP1-like	
					108993282	22,793,390 - 22,797,398	G-type lectin S-receptor-like serine/threonine-protein kinase At4g27290	2
					108993306	22,807,408 - 22,811,134	NAD(P)H dehydrogenase (quinone) FQR1-like	
Nut Volume	AX-170810656	7	38,570,409	38,568,209 - 38,618,466	108990569	38,582,706 - 38,586,602	oxidation resistance protein 1-like	
					108990557	38,588,002 - 38,588,068	uncharacterized LOC108990557	
					109019525	38,588,069 - 38,589,777	pre-mRNA-processing factor 19-like	
					108990558	38,590,516 - 38,591,582	uncharacterized LOC108990558	
					109019525	38,592,948 - 38,601,927	pre-mRNA-processing factor 19-like	
					109019524	38,617,727 - 38,621,795	uncharacterized LOC109019524	
Nut Volume	AX-171001086	6	30,535,497	No block	108998609	30,535,124 - 30,535,841	**L-Ala-D/L-amino acid epimerase-like**	3
Nut Volume	AX-170719786	2	31,237,408	31,218,056 - 31,252,334	109014448	31,222,934 - 31,225,930	uncharacterized LOC109014448	
					109014447	31,227,162 - 31,229,778	probable LRR receptor-like serine/threonine-protein kinase At1g63430	
					109004211	31,230,900 - 31,231,056	probable LRR receptor-like serine/threonine-protein kinase At1g63430	
					109014447	31,227,162 - 31,229,778	probable LRR receptor-like serine/threonine-protein kinase At1g63430	
					109014446	31,244,186 - 31,248,863	O-glucosyltransferase rumi homolog	
					109014525	31,250,638 - 31,251,743	uncharacterized LOC109014525	
Kernel Volume	AX-170690867	14	1,248,953	1,248,139 - 1,248,953	109012316	1,248,505 - 1,262,472	**beta-galactosidase**	1

Kernel Volume	AX-170640054	5	10,793,840	10,792,541 - 10,794,446	No gene	-	-	
Kernel Volume	AX-170940799	10	6,432,676	6,429,804 - 6,441,312	108996735	6,430,658 - 6,432,500	tubulin beta-1 chain-like	
					108996755	6,433,737 - 6,434,513	uncharacterized LOC108996755	
Empty Space Volume	AX-170593629	12	24,730,815	24,688,102 - 24,738,026	109005084	24,689,734 - 24,690,724	early light-induced protein 1, chloroplastic-like	
					109005081	24,691,824 - 24,699,440	dymeclin-like	
					109005079	24,705,107 - 24,708,140	heat stress transcription factor A-6b-like isoform X1	
					109005080	24,709,034 - 24,712,463	1,2-dihydroxy-3-keto-5-methylthiopentene dioxygenase 2-like	
					109005077	24,723,407 - 24,726,043	1,2-dihydroxy-3-keto-5-methylthiopentene dioxygenase 1-like	
					109005076	24,728,025 - 24,733,214	**protein TIFY 4B-like**	
Empty Space Volume	AX-170984383	9	22,804,956	22,785,041 - 22,811,656	108992222	22,786,344 - 22,787,141	calcium-binding protein PBP1-like	
					108993282	22,793,390 - 22,797,398	G-type lectin S-receptor-like serine/threonine-protein kinase At4g27290	2
					108993306	22,807,408 - 22,811,134	NAD(P)H dehydrogenase (quinone) FQR1-like	
Empty Space Volume	AX-171514726	7	50,189,503	50,141,099 - 50,199,957	108986377	50,141,347 - 50,143,302	actin-depolymerizing factor 2-like	
					108986297	50,144,101 - 50,146,776	pentatricopeptide repeat-containing protein At2g39620	
					108986366	50,147,829 - 50,153,101	elongator complex protein 1	
					108986348	50,154,980 - 50,167,794	uncharacterized LOC108986348	
					109015291	50,169,575 - 50,170,416	THO complex subunit 4B-like	
					108986386	50,170,703 - 50,172,827	THO complex subunit 4A-like	
					108986360	50,173,237 - 50,195,465	**uncharacterized LOC108986360**	
Empty Space Volume	AX-171017658	1	10,234,077	10,227,829 - 10,244,488	108991088	10,243,722 - 10,245,723	E3 ubiquitin-protein ligase At4g11680-like	
Kernel Filling Ratio	AX-171048523	10	19,688,911	19,686,439 - 19,688,911	No gene	-	-	
Kernel Filling Ratio	AX-171583500	3	3,603,319	3,596,654 - 3,609,718	108980538	3,597,955 - 3,599,758	uncharacterized LOC108980538	
					109014312	3,607,107 - 3,609,769	transcription initiation factor IIB-2	
Kernel Filling Ratio	AX-171540343	13	6,539,102	6,523,102 - 6,557,694	108993396	6,523,328 - 6,523,663	probable aquaporin PIP2-8	
					108993404	6,525,302 - 6,528,394	pre-mRNA-splicing factor SYF2	
					108999651	6,532,256 - 6,532,777	uncharacterized LOC108999651	
					108993333	6,532,887 - 6,533,184	uncharacterized LOC108993333	
					108999651	6,535,004 - 6,535,132	uncharacterized LOC108999651	

					108993333	6,535,540 - 6,536,367	uncharacterized LOC108993333	
					108993332	6,540,321 - 6,543,295	transcription factor ABORTED MICROSPORES-like	
					108993314	6,543,708 - 6,546,854	chemocyanin-like	
					108993356	6,549,645 - 6,550,413	uncharacterized LOC108993356	
					108993378	6,552,037 - 6,558,488	probable ubiquitin-conjugating enzyme E2 23	
Kernel Filling Ratio	AX-171117250	9	12,230,952	12,208,938 - 12,255,984	108982558	12,251,626 - 12,255,901	serine/threonine-protein kinase ATM	
Weight related traits								
Weight of 100 Walnuts	AX-171175345	8	26,618,252	26,611,559 - 26,618,252	108985800	26,611,675 - 26,613,601	phosphoenolpyruvate carboxykinase [ATP]-like	
					108985801	26,615,053 - 26,618,277	**uncharacterized LOC108985801**	
Weight of 100 Walnuts	AX-170573680	3	7,355,283	7,338,541 - 7,356,913	109003589	7,339,267 - 7,342,354	14 kDa zinc-binding protein	
					109003588	7,343,193 - 7,344,784	sulfiredoxin, chloroplastic/mitochondrial	
					109003587	7,345,910 - 7,351,460	phosphoglycerate kinase, cytosolic	
					109003586	7,354,174 - 7,354,840	chaperonin 60 subunit alpha 2, chloroplastic	
Weight of 100 Walnuts	AX-171083810	6	31,463,066	31,364,479 - 31,464,703	108985243	31,384,084 - 31,385,250	myb-related protein 308-like	
					108985242	31,425,116 - 31,426,216	transcription repressor MYB6-like	
Weight of 100 Walnuts	AX-170973887	12	3,304,218	3,296,379 - 3,352,059	108998258	3,304,683 - 3,341,673	peroxisome biogenesis protein 1	
					108998259	3,344,378 - 3,345,799	FHA domain-containing protein FHA2	
Weight of 100 Walnuts	AX-170605550	1	15,926,539	15,898,312 - 15,966,124	109011118	15,898,355 - 15,906,694	ER lumen protein-retaining receptor	
					109006016	15,940,330 - 15,943,514	protein XRI1	
					109011150	15,943,581 - 15,943,832	protein XRI1-like	
					109011117	15,947,604 - 15,951,327	translation initiation factor IF-2-like	
					109006017	15,953,917 - 15,961,294	purple acid phosphatase 15-like	
					109006018	15,965,164 - 15,966,365	uncharacterized LOC109006018	
X3 Extreme Groups	AX-171170293	14	1,248,839	1,248,139 - 1,248,953	109012316	1,248,505 - 1,262,472	**beta-galactosidase**	1
X3 Extreme Groups	AX-170834489	1	19,207,870	19,202,146 - 19,218,695	108991966	19,217,134 - 19,217,736	uncharacterized LOC108991966	
X3 Extreme Groups	AX-170723157	10	4,802,938	4,802,821 - 4,802,938	No gene	-	-	
Weight of 50 Kernels	AX-171207844	1	39,963,556	39,934,883 - 39,963,556	109003196	39,944,022 - 39,944,119	endochitinase 2-like	
					109003484	39,947,326 - 39,948,386	uncharacterized mitochondrial protein AtMg00810-like	
					109003196	39,948,866 - 39,949,391	endochitinase 2-like	
					109003737	39,952,078 - 39,957,851	uncharacterized LOC109003737	

					109006297	39,960,250 - 39,962,011	uncharacterized LOC109006297	
Weight of 50 Kernels	AX-171547969	7	3,283,684	3,279,632 - 3,290,086	108995343	3,280,546 - 3,286,458	**BEL1-like homeodomain protein 4**	
Weight of 50 Kernels	AX-170806411	4	23,487,069	23,484,729 - 23,487,069	No gene	-	-	
Breaking Yield	AX-171005810	14	26,821,528	26,821,528 - 26,851,219	No gene	-	-	
Breaking Yield	AX-170746651	16	17,458,649	No block	No gene	-	-	
Cracking related traits								
Nut Face Strength	AX-170722428	11	13,511,848	13,503,419 - 13,528,275	109000368	13,524,920 - 13,527,615	putative GPI-anchor transamidase	
Nut Face Strength	AX-170865366	2	32,096,896	32,095,993 - 32,104,213	109010833	32,096,186 - 32,100,753	**protein TPX2-like, transcript variants X1 to X3**	
Nut Suture Strength	AX-170748526	5	13,024,522	13,017,020 - 13,109,943	108986888	13,045,730 - 13,046,379	lamin-like protein	
					108986895	13,046,658 - 13,056,810	protein argonaute 4A-like	
					108986890	13,104,784 - 13,106,382	uncharacterized LOC108986890	
Nut Suture Strength	AX-170908256	11	12,909,083	12,884,235 - 12,916,294	No gene	-	-	
Shell Thickness	AX-170865411	2	32,128,087	32,126,341 - 32,131,231	109010875	32,127,411 - 32,130,937	**uncharacterized LOC109010875**	
Molecules of interest contents								
MUFA	AX-170701328	14	26,730,713	26,728,860 - 26,731,202	No gene	-	-	
MUFA	AX-171494826	9	21,100,618	No block	No gene	-	-	
MUFA	AX-170973656	12	3,155,205	3,139,539 - 3,155,205	108998246	3,141,707 - 3,148,716	protein DETOXIFICATION 27-like	
					108998216	3,151,562 - 3,153,143	putative clathrin assembly protein At4g40080	
MUFA	AX-170996807	14	21,159,561	No block	No gene	-	-	
MUFA	AX-171536319	7	47,712,322	No block	108992563	47,712,191 - 47,712,352	**calmodulin-7**	
Tocopherols	AX-171134134	10	34,010,861	34,010,496 - 34,015,586	108983891	34,010,620 - 34,011,067	**synaptotagmin-5-like**	4
					108983893	34,013,577 - 34,014,069	receptor-like protein 12	
Tocopherols	AX-170893393	9	15,280,292	15,271,076 - 15,281,282	108994350	15,271,151 - 15,280,670	**E3 ubiquitin-protein ligase HOS1-like**	5
					108994314	15,280,768 - 15,281,289	2-oxoglutarate-Fe(II) type oxidoreductase-like	
Tocopherols	AX-170775823	13	23,062,305	23,062,229 - 23,062,659	No gene	-	-	
Tocopherols	AX-170615496	8	29,149,403	29,145,942 - 29,153,435	109006750	29,146,407 - 29,147,252	uncharacterized mitochondrial protein AtMg00810-like	
					109006761	29,150,909 - 29,152,572	uncharacterized LOC109006761	
Tocopherols	AX-171592220	5	16,978,410	16,976,697 - 16,991,665	108983737	16,978,443 - 16,980,743	glucan endo-1,3-beta-glucosidase 14-like	
Tocopherols	AX-170909834	7	3,266,820	3,265,033 - 3,267,053	No gene	-	-	
Tocopherols	AX-171595297	14	7,508,539	7,508,539 - 7,508,805	No gene	-	-	

Vitamin E Activity	AX-171134134	10	34,010,861	34,010,496 - 34,015,586	108983891 108983893	34,010,620 - 34,011,067 34,013,577 - 34,014,069	**synaptotagmin-5-like** receptor-like protein 12	4
Vitamin E Activity	AX-170893393	9	15,280,292	15,271,076 - 15,281,282	108994350 108994314	15,271,151 - 15,280,670 15,280,768 - 15,281,289	**E3 ubiquitin-protein ligase HOS1-like** 2-oxoglutarate-Fe(II) type oxidoreductase-like	5

[a] Chr, abbreviation for Chromosome

[b] Physical position given in bp

[b] LD blocks are defined using the "solid spine of LD" method

The candidate genes in **bold** overlap the physical position of the associated SNP

Discussion

The INRAE Walnut Germplasm Collection Has a High Degree of Fruit Phenotypic Variability

The walnut genetic resources of INRAE form an *ex situ* orchard near Bordeaux, France, and gather numerous accessions from worldwide. For the need of his breeding program, Eric Germain brought accessions from 23 countries, including European (France, Spain, Germany, Switzerland) American (California, Chile), and Asian (China, India, Japan) continents. The main goal was to collect a wide range of particular and interesting phenotypes regarding all traits economically important for walnut industry, and constituting a reservoir of potential genitors. By comparison with other panels used for GWAS, we observed a larger degree of fruit phenotypic variability. For instance, in the panel used in Iran, the range for the weight of a walnut was between 7.71 and 20.11 g (Arab et al., 2019), whereas in our panel, the range was between 5.22 and 22.51 g, all years combined. Similarly, in the work on suture strength of the University of Davis, California, the use of texture analyzer showed a variation between 6.94 to 63.14 kg-force (i.e. between 68.06 to 619.19 Newtons) among the 556 accessions mainly from 39 biparental progenies, so a variation of 551.13 N (Sideli et al., 2020). In our germplasm, with only 170 accessions all years combined, the variation extends from 74.40 to 776.97, i.e. a variation of 702.57 N. By increasing the phenotypic variability and the origin diversity of the panel used for GWAS analysis, knowing that most of these traits are highly quantitative, we may expect to find more easily the major loci involved in their variation.

Marker-Trait Associations of Correlated Traits Are Not Impacted by Different Phenotyping Methods and Give Major Loci

In our study, we have tried to improve the accuracy of measurement of several traits from one phenotyping campaign to the following. This is the case for three of the nut morphometrics related traits: in 2017, we used an electronic caliper to measure the nut length, the nut face diameter and the nut profile diameter, but we decided to develop a X-ray CT method for the 2018 campaign. We had initially expressed some doubts about changing our phenotyping method on GWAS results. However, the resulting BLUPs appear to be reliable since we found the association on Chr 14 for the nut face diameter which is also detected for the nut volume and the kernel volume (the three traits are significantly positively correlated), only measured in 2018 and for which no BLUPs were calculated.

Performing GWAS on correlated traits can give clues to dissect the complex genetic architecture of such quantitative traits like a nut size. Nut face diameter, nut volume and kernel volume are all correlated traits determined in part by the SNP AX-170690867 (Chr 14, 1,248,953 bp), the alleles of which are G (minor allele, frequency 0.30) and T (major allele), having a positive effect (+0.74 BLUPs unit, +2,101.08 mm^3 and +521.01 mm^3, respectively). The trait "x3 extreme groups" is influenced by the SNP AX-171170293 (Chr 14, 1,248,839 bp), with the alleles A (minor, frequency 0.47) and G (major), having a negative effect (-9.83 BLUPs unit). They belong to a haplotype blocks made of 4 SNPs with 5 haplotypes, whose haplotype of interest 'NNAT' has a frequency of only 0.23. The two SNPs are in moderate LD (r^2 = 0.37) and belong to the sequence of a *beta-galactosidase* encoding gene. This finding may suggests allelic heterogeneity.

In addition to probable allelic heterogeneity, a complex trait such as the nut size would depend on numerous factors. Since the previously described association on Chr 14 is involved in three different dimensions of the nut (nut face diameter, nut volume and kernel volume), it may be the major locus controlling the global nut size. The values of R^2 (percentage explained

variance) are agreement with this hypothesis. The higher R^2 is obtained for each of the three traits for this association (i.e. for nut face diameter: 26.36 vs. 19.80, 5.88 and 3.22). However, by phenotyping close and correlated traits, we were able to detect minor and more specific loci involved in the nut size. For instance, if nut volume and kernel volume are both traits sharing the major association on Chr 14, they both have specific other signals.

Close Germplasms Lead to Similar Results

As previously explained, we decided to go further with our GWAS results regarding the cracking related traits. The recent work published by the University of Davis highlighted a locus on Chr 5 controlling the suture strength (Sideli et al., 2020). They used among other manual evaluation using a hammer (scoring scale from 1 to 9) and texture analyzer recording the initial rupture, the maximum force and the integral of the force applied. Using manual evaluation and FarmCPU model, the team found the SNP AX-170748528 (Chr 5, 13,023,760 bp) associated. This SNP is only 762 bp apart from our peak SNP of the no-significant signal for the suture strength initial rupture. We both proposed the *lamin-like protein* encoding gene as candidate gene for the suture strength variation. Similarly, the team of the University of Davis found three SNPs associated on the Chr 11 using FarmCPU model, depending on the phenotype recorded with texture analyzer. The genomic region spans from 10,995,387 bp (maximum force) to 13,714,234 bp (initial rupture). We found a no-significant signal at 12,909,083 bp. The INRAE walnut germplasm and the UCDavis germplasm share numerous accession, such as ‘Chandler’, ‘Robert Livermore’, ‘Vina’ and diverse representing accessions from China, Japan and France. We can assume that all following factors combined are the reason we found the same major loci: shared accessions, similar quantitative phenotyping method and same SNP array.

On the contrary, we did not find any association in common by comparing our results with those of the team of the University of Tehran, Iran, using the same array (Arab et al. 2019). They found three significant marker-trait associations on Chr 7 for the nut weight, but we did not found any association on Chr 7 for this trait. For the nut length, we found associations on Chr 5, 8, 11 and 15, while they found associations on Chr 10 and 13. We unsuccessfully compared all our 11 traits in common. This may highlights the crucial role of the plant material used in a GWAS panel. The team of the University of Tehran, used a panel of 95 genotypes representing local populations across four different Iranian areas from valleys to mountains. If the panel represents a large part of Iranian walnut populations' genetic diversity, it is not surprising for us to not be able to confirm associations previously found. The marker-traits associations found by the team of the University of Tehran may be more specific to their germplasm. This finding proves the difficulty to develop markers for selection on a global scale (Mohammadi et al., 2020).

Contribution for the Better Understanding of the Genetic Control of Fruit Quality

The different found candidate genes overlapping the SNPs associated give clues to decipher the genetic architecture of the traits studied. The global nut size is a result of complex signaling pathways processes involved in cell cycle, including cell expansion. Serine/threonine kinases are enzymes catalyzing the phosphorylation from ATP to a amino acid residue (Diallo and Prigent, 2011). Several studies suggest that receptor-like cytosolic kinases contribute to plant development, particularly in cell wall function (Muto et al., 2004; Steinwand and Kieber, 2010) and cell morphogenesis (Becraft, 2002). Similarly, beta-galactosidases are enzymes catalyzing hydrolyzation of galactosyl residues of hemicellulose and pectin from the cell wall (Smith and Gross, 2000; Yang et al., 2018), and L-Ala-D/L-

amino acid epimerases are enzymes catalyzing the epimerization of various dipeptides (Lukk et al., 2012), involved in the peptidoglycan pathway essential for cell wall integrity in bacteria, but also found in land plants, although lacking in those peptidoglycans (Yang et al., 2013). The *beta-galactosidase*, *L-Ala-D/L-amino acid epimerase-like* and *receptor-like cytosolic serine/threonine-protein kinase RBK1* are therefore interesting candidate genes involved in nut morphometrics.

Apart from the nut size, the weight of the kernel, the seed of the pseudodrupe, and the ease of cracking are ones of the most important traits in walnut quality. We found a *BEL1-like homeodomain protein 4* encoding gene associated with the weight of 50 kernels. In *Arabidopsis thaliana*, BEL1 protein is a commonly found transcription factor required for ovule morphogenesis (Modrusan et al., 1994) and in potato, BEL1-type proteins enhance tuberization (Sharma et al., 2014), making our candidate gene likely involved in kernel weight. In addition, we found a *protein TPX2-like* encoding candidate gene for the nut face strength. TPX2 is a protein phosphorylated during mitosis interphase acting as a spindle assembly factor for microtubules in the nucleus, and is therefore crucial for cell division.

Using the "solid spine of LD" for the definition of the LD blocks, we highlighted complementary candidate genes for several marker-trait associations. The nut volume is the trait for which we listed the most candidate genes. Apart from the gene coding for a *beta-galactosidase* related to the most significant association on Chr 14, we found an *expansin-A4-like* and a *scarecrow-like protein 14* encoding genes on Chr 7, involved in loosening of cell wall during growth (Cosgrove, 2000) and flower bud transition (Quan et al., 2019). We also found a *calcium-binding protein PBP1-like* encoding gene on Chr 9 and calcium is known to be important during fruit growth whose uptake increases (Song et al., 2018). Like the team of the University of Davis, we identified a gene encoding for a *lamin-like protein* for the nut

suture strength. Lamins are proteins part of the lamina, providing a mechanical support to the nuclear envelope and also acting on nucleus size and shape (Ciska and Moreno Díaz de la Espina, 2014).

The selection of appropriate model and threshold is crucial in GWAS. We mainly found all our significant marker-trait associations using FarmCPU model and they were unsurprisingly more numerous using a 5% Bonferroni threshold. Nowadays, FarmCPU seems to better perform in controlling both false positives without compromising false negatives; it uses a MLMM divided into fixed effect and random effect model iteratively (Liu et al., 2016). FarmCPU was not widely used for complex traits in crops because of a lacking comparison with previous existing models. However, MLMM and FarmCPU models have been recently compared using soybean and maize well-known GWAS dataset and authors confirm that it controlled both false positives and false negatives based on Q-Q plots analysis and identification of the known associations (Kaler et al., 2020). In addition, the authors claim that MLMM did not find any significant marker using to conservative multiple comparison adjustment method. That is for this reason we decided to also retain associations using a 5% Bonferroni threshold to prevent from false negatives. In Manhattan plots, the $-\log_{10}$ of the p-value decreases from 7.56 automatically implemented to 6.86 for significance for 5% Bonferroni threshold, giving 58 significant SNPs in total in place of 37 using 1%.

First List of Candidate Genes Involved in Fatty Acids and Vitamin E Contents Variation in Kernel

For the first time in walnut, we discovered SNPs and candidate genes associated with monounsaturated fatty acids and tocopherols contents. Primary metabolites in plants such as lipids and vitamins are synthesized by numerous multi-enzymatic complexes and we found a

gene coding for a *calmodulin-7* overlapping the SNP associated on Chr7 with monounsaturated fatty acids content. In *de novo* fatty acids synthesis, the fatty acid synthase requires NAD(P)H as cofactor and the dependant nicotinamide adenine dinucleotide (NAD) signaling is thought to be highly regulated by calmodulin (Tai et al., 2019). However, even supposed to be inheritable, those traits need to be phenotyped several years to confirm the candidate genes, since they are probably influenced by environment.

Conclusions

By combining a highly accurate characterization of crucial traits on a diverse walnut germplasm collection, a dense SNP genotyping and newly available multi-locus models, our GWAS showed numerous marker-trait associations, particularly for traits related to the global nut size. We confirmed the major loci involved in the suture strength and we proposed candidate genes helping to decipher the genetic architecture of fruit quality, mainly linked to cell wall function for the nut size and linked to calcium signaling for the fatty acids contents. After phenological traits, the INRAE walnut germplasm repository proved its suitability for GWAS and will permit to release new walnut cultivars meeting the criteria required by both consumers and producers.

Data availability

The phenotypic raw dataset generated and used in this work is available in S2 Table. The SNP genotyping raw dataset in “hapmap” format is freely and openly accessed on the “Portail Data INRAE” official institute repository, via the identifier “INRA’s Walnut Hapmap” and the following Digital Object Identifier (DOI): https://doi.org/10.15454/XPKII8. The dataset is called “GWAS_hapmap.txt”. The additional file called “List of ID.tab” allows to link the

array identifier name of the accessions used in this study with the identifier name of the INRAE walnut germplasm collection.

Author contributions

AB performed all the genetic analysis and wrote the manuscript. JC assisted AB in performing, analyzing and interpreting the data of weight and cracking related traits. AD provided his experience in GWAS to interpret the results. ED, FL and AB conceived and coordinated the research. All coauthors revised the manuscript.

Funding

This work has been mainly funded by the “Région Nouvelle-Aquitaine” in the project “INNOV’noyer”, coordinated by the CTIFL, and in partnership with the INRAE of Bordeaux. This work has been also funded by the “Cifre” convention number 2016/1558 of ANRT (Agence Nationale de la Recherche et de la Technologie). To finish, Anthony Bernard thanks the “Initiative d’Excellence” (IdEx) program of the University of Bordeaux and the “Dufrenoy-Crédit Agricole d’Ile-de-France Mécénat” grant for contributing to the funding of his PhD.

Conflict of interest statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgments

First, we would like to thank the late Eric Germain, former head of the breeding program at the INRAE of Bordeaux from 1977 to 2007. His remarkable work, then continued by Francis

Delort, has given us the opportunity to study a rich set of plant material. We thank the Fruit Tree Experimental Unit of the INRAE in Toulenne and the *Prunus/Juglans* Genetic Resources Center for the maintenance of the collection and for helping us to collect the fruits. We want to thank the GIS Fruits consortium for the funding of the Master's degree fellowship of Julie Crabier. We acknowledge the GEVES for X-ray CT analyses. The measurements were performed on a phenotyping platform, PHENOTIC, dedicated to instrumentation and imagery for seeds, seedlings and plants in Angers, France (https://www6.inrae.fr/phenotic/; https://doi.org/10.15454/U2BWFJ). We acknowledge the ITERG laboratory in Canéjan, France, for the fatty acids and tocopherols quantifications (http://iterg.com/index.php/fr/home/). Then, the CTIFL, holder of the project "INNOV'noyer", in partnership with the INRAE of Bordeaux, want to thank the "Région Nouvelle-Aquitaine", and "Cifre" convention of "ANRT" (Agence Nationale de la Recherche et de la Technologie). It is also important to note that the project is supported by the "Agri Sud-Ouest Innovation" competitiveness cluster.

References

Ahandani, E. A., Ramandi, H. D., Sarmad, J., Samani, M. A., Yavari, A., Ahandani, R. A., et al. (2014). Evaluation of morphological diversity among some Persian walnut accessions (*Juglans regia* L.) in Guilan, northern Iran. *Int. J. Plant Biol. Res.* 2, 1015–1022.

Arab, M. M., Marrano, A., Abdollahi-Arpanahi, R., Leslie, C. A., Askari, H., Neale, D. B., et al. (2019). Genome-wide patterns of population structure and association mapping of nut-related traits in Persian walnut populations from Iran using the Axiom *J. regia* 700K SNP array. *Sci. Rep.* 9, 6376.

Aradhya, M. K., Potter, D., Simon, C. J. (2006). Cladistic Biogeography of *Juglans* (*Juglandaceae*) Based on Chloroplast DNA Intergenic Spacer Sequences. In *Darwins Harvest*

New Approaches Orig. Evol. Conserv. Crops. Eds. Motley, T. J., Zerega, N., Cross, H. (New-York, Columbia University Press), 143–170.

Barrett, J. C., Fry, B., Maller, J., Daly, M. J. (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics.* 21, 263–265.

Bates, D., Maechler, M., Bolker, B., Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Softw*. 67, 1–48.

Becraft, P. W. (2002). Receptor Kinase Signaling in Plant Development. *Annu. Rev. Cell Dev. Biol*. 18, 163–192.

Bernard, A., Lheureux, F., Dirlewanger, E. (2018). Walnut: past and future of genetic improvement. *Tree Genet. Genomes*. 14, 1.

Bernard, A., Marrano, A., Donkpegan, A., Brown, P. J., Leslie, C. A., Neale, D. B., et al. (2020). Association and linkage mapping to unravel genetic architecture of phenological traits and lateral bearing in Persian Walnut (*Juglans regia* L.). *BMC Genomics.* 21, 203.

Burghardt, L. T., Young, N. D., Tiffin, P. (2017). A Guide to Genome-Wide Association Mapping in Plants. *Curr. Protoc. Plant Biol.* 2, 22–38.

Ciska, M., Moreno Díaz de la Espina, S. (2014). The intriguing plant nuclear lamina. *Front. Plant Sci.* 5, 166.

Cosgrove, D. J. (2000). Loosening of plant cell walls by expansins. *Nature*. 407, 321–326.

Diallo, A., Prigent, C. (2011). The serine/threonine kinases that control cell cycle progression as therapeutic targets. *Bulletin du Cancer*. 98, 1335–1345.

Eskandari, S., Hassani, D., Abdi, A. (2005). Investigation on genetic diversity of Persian walnut and evaluation of promising genotypes. *Acta Hortic.* 705, 159–166.

Frichot, E., François, O. (2015). LEA: An R package for landscape and ecological association studies. *Methods Ecol. Evol.* 6, 925–929.

Furbank, R. T., Tester, M. (2011). Phenomics – technologies to relieve the phenotyping bottleneck. *Trends Plant Sci.* 16, 635–644.

Ghasemi, M., Arzani, K., Hassani, D. (2012). Evaluation and identification of walnut (*Juglans regia* L.) genotypes in Markazi province of Iran. *Crop Breed. J.* 2, 119–124.

Harrison, R. D., Gardner, W. A., Tollner, W. E., Kinard, D. J. (1993). X-ray computed tomography studies of the burrowing behaviour of fourth-instar pecan weevil (*Coleoptera*: *Curculionidae*). *J. Econ. Entomol.* 86, 1714–1719.

Kaler, A. S., Gilman, J. D., Beissinger, T., Purcell L. C. (2020). Comparing Different Statistical Models and Multiple Testing Corrections for Association Mapping in Soybean and Maize. *Front. Plant Sci.* 10, 1794.

Khadivi-Khub, A., Ebrahimi, A., Sheibani, F., Esmaeili, A. (2015). Phenological and pomological characterization of Persian walnut to select promising trees. *Euphytica.* 205, 557–567.

Khosa, I., Pasero, E. (2014). Feature extraction in X-ray images for hazelnuts classification. *Proc. Int. Jt. Conf. Neural. Netw.* 2354-2360.

Kim, S., Schatzki, T. (2001). Detection of pinholes in almonds through X-ray imaging. *Trans. ASAE.* 44, 997–1003.

Kotwaliwale, N., Singh, K., Kalne, A., Jha, S. N., Seth, N., Kar, A., et al. (2014). X-ray imaging methods for internal quality evaluation of agricultural produce. *J. Food Sci. Technol.* 51, 1–15.

Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P.J., et al. (2012). GAPIT: genome association and prediction integrated tool. *Bioinformatics.* 28, 2397–2399.

Liu, X., Huang, M., Fan, B., Buckler, E. S., Zhang, Z. (2016). Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. *PLoS Genet.* 12, e1005767.

Lukk, T., Sakai, A., Kalyanaraman, C., Brown, S. D., Imker, H. J., Song, L., et al. (2012). Homology models guide discovery of diverse enzyme specificities among dipeptide epimerases in the enolase superfamily. *Proc. Natl. Acad. Sci. USA.* 109, 4122–4127.

Mahmoodi, R., Dadpour, M. R., Hassani, D., Zeinalabedini, M., Vendramin, E., Micali, S., et al. (2019). Development of a core collection in Iranian walnut (*Juglans regia* L.) germplasm using the phenotypic diversity. *Sci. Hortic.* 249, 439–448.

Manichaikul, A., Mychaleckyj, J. C., Rich, S. S., Daly, K., Sale, M., Chen, W. M. (2010). Robust relationship inference in genome-wide association studies. *Bioinformatics.* 26, 2867–2873.

Marrano, A., Britton, M., Zaini, P. A., Zimin, A. V., Workman, R. E., et al. (2020). High-quality chromosome-scale assembly of the walnut (*Juglans regia* L) reference genome. *bioRxiv* 809798 [Preprint], doi:https://doi.org/10.1101/809798.

Marrano, A., Martínez-García, P. J., Bianco, L., Sideli, G. M., Di Pierro, E. A., Leslie, C. A., et al. (2019a). A new genomic tool for walnut (*Juglans regia* L.): development and validation of the high-density Axiom™ *J. regia* 700K SNP genotyping array. *Plant Biotechnol. J.* 17, 1027–1036.

Marrano, A., Sideli, G. M., Leslie, C. A., Cheng, H., Neale, D. B. (2019b). Deciphering of the Genetic Control of Phenology, Yield and Pellicle Color in Persian Walnut (*Juglans regia* L.). *Front. Plant Sci.* 10, 1140.

McGranahan, G., Leslie, C. (2009). Breeding Walnuts (*Juglans regia*). In *Breeding Plantation Tree Crops: Temperate Species*. Eds. Jain, S. M., Priyadarshan, P. M. (Springer Science), 249–273.

Modrusan, Z., Reiser, L., Feldmann, K. A., Fischer, R. L., Haughn, G. W. (1994). Homeotic transformation of ovules into carpel-like structures in *Arabidopsis*. *Plant Cell*. 6, 333–349.

Mohammadi, M., Xavier, A., Beckett, T., Beyer, S., Chen, L., Chikssa, H., et al. (2020). Identification, deployment, and transferability of quantitative trait loci from genome-wide association studies in plants. *Curr. Plant Biol.* In Press, available online 27 March 2020.

Muto, H., Yabe, N., Asami, T., Hasunuma, K., Yamamoto, K. T. (2004) Overexpression of constitutive differential growth 1 gene, which encodes a RLCK VII- subfamily protein

kinase, causes abnormal differential and elongation growth after organ differentiation in *Arabidopsis*. *Plant Physiol*. 136, 3124–3133.

Poggetti, L., Ermacora, P., Cipriani, G., Pavan, F., Testolin, R., et al. (2017). Morphological and carpological variability of walnut germplasm (*Juglans regia* L.) collected in North-Eastern Italy and selection of superior genotypes. *Sci. Hortic.* 225, 615–619.

Quan, S., Niu, J., Zhou, L., Xu, H., Ma, L., Qin, Y. (2019). Genome-wide Identification, Classification, Expression and Duplication Analysis of GRAS Family Genes in *Juglans regia* L. *Sci. Rep.* 9, 11643.

Rondeau, A.-S. (1997). Le Noyer. Eds. Actes Sud.

Segura, V., Vilhjálmsson, B. J., Platt, A., Korte, A., Seren, Ü., Long, Q., et al. (2012). An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nat. Genet.* 44, 825–830.

Sharma, P., Lin, T., Grandellis, C., Yu, M., Hannapel, D. J. (2014). The BEL1-like family of transcription factors in potato. *J. Exp. Bot.* 65, 709–723.

Sideli, G. M., Marrano, A., Montanari, S., Leslie, C. A., Allen, B. J., et al. (2020). Quantitative phenotyping of shell suture strength in walnut (*Juglans regia* L.) enhances precision for detection of QTL and genome-wide association mapping. *PLoS ONE*. 15(4), e0231144.

Smith, D. L., Gross, K. C. (2000). A family of at least seven β-galactosidase genes is expressed during tomato fruit development. *Plant Physiol*. 123, 1173–1183.

Steinwand, B. J., Kieber, J. J. (2010). The Role of Receptor-Like Kinases in Regulating Cell Wall Function. *Plant Physiol*. 153, 479–484.

Tai, L., Li, B.-B., Nie, X.-M., Zhang, P.-P., Hu, C.-H., Zhang, L., et al. (2019). Calmodulin Is the Fundamental Regulator of NADK-Mediated NAD Signaling in Plants. *Front. Plant Sci.* 10, 681.

Vahdati, K., Arab, M. M., Sarikhani, S., Sadat-Hosseini, M., Leslie, C. A., Brown, P. J. (2019). Advances in Persian Walnut (*Juglans regia* L.) Breeding Strategies. In *Advances in Plant Breeding Strategies: Nut and Beverage Crops*. Eds. Al-Khayri, J., Jain, S. M., Johnson, D. V. (Switzerland, Springer), 401–472.

Wei, T., Simko, V. R. (2017). Package "corrplot": Visualization of a Correlation Matrix (Version 0.84). https://github.com/taiyun/corrplot.

Wickham, H. (2017). Tidyverse: Easily Install and Load the 'Tidyverse'. R package version 1.2.1. https://CRAN.R-project.org/package=tidyverse.

Woodworth, R. H. (1930). Meiosis of micro-sporogenesis within the *Juglandaceae*. *Am. J. Bot.* 17, 863–869.

Yang, H., Liu, J., Dang, M., Zhang, B., Li, H., Meng, R., et al. (2018). Analysis of β-Galactosidase During Fruit Development and Ripening in Two Different Texture Types of Apple Cultivars. *Front. Plant Sci*. 9, 539.

Yang, Z., Wang, Y., Zhou, Y., Gao, Q., Zhang, E., Zhu, L., et al. (2013). Evolution of land plant genes encoding L-Ala-D/L-Glu epimerases (AEEs) via horizontal gene transfer and positive selection. *BMC Plant Biol*. 13, 34.

Zeneli, G., Kola, H., Dida, M. (2005). Phenotypic variation in native walnut populations of Northern Albania. *Sci. Hortic.* 105, 91–100.

5. Contribution au management des ressources génétiques du noyer

Un des challenges dans le management d'une collection est le fait de préserver le maximum de diversité génétique et de variabilité phénotypique au sein du plus petit nombre d'accessions puisque tout gestionnaire possède un budget limité (consommables, terrain alloué et personnel). Avant de construire ces core collections, une étape clé est la détermination de la structure de la population artificielle et des apparentements entre les accessions. Dans ces études, le choix approprié du type de marqueur moléculaire a son importance.

Les SNP ayant connu une forte progression dans notre ère des outils génomiques, un débat est né quant à l'utilité des SSRs pourtant très utilisés pour ce type d'études précédemment cité. Ainsi, sur un panel de 150 des 170 accessions précédemment utilisées en GWAS, il a été proposé de comparer la puissance des SSRs par rapport à l'ensemble des SNP de la puce à haute densité de 700K SNP. Le premier niveau de structure K=2, lié à l'origine géographique des accessions, est détecté de manière équivalente avec les deux types de marqueurs, soit avec 13 SSRs ou 364 275 SNP. Toujours pour K=2, il n'y a aucune modification en termes d'accessions dans la constitution des deux groupes lorsque les SSR et les SNP sont comparés. De plus, il a été montré qu'une centaine de SNP choisis aléatoirement le long du génome donnent une structure basée sur des ACoP comparable aux SSRs. Ensuite, plusieurs core collections ont été proposées, en utilisant les deux types de marqueurs, deux méthodes de construction, puis deux seuils différents (basé tantôt sur un nombre minimal d'accessions à retenir, tantôt sur un pourcentage de diversité allélique total à conserver). Les deux types de marqueurs se sont avérés être toujours appropriés pour la construction de core collections. Un fort pourcentage d'accessions en commun est retrouvé dans la core collection quel que soit le marqueur, la méthode de construction et le seuil. Au final, le type de marqueur à préférer dépendra évidemment du but recherché et également du coût de développement et de génotypge de celui-ci

L'article suivant a été publié en ligne le 17 septembre 2020 dans le journal Tree Genetics and Genomes, volume 16:76, sous le nom de « Comparison of structure analyses and core collections for the management of walnut genetic resources » (DOI 10.1007/s11295-020-01469-5).

Tree Genetics & Genomes (2020) 16:76
https://doi.org/10.1007/s11295-020-01469-5

ORIGINAL ARTICLE

Comparison of structure analyses and core collections for the management of walnut genetic resources

Anthony Bernard[1,2] · Teresa Barreneche[1] · Armel Donkpegan[1] · Fabrice Lheureux[2] · Elisabeth Dirlewanger[1]

Received: 27 July 2020 / Revised: 2 September 2020 / Accepted: 3 September 2020
© The Author(s) 2020

Abstract
The preservation of maximum diversity within the smallest number of accessions is one of the challenges of germplasm management, and the assessment of the population structure, the relationships between the accessions, and the construction of core collections are the key steps. The choice of suitable molecular markers is often the starting point. In this study, we analyzed a part of the INRAE walnut germplasm collection, which is highly diverse and unique in Europe, consisting of 150 accessions from American, European, and Asian continents. Based on genotypic data already available, using 13 SSRs or 364,275 SNPs, we showed that the first level of structure is detected equally. We also highlighted empirically that approximately 100 SNPs are needed to obtain similar clustering to 13 SSRs in Principal Coordinate Analysis (PCoA). We constructed eight core collections following two strategies (percentage of total allelic diversity or number of accessions) and two construction methods based on different algorithms ("maximum length subtree" and "entry-to-nearest-entry" methods). We showed that core collections based on few SSR markers are able to capture at least 99.5% of the SNP allelic diversity, irrespective of the construction method used. Then, core collections based on each marker type are highly similar, using both construction methods. All these steps are crucial to identify the suitable tools and methods to improve plant genetic resources management.

Keywords Germplasm management · SNP · SSR · Core collection · Walnut

Introduction

In the context of climate change and human population growth, plant genetic resources (PGR) are of the utmost importance. They constitute the foundations of agricultural sustainability and global food safety (Ulukan 2011; Ogwu et al. 2014), but they are facing conservation issues. Regarding ex situ management of PGR, this mainly involves seeds and some collections are at risk, due in part to the fact that they are generally underfunded and that evaluation and characterization are often imprecise or inadequate (Fu 2017). In that respect, careful PGR management is crucial from storage to use (Maxted et al. 1997). For clonally propagated perennial species, such as fruit and nut tree species, PGR are generally conserved in ex situ orchards as grafted cultivars, which has its pros and cons: the main advantages are that they can be stored under the climate conditions of their intended use, and can be evaluated during storage; but on the other hand, they require a lot of space and the cost of conservation is significant (Hammer and Teklu 2008).

Nowadays, molecular tools contribute to PGR management (Wambugu et al. 2018), since they can help to find genetically close or synonymous accessions and to create "core collections" which will contain maximum genetic diversity within the smallest number of accessions, leading to a reduction in conservation costs, among other things. Simple Sequence Repeats (SSRs) were frequently used because of their high polymorphism, but the development of high-throughput genomics tools has led to the use of Single-Nucleotide Polymorphisms (SNPs). In the Persian walnut (*Juglans regia* L.), a species which is widely disseminated and grown in many temperate regions, more than 20 publications mention the use of SSRs (Bernard et al. 2018a). This type of markers was

Communicated by M. Troggio

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11295-020-01469-5) contains supplementary material, which is available to authorized users.

✉ Elisabeth Dirlewanger
elisabeth.dirlewanger@inrae.fr

[1] University Bordeaux, INRAE, Biologie du Fruit et Pathologie, UMR 1332, F-33140 Villenave d'Ornon, France

[2] CTIFL, Centre Opérationnel de Lanxade, 24130 Prigonrieux, France

Published online: 16 September 2020

particularly used to decipher the genetic diversity and the structure of numerous walnut accessions at a continental scale (Ebrahimi et al. 2016). SSRs were also used to perform a landscape genetic analysis across the Asian range of Persian walnut, species native to the foothills of the Himalayas (Pollegioni et al. 2014). The authors showed that reservoirs of high genetic diversity still exist, confirming the importance of the conservation of walnut genetic resources in this area. Recently, a high-density Axiom™ *J. regia* 700K SNP genotyping array was developed and validated, initiating a novel genomic era in walnut (Marrano et al. 2019a). This genomics tool was used for instance to investigate the genome-wide population structure in Iranian walnut populations (Arab et al. 2019), and to perform genome-wide association studies (Bernard et al. 2020; Marrano et al. 2019b).

As a result, it is fair to wonder how to conduct the tasks related to germplasm management, such as core collections construction, and the choice of the type of marker to be used is one of the first questions. Neutral SSR loci, due to slippage during replication, usually mutate much more frequently than SNP loci, leading to population-specific alleles that are useful for revealing population structure (Tsykun et al. 2017). On the other hand, since they are usually developed in small numbers for one species, they therefore may not reflect genome-wide genetic diversity contrary to SNP loci that are much more frequent in the genome of most species (Ljungqvist et al. 2010). These two types of marker bring different views of the structure, and the merits of each are listed in Guichoux et al. (2011). SSRs and SNPs were compared in order to assess population structure and relatedness in short-lived species such as rice (Singh et al. 2013), maize (Hamblin et al. 2007; Van Inghelandt et al. 2010; Yang et al. 2011), sunflower (Filippi et al. 2015), bean (Müller et al. 2015), and cowpea (Desalegne et al. 2017). These comparisons are rarer for perennials (grapes (Emanuelli et al. 2013) and jujube (Chen et al. 2017)). A recent work in an apple gene bank compared population structure and relatedness using 15 SSRs and more than 15,000 SNPs. However, studies on core collections construction are lacking.

The aim of this paper is to identify the type of marker and the core collections construction methods more relevant for routine walnut germplasm management.

Materials and methods

Plant materials and genotypic data

The plant material consists in 150 unique accessions of *Juglans regia* from American, European, and Asian continents. The trees are maintained at the *Prunus* and *Juglans* Genetic Resources Center located in the Fruit Experimental Unit of INRAE in Toulenne (latitude 44° 34′ 37.442″ N–longitude 0° 16′ 51.48″ W), near Bordeaux, France (S1 Table). The 150 accessions were selected following two steps. (*i*) We created a GWAS panel of 170 accessions (Bernard et al. 2020) from SSR and SNP genotyping were described and made publicly available (Bernard et al. 2018b; Bernard et al. 2020).

Structure analyses

Principal Coordinate Analysis (PCoA) was used to determine the structure patterns among the 150 accessions (Fig. 1). Dissimilarities, based on allelic data, were calculated with 10,000 bootstraps, and transformed into Euclidean distances using a power transformation of 0.5. PCoA was performed using the DARwin 6.0.14 software (Perrier and Jacquemoud-Collet 2006), supplemented by the scatterplot3d R package for 3D visualization.

As linked SNPs can account for too much in the population structure variance, particularly in linkage disequilibrium (LD) regions (Price et al. 2008), a pruned subset of SNPs was also used for the PCoA. This filtering was completed using the PLINK 1.9 software, keeping only the SNPs with the higher minor allele frequency, and based on a threshold of $r^2 = 0.2$ (command–indep-pairwise 50 5 0.2). A subset of 100 SNPs, a number similar to that of SSR alleles, randomly selected, was also tested using the PLINK 1.9 software (command–thin-count 100) to compare PCoA results.

The genetic structure of our panel was also investigated using two types of analyses depending on the markers. Bayesian model–based analysis using the STRUCTURE 2.3.4 software was implemented for the SSR markers (Pritchard et al. 2000), following the method of our previous work (Bernard et al. 2018b). The ΔK method (Evanno et al. 2005), implemented in STRUCTURE harvester (Earl and von Holdt 2012), was used to determine the most likely K.

Using SNP markers, a sparse non-negative matrix factorization algorithm was implemented using the sNMF 2.0 software, available as a function of the LEA R package (Frichot and François 2015). This software has a fast and efficient program for estimating individual admixture coefficients from large genomic datasets, and produces results very close to Bayesian clustering programs such as STRUCTURE (Frichot et al. 2014). The choice of the best number of clusters (K) is based on a cross-entropy criterion implemented in the LEA R package. For SSR and SNP markers, thresholds of 0.8 and 0.7 for admixture coefficients, respectively, were chosen to consider one accession as admixed.

Next, the genetic relationships between the 150 accessions were also assessed with the Neighbor-Joining method (Saitou and Nei 1987) using the DARwin 6.0.14 software, for both the SSR and SNP datasets. The Unweighted Neighbor-Joining option was used to build the trees.

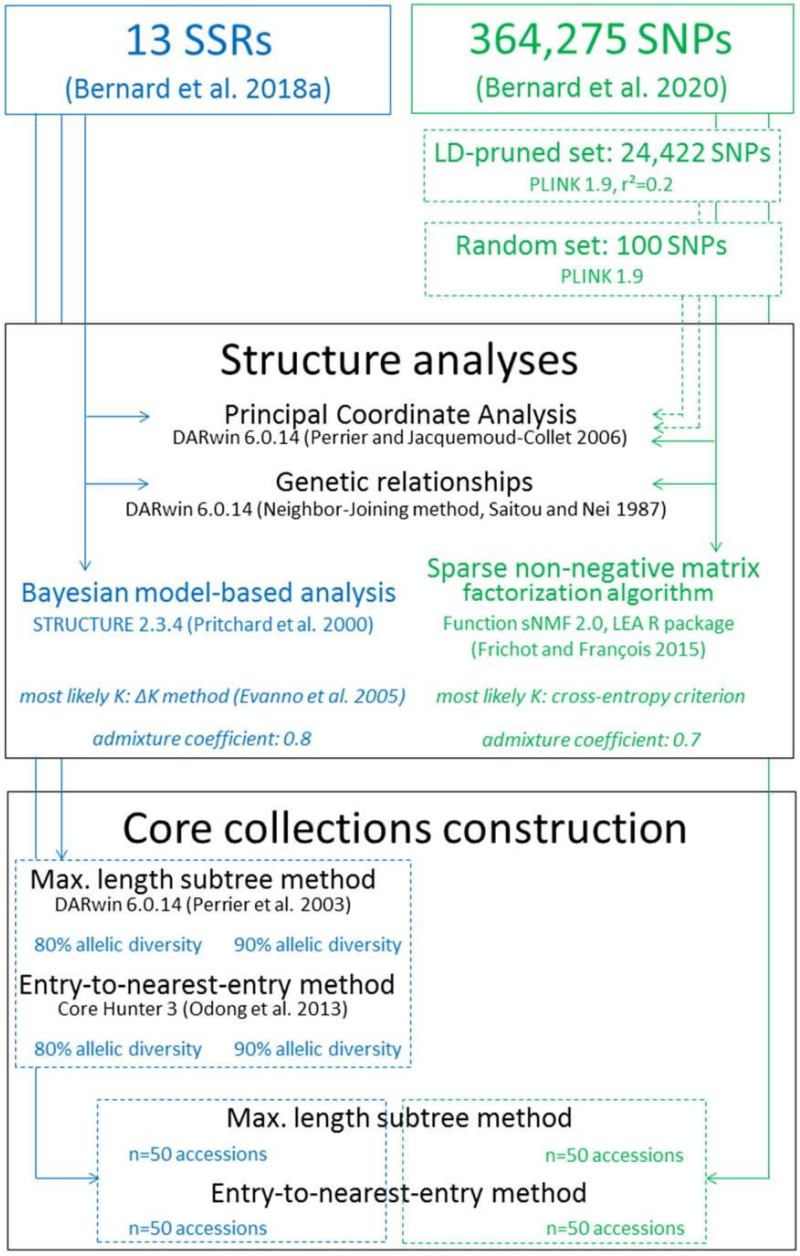


Fig. 1 Workflow of the structure analyses and the core collections construction

Core collections construction

Several core collections based on both marker types were constructed using two methods and two selection thresholds (Fig. 1). (*i*) The two methods: we used the "maximum length sub tree" function (Perrier et al. 2003) of the DARwin 6.0.14 software, which looks for a subset of accessions minimizing the redundancy between them, and limiting where possible the loss of diversity (diversity here is expressed by the tree built), and the "entry-to-nearest-entry" method (Odong et al. 2013), implemented in the Core Hunter 3 software (De Beukelaer et al. 2018), which looks for a subset of accessions as different as possible from each other, avoiding selecting a few clusters of similar accessions at the extreme ends of the distribution. (*ii*) The two selection thresholds: using the two methods mentioned above, we first used the SSR dataset to capture at least 80% or 90% of total allelic diversity, and the number of alleles captured in one SSR-based core collection was also estimated in the SNP dataset (selection threshold based on percentage of total allelic diversity). Secondly, we used a sampling intensity of 33% (n = 50/150) to observe the differences in the selected accessions, using these two methods (selection threshold based on a number of selected accessions).

Results

Comparison of the first level of structure using SSR and SNP markers

Using SSR markers, the higher drop of ΔK is for $K = 2$, followed by a rise for $K = 6$, and very similar results were obtained using SNP markers, since the higher drop of the cross-entropy criterion is for $K = 2$, with a curve slope starting for $K = 6$ (Fig. 2). To compare the results using SSR and SNP markers, we focused on the first level of structure for $K = 2$ (S2 Table). The individual admixture coefficients are shown for each of the 150 accessions (Fig. 3). For both marker types, the clustering is linked to the geographic origin of the accessions, as we already described (Bernard et al. 2018b; Bernard et al. 2020). Group A contains the accessions from Eastern Europe and Asia (named "E"), and group B contains the accessions from Western Europe and America (named "W").

Using SSR markers, with a threshold of 0.8 for the individual admixture coefficient, we found 17 admixed accessions, so 88.7% of the accessions were assigned to a group (Table 1). They include 11 French and Californian modern hybrids and 6 accessions from Eastern Europe, Israel, and Central Asia, confirming our previous SSR analysis.

Using SNP markers and a threshold of 0.7, we found 24 admixed accessions (84% of assignment), including 9 of the

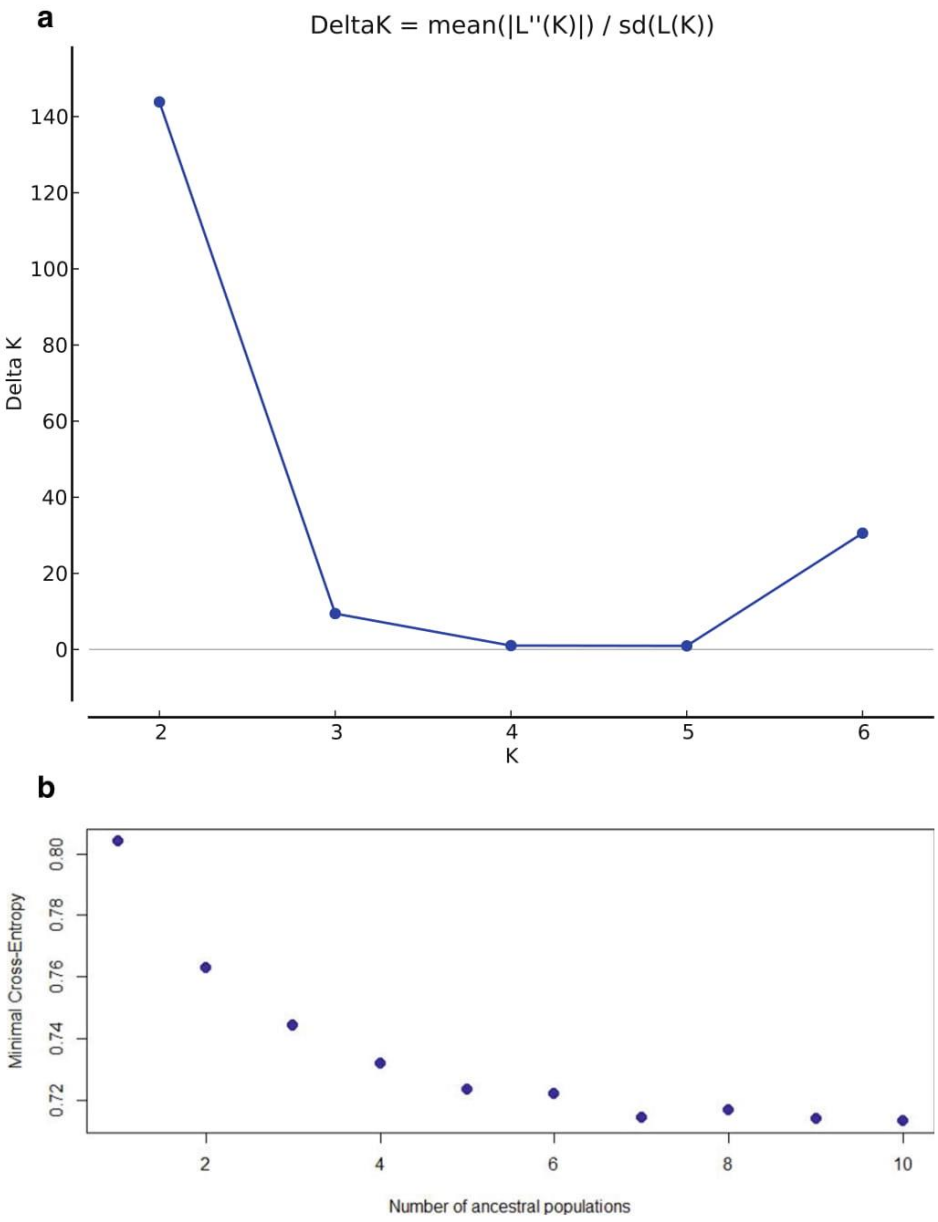


Fig. 2 The most likely K subpopulations. K was evaluated with (**a**) the ΔK method using SSRs, and (**b**) the cross-entropy criterion using SNPs

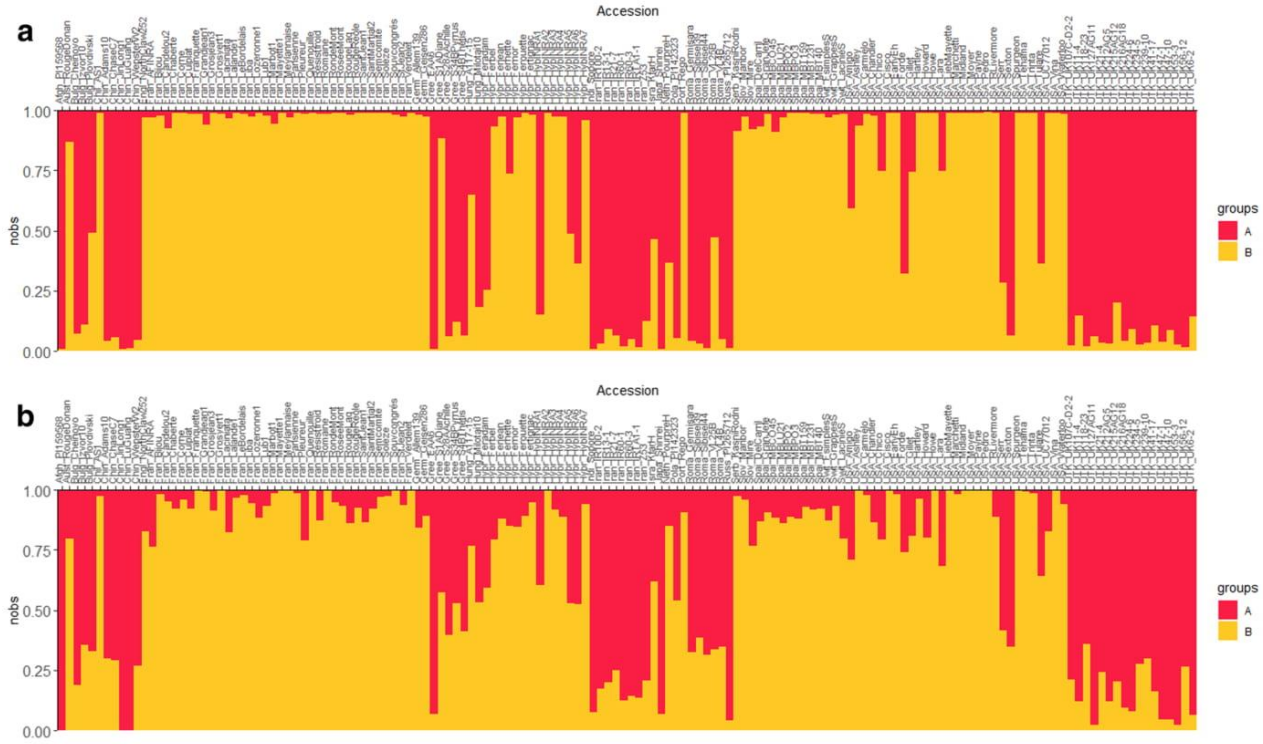


Fig. 3 Bar plots showing the individual admixture coefficients of the 150 accessions for $K = 2$. Structure was assessed (**a**) using SSRs, and (**b**) using SNPs. The accessions are ordered by their country of origin, in alphabetical order. Group A in red contains the accessions from Eastern Europe and Asia ("E"), whereas group B in yellow contains the accessions from Western Europe and America ("W") and hybrids from INRAE

17 found using SSR markers (Fig. 4). From these admixed accessions, 7 are now clustered into group B, and "UK 215AG12" is now in group A. In addition, 14 accessions from group A, and 1 from group B, based on SSR clustering, are found admixed using SNP markers (Fig. 4). Therefore, only 23/150 accessions (15%) are differently clustered. At all events, we did not find any group exchange from A to B, or from B to A, when comparing the clustering based on SSR or on SNP markers.

When using a threshold of 0.8 for the SNPs, the percentage of population assignment for $K = 2$ is low (70.7%) but is still high for a threshold of 0.7 (84.0%) (Table 1), showing that 0.7 and 0.8 are more appropriate thresholds for SNPs and SSRs, respectively. In addition, we ran a Spearman rank correlation test for $K = 2$ and found that the clustering between SSR and SNP markers is highly correlated, up to 84%.

Table 1 Percentage of population assignment to a group from $K = 2$ to $K = 6$, using SSR and SNP markers

	$K = 2$	$K = 3$	$K = 4$	$K = 5$	$K = 6$
SSR (admixture threshold = 0.8)	88.7	85.3	73.3	59.3	68.7
SNP (admixture threshold = 0.7)	84.0	62.7	56.0	49.3	44.7
SNP (admixture threshold = 0.8)	70.7	47.3	36.0	26.7	22.0

The PCoAs constructed in 2D and 3D show the clustering of the 150 accessions following the results obtained from $K = 2$ (Fig. 5). For $K = 2$, the PCoA results are in agreement with the structure based on SSR or SNP markers, since the scatterplots for groups A and B are well-defined by the first principal component. The admixed accessions are positioned mainly between the two groups. For both methods, the three Manregian walnuts ("Chase C7", "Wepster W2," and "Adams 10") are found to be genetically diverse. Regarding the percentage of explained variation, they are also comparable using both types of markers. The first three axes (x, y, and z) explain 21.86% of the cumulative variation for SSRs, and 14.91% for SNPs.

However, we found some differences between the two types of marker. Using SSR markers, the scatterplot corresponding to group A is more extensive, whereas group B is more scattered using SNP markers. Additionally, this group is split in two, showing the French accessions on one side and the Californian ones on the other.

The Neighbor-Joining method implemented in DARwin 6.0.14 was used to construct grouping trees with the 150 accessions (Fig. 6). The main branching groups of the trees

obtained with both markers are in agreement with the structure results ($K = 2$), since they are mainly defined by groups A and B. The two accessions "Jin Long 1" from China and "PI 15 95 68" from Afghanistan are among the longest length branches. However, few differences were detected between the structure and the tree, with accessions from group A found in the branching group mainly corresponding to group B: 9 accessions using SSR markers, and 4 accessions using SNP markers.

We then inferred the individual admixture coefficients of the 150 accessions from $K = 3$ to $K = 6$ for both markers (S2 Table). Bar plots are given for information in S1-S4 Figs.

Comparison of three sets of SNPs for PCoA assessment

In addition, for PCoA assessment, we compared the entire set of 364,275 SNPs with a subset of SNPs filtered for LD, with a

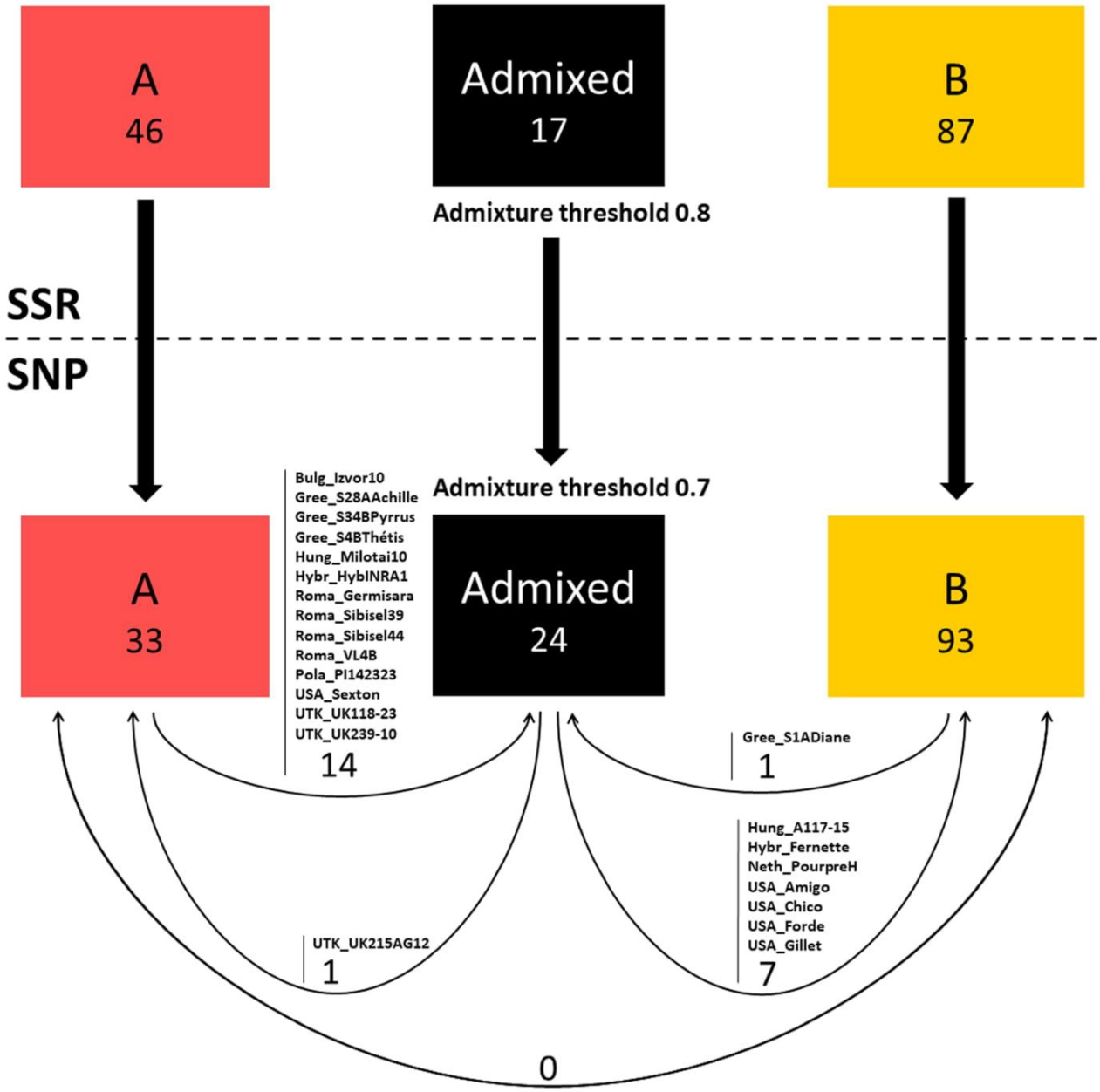


Fig. 4 SNP-based clustering results, compared with SSR-based clustering results. For a threshold of 0.7 for admixture, SNPs clustered 14 accessions found in group A, and one found in group B, compared with SSRs, into the admixed group. Conversely, SNPs clustered one and seven accessions found to be admixed, compared with SSRs, into groups A and B, respectively. There was no clustering exchange between groups A and B, when comparing both markers

threshold of $r^2 = 0.2$ (Fig. 7). We then retained 24,422 SNPs, or 6.7% of the entire set. Interestingly, the results found for both datasets are strongly similar, with the scatterplots well-distinguished, according to the $K = 2$ results (Fig. 7). We still distinguish the French accessions from the Californian accessions within group B. The main difference is that the variance is better explained by the first three axes using the LD-pruned set (18.90% vs. 14.91%). By comparing the entire set with a random subset of 100 SNPs, a range comparable with the total number of SSR alleles, we found that the scatterplots are less well-defined, but still in agreement with the $K = 2$ results (Fig. 7). In this case, we cannot distinguish the French accessions from the Californian accessions within group B.

Comparison of core collections using SSR and SNP markers

In light of the results obtained concerning structure analyses (13 SSRs give similar results compared to 364,275 SNPs), we decided to construct first four core collections based on SSR markers. We created them with both core collections construction methods and two thresholds, to capture at least 80% or 90% of total allelic diversity. To capture at least 80% of total SSR allelic diversity, we need 16 accessions using the method based on the "maximum length sub tree" function of DARwin 6.0.14, and 15 accessions using the "entry-to-nearest-entry" method of Core Hunter 3 (Table 2). Next, to capture at least 90% of allelic diversity, we need 27 and 32 accessions for each construction method, respectively. The accessions

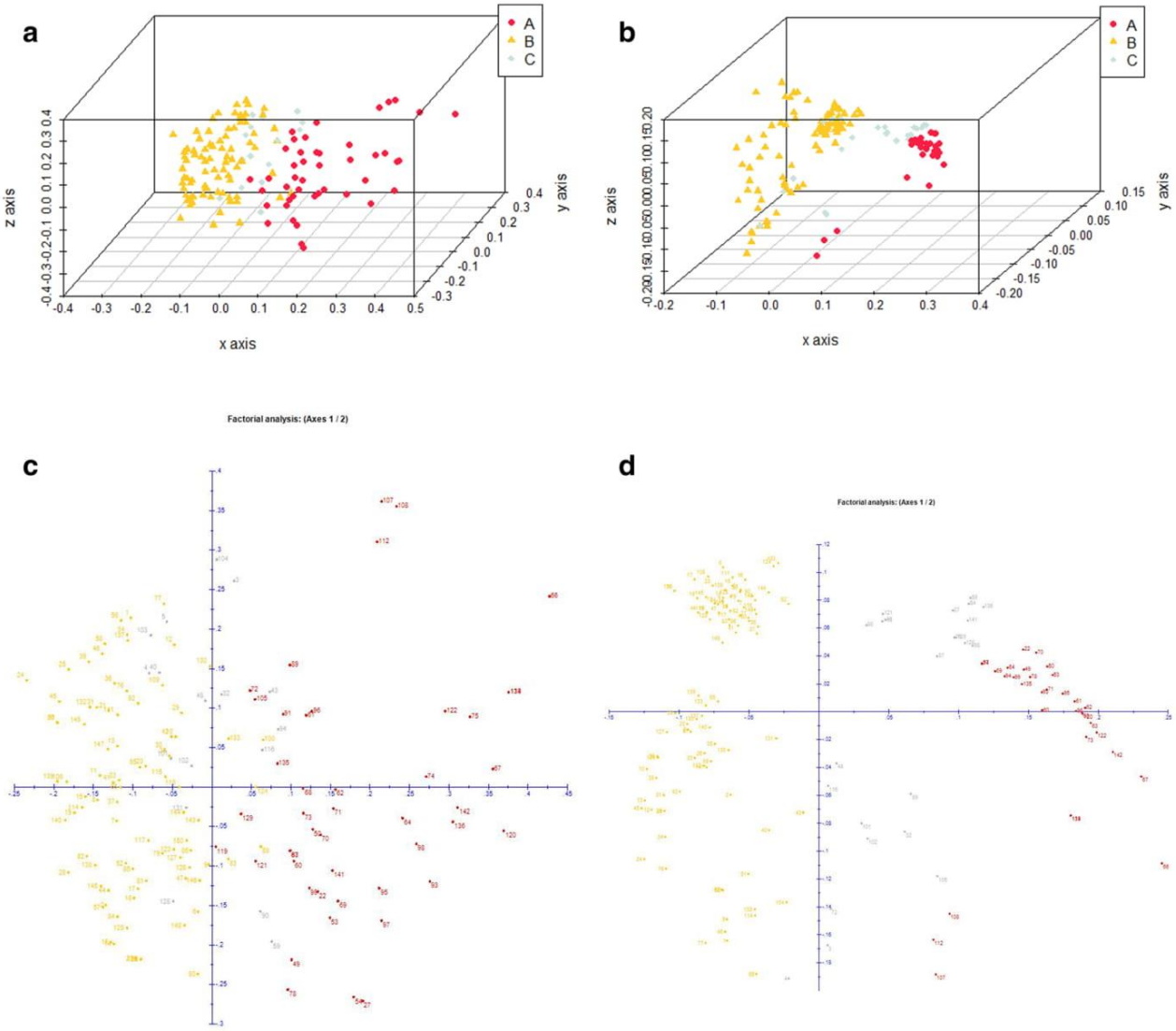


Fig. 5 Principal Coordinate Analysis scatterplots. PCoAs were constructed in 3D using (**a**) SSR, and (**b**) SNP markers, and in 2D using (**c**) SSR, and (**d**) SNP markers. The 150 accessions are colored following the $K = 2$ results: group A in red, group B in yellow, and admixed in grey

belong mainly to group A, from Eastern Europe and Asia, known to be more diverse: 11/16 and 12/15 respectively for 80%, and 20/27 and 21/32 respectively for 90%. They include, for instance, Iranian accessions, the Indian "Sopore" and several accessions from the Botanical Garden of Kiev. Regarding the French accessions, few landraces such as "Corne," "Marbot" and "Grandjean1" were kept only when capturing 90% of allelic diversity using "Core Hunter 3".

Moreover, the consistency of the results between both percentages of total captured allelic diversity was checked for the two core collection construction methods. Using Darwin 6.0.14, the 16 accessions selected to capture 80% of allelic diversity were found among the 27 selected to capture 90% (Table 2). However, using Core Hunter 3, 2/15 accessions selected to capture 80% were not found among the 32 selected to capture 90%. Hence, for 80%, 9 accessions (among 16 or 15) are in common between the two methods, and 19 accessions for 90% (among 27 or 32). We then estimated the number of alleles captured in the SNP dataset for each core collection. For the two SSR-based core collections capturing at least 80% of total allelic diversity, they also capture more than 99.5% of total SNP allelic diversity. For 90%, they reach more than 99.9%.

Next, we constructed four additional core collections using the same sampling intensity threshold of 33% to select 50/150 accessions, with both methods and both markers. Results show that 34/50 accessions are in common between the SSR and SNP markers using Darwin 6.0.14, and 32/50 accessions using Core Hunter 3 (Table 3). The accessions belong mainly to group A, from Eastern Europe and Asia, known to be more diverse (between 30 and 31/50 using SSRs, and between 22 and 23/50 using SNPs). They both include the Iranian accessions, the Chinese "Jin Long 1," "PI 15 95 68" from Afghanistan, the Indian "Sopore," the Bulgarian "Izvor 10" and "Plovdivski," and several accessions from the Botanical Garden of Kiev.

We also checked the consistency of the results between the two core collection construction methods for both markers. Using SSR markers, 39/50 accessions are in common between the two methods, and 45/50 accessions using SNP markers. Regarding the number of alleles, for SSR markers, we captured 97.1% and 94.3% of the 105 total alleles found within the entire collection, using the "maximum length subtree" and the "entry-to-nearest entry" methods, respectively. Using SNP markers, we captured more than 99.9% of the 728,550 total alleles found with the two methods (Table 3).

Discussion

The use of either SSR or SNP markers shows comparable results for the collection structure

By genotyping the panel of 150 *J. regia* accessions using 13 SSRs and more than 300,000 SNPs, we obtain similar results. Both types of marker showed a first level of structure for $K = 2$, with no exchange of accession between the main groups (A and B), which are related to the geographical origin of the accessions. Similar findings were reported in Ebrahimi et al. (2016), with $K = 2$ mainly corresponding to a group from Asia, and a group of European and North Africa. The exchanges observed between both types of marker only concern

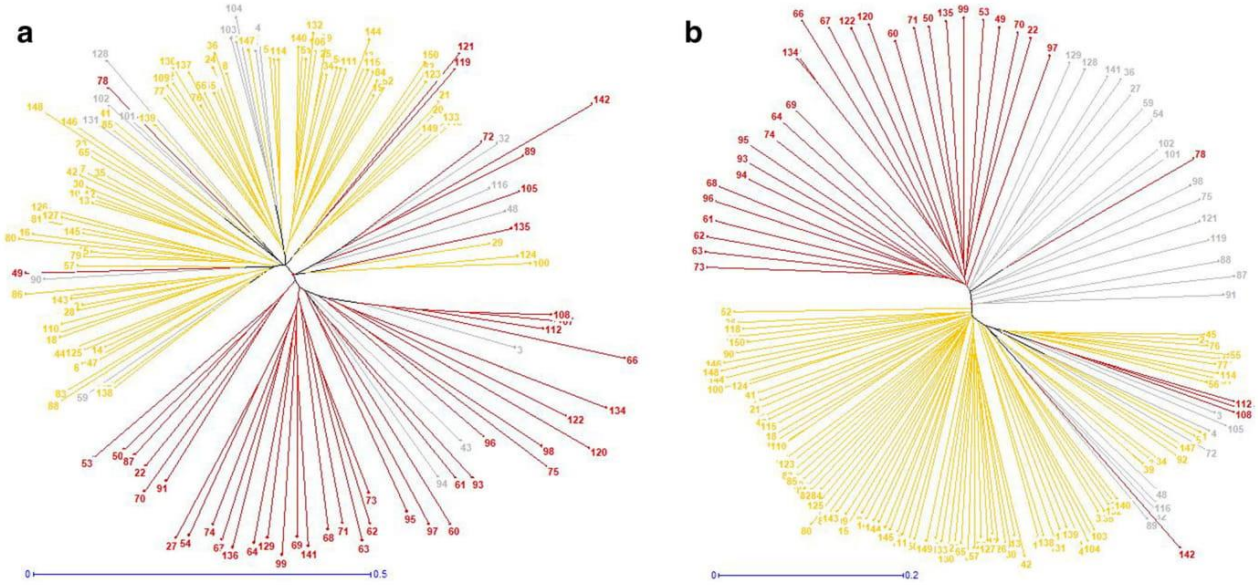


Fig. 6 Neighbor-Joining trees. Trees were constructed using (**a**) SSR, and (**b**) SNP markers. The 150 accessions are colored following the $K = 2$ results: group A in red, group B in yellow, and admixed in grey

accessions that switched from one group to the admixed cluster, or vice versa. From $K = 3$ to $K = 6$, the results remain highly comparable, with the highlighting of substructures linked, for instance, to "Payne" pedigree for the Californian

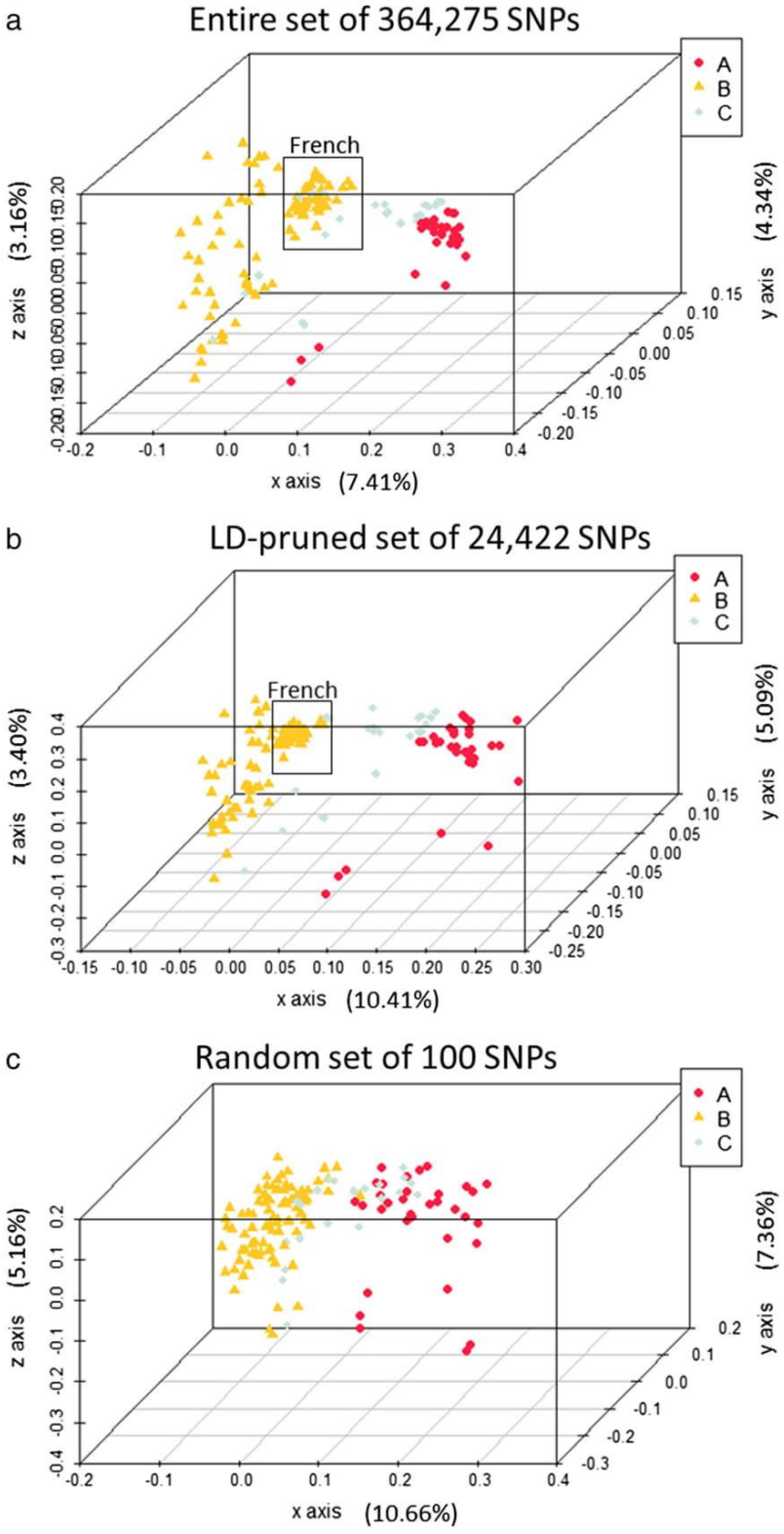
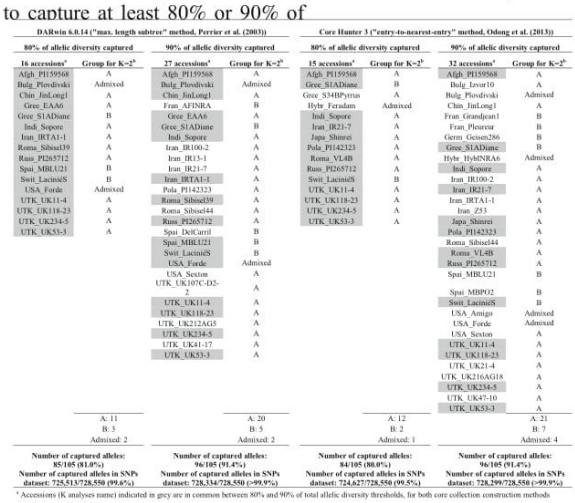


Fig. 7 Comparison of SNP sets for Principal Coordinate Analysis. PCoA were constructed using (**a**) the entire set of 364,275 SNPs, (**b**) the LD-pruned subset of 24,422 SNPs, and (**c**) the random set of 100 SNPs

Table 2 Construction of the core collections using SSRs and two different methods of construction, to capture at least 80% or 90% of total allelic diversity



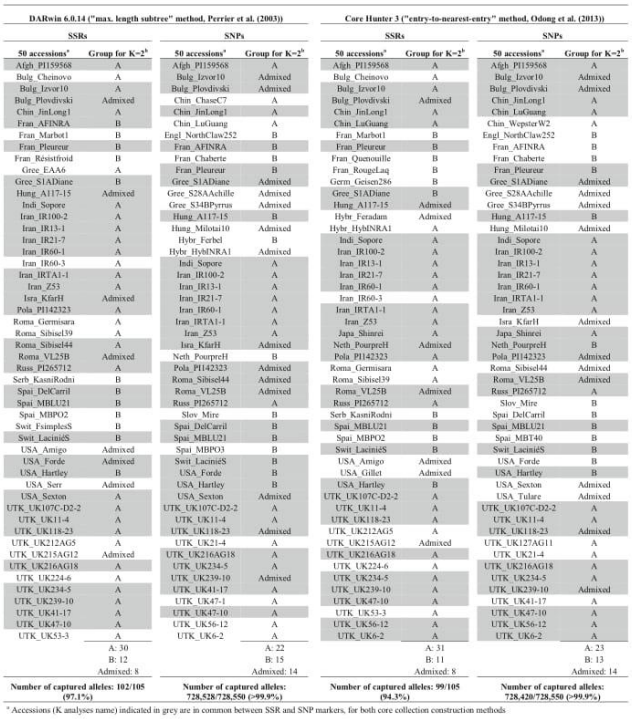
DARwin 6.0.14 ("max. length subtree" method, Perrier et al. (2003))					Core Hunter 3 ("entry-to-nearest-entry" method, Odong et al. (2013))			
80% of allelic diversity captured		90% of allelic diversity captured			80% of allelic diversity captured		90% of allelic diversity captured	
16 accessions[a]	Group for K=2[b]	27 accessions[a]	Group for K=2[b]		15 accessions[a]	Group for K=2[b]	32 accessions[a]	Group for K=2[b]
Afgh_PI159568	A	Afgh_PI159568	A		Afgh_PI159568	A	Afgh_PI159568	A
Bulg_Plovdivski	Admixed	Bulg_Plovdivski	Admixed		Gree_S1ADiane	B	Bulg_Izvor10	A
Chin_JinLong1	A	Chin_JinLong1	A		Gree_S34BPyrrus	A	Bulg_Plovdivski	Admixed
Gree_EAA6	A	Fran_AFINRA	B		Hybr_Feradam	Admixed	Chin_JinLong1	A
Gree_S1ADiane	B	Gree_EAA6	A		Indi_Sopore	A	Fran_Grandjean1	B
Indi_Sopore	A	Gree_S1ADiane	B		Iran_IR21-7	A	Fran_Pleureur	B
Iran_IRTA1-1	A	Indi_Sopore	A		Japa_Shinrei	A	Germ_Geisen286	B
Roma_SibiselJ9	A	Iran_IR100-2	A		Pola_PI142323	A	Gree_S1ADiane	B
Russ_PI265712	A	Iran_IR13-1	A		Roma_VL4B	A	Hybr_HybINRA6	Admixed
Spai_MBLU21	B	Iran_IR21-7	A		Russ_PI265712	A	Indi_Sopore	A
Swit_LaciniéS	B	Iran_IRTA1-1	A		Swit_LaciniéS	B	Iran_IR100-2	A
USA_Forde	Admixed	Pola_PI142323	A		UTK_UK11-4	A	Iran_IR21-7	A
UTK_UK11-4	A	Roma_SibiselJ9	A		UTK_UK118-23	A	Iran_IRTA1-1	A
UTK_UK118-23	A	Roma_SibiselJ44	A		UTK_UK234-5	A	Iran_Z53	A
UTK_UK234-5	A	Russ_PI265712	A		UTK_UK53-3	A	Japa_Shinrei	A
UTK_UK53-3	A	Spai_DelCarril	B				Pola_PI142323	A
		Spai_MBLU21	B				Roma_SibiselJ44	A
		Swit_LaciniéS	B				Roma_VL4B	A
		USA_Forde	Admixed				Russ_PI265712	A
		USA_Sexton	A				Spai_MBLU21	B
		UTK_UK107C-D2-2	A				Spai_MBPO2	B
		UTK_UK11-4	A				Swit_LaciniéS	B
		UTK_UK118-23	A				USA_Amigo	Admixed
		UTK_UK212AG15	A				USA_Forde	Admixed
		UTK_UK234-5	A				USA_Sexton	A
		UTK_UK41-17	A				UTK_UK11-4	A
		UTK_UK53-3	A				UTK_UK118-23	A
							UTK_UK21-4	A
							UTK_UK216AG18	A
							UTK_UK234-5	A
							UTK_UK47-10	A
							UTK_UK53-3	A
A: 11 B: 3 Admixed: 2		A: 20 B: 5 Admixed: 2			A: 12 B: 2 Admixed: 1		A: 21 B: 7 Admixed: 4	
Number of captured alleles: 85/105 (81.0%) Number of captured alleles in SNPs dataset: 725,513/728,550 (99.6%)		Number of captured alleles: 96/105 (91.4%) Number of captured alleles in SNPs dataset: 728,334/728,550 (>99.9%)			Number of captured alleles: 84/105 (80.0%) Number of captured alleles in SNPs dataset: 724,627/728,550 (99.5%)		Number of captured alleles: 96/105 (91.4%) Number of captured alleles in SNPs dataset: 728,299/728,550 (>99.9%)	

[a] Accessions (K analyses name) indicated in grey are in common between 80% and 90% of total allelic diversity thresholds, for both core collection construction methods

[b] Admixture threshold of 0.8

modern hybrids, or to geographical area (South-West vs. South-Est) for French landraces. When comparing structure results for $K = 2$ with PCoAs, and with grouping trees, the results are still consistent. More precisely, we found as the most diverse accessions, using both markers, the three Manregian walnuts originating from seed collected in north-eastern China and the two accessions "PI 15 95 68" and "Jin Long 1," both coming from Asia (Bernard et al. 2018b).

Table 3 Construction of the core collections using both marker types, SSRs and SNPs, and two different methods, to keep $n = 50$ accessions



DARwin 6.0.14 ("max. length subtree" method, Perrier et al. (2003))				Core Hunter 3 ("entry-to-nearest-entry" method, Odong et al. (2013))			
SSRs		SNPs		SSRs		SNPs	
50 accessions[a]	Group for K=2[b]	50 accessions[a]	Group for K=2[b]	50 accessions[a]	Group for K=2[b]	50 accessions[a]	Group for K=2[b]
Afgh_PI159568	A	Afgh_PI159568	A	Afgh_PI159568	A	Afgh_PI159568	A
Bulg_Cheinovo	A	Bulg_Izvor10	Admixed	Bulg_Cheinovo	A	Bulg_Izvor10	Admixed
Bulg_Izvor10	A	Bulg_Plovdivski	Admixed	Bulg_Izvor10	A	Bulg_Plovdivski	Admixed
Bulg_Plovdivski	Admixed	Chin_ChaseC7	A	Bulg_Plovdivski	Admixed	Chin_JinLong1	A
Chin_JinLong1	A	Chin_JinLong1	A	Chin_JinLong1	A	Chin_LuGuang	A
Fran_AFINRA	B	Chin_LuGuang	A	Chin_LuGuang	A	Chin_WepsterW2	A
Fran_Marbot1	B	Engl_NorthClaw252	B	Fran_Marbot1	B	Engl_NorthClaw252	B
Fran_Pleureur	B	Fran_AFINRA	B	Fran_Pleureur	B	Fran_AFINRA	B
Fran_Resistfroid	B	Fran_Chaberte	B	Fran_Quenouille	B	Fran_Chaberte	B
Gree_EAA6	A	Fran_Pleureur	B	Fran_RougeLaq	B	Fran_Pleureur	B
Gree_S1ADiane	B	Gree_S1ADiane	Admixed	Germ_Geisen286	B	Gree_S1ADiane	Admixed
Hung_A117-15	Admixed	Gree_S28AAchille	Admixed	Gree_S1ADiane	B	Gree_S28AAchille	Admixed
Indi_Sopore	A	Gree_S34BPyrrus	Admixed	Hung_A117-15	Admixed	Gree_S34BPyrrus	Admixed
Iran_IR100-2	A	Hung_A117-15	B	Hybr_Feradam	Admixed	Hung_A117-15	B
Iran_IR13-1	A	Hung_Milotai10	Admixed	Hybr_HybINRA1	A	Hung_Milotai10	Admixed
Iran_IR21-7	A	Hybr_Ferhel	B	Indi_Sopore	A	Indi_Sopore	A
Iran_IR60-1	A	Hybr_HybINRA1	Admixed	Iran_IR100-2	A	Iran_IR100-2	A
Iran_IR60-3	A	Indi_Sopore	A	Iran_IR13-1	A	Iran_IR13-1	A
Iran_IRTA1-1	A	Iran_IR100-2	A	Iran_IR21-7	A	Iran_IR21-7	A
Iran_Z53	A	Iran_IR13-1	A	Iran_IR60-1	A	Iran_IR60-1	A
Isra_KfarH	Admixed	Iran_IR21-7	A	Iran_IR60-3	A	Iran_IRTA1-1	A
Pola_PI142323	A	Iran_IR60-1	A	Iran_IRTA1-1	A	Iran_Z53	A
Roma_Germisara	A	Iran_IRTA1-1	A	Iran_Z53	A	Isra_KfarH	Admixed
Roma_SibiselJ39	A	Iran_Z53	A	Japa_Shinrei	A	Japa_Shinrei	A
Roma_SibiselJ44	A	Isra_KfarH	Admixed	Neth_PourpreH	Admixed	Neth_PourpreH	B
Roma_VL25B	Admixed	Neth_PourpreH	B	Pola_PI142323	A	Pola_PI142323	Admixed
Russ_PI265712	A	Pola_PI142323	Admixed	Roma_Germisara	A	Roma_SibiselJ44	Admixed
Serb_KasniRodni	B	Roma_SibiselJ44	Admixed	Roma_SibiselJ39	A	Roma_VL25B	Admixed
Spai_DelCarril	B	Roma_VL25B	Admixed	Roma_VL25B	Admixed	Russ_PI265712	A
Spai_MBLU21	B	Russ_PI265712	A	Russ_PI265712	A	Slov_Mire	B
Spai_MBPO2	B	Slov_Mire	B	Serb_KasniRodni	B	Spai_DelCarril	B
Swit_FsimplexS	B	Spai_DelCarril	B	Spai_MBLU21	B	Spai_MBLU21	B
Swit_LaciniéS	B	Spai_MBLU21	B	Spai_MBPO2	B	Spai_MBT40	B
USA_Amigo	Admixed	Spai_MBPO3	B	Swit_LaciniéS	B	Swit_LaciniéS	B
USA_Forde	Admixed	Swit_LaciniéS	B	USA_Amigo	Admixed	USA_Forde	B
USA_Hartley	B	USA_Forde	B	USA_Gillet	Admixed	USA_Hartley	B
USA_Serr	Admixed	USA_Hartley	B	USA_Hartley	B	USA_Sexton	Admixed
USA_Sexton	A	USA_Sexton	Admixed	UTK_UK107C-D2-2	A	USA_Tulare	Admixed
UTK_UK107C-D2-2	A	UTK_UK107C-D2-2	A	UTK_UK11-4	A	UTK_UK107C-D2-2	A
UTK_UK11-4	A	UTK_UK11-4	A	UTK_UK118-23	A	UTK_UK11-4	A
UTK_UK118-23	A	UTK_UK118-23	Admixed	UTK_UK212AG5	A	UTK_UK118-23	Admixed
UTK_UK212AG5	A	UTK_UK21-4	A	UTK_UK215AG12	Admixed	UTK_UK127AG11	A
UTK_UK215AG12	Admixed	UTK_UK216AG18	A	UTK_UK216AG18	A	UTK_UK21-4	A
UTK_UK216AG18	A	UTK_UK234-5	A	UTK_UK224-6	A	UTK_UK216AG18	A
UTK_UK224-6	A	UTK_UK239-10	Admixed	UTK_UK234-5	A	UTK_UK234-5	A
UTK_UK234-5	A	UTK_UK41-17	A	UTK_UK239-10	A	UTK_UK239-10	Admixed
UTK_UK239-10	A	UTK_UK47-1	A	UTK_UK47-10	A	UTK_UK41-17	A
UTK_UK41-17	A	UTK_UK47-10	A	UTK_UK53-3	A	UTK_UK47-10	A
UTK_UK47-10	A	UTK_UK56-12	A	UTK_UK56-12	A	UTK_UK56-12	A
UTK_UK53-3	A	UTK_UK6-2	A	UTK_UK6-2	A	UTK_UK6-2	A
A: 30 B: 12 Admixed: 8		A: 22 B: 15 Admixed: 14		A: 31 B: 11 Admixed: 8		A: 23 B: 13 Admixed: 14	
Number of captured alleles: 102/105 (97.1%)		Number of captured alleles: 728,528/728,550 (>99.9%)		Number of captured alleles: 99/105 (94.3%)		Number of captured alleles: 728,420/728,550 (>99.9%)	

[a] Accessions (K analyses name) indicated in grey are in common between SSR and SNP markers, for both core collection construction methods

[b] Admixture threshold of 0.8 using SSRs, and 0.7 using SNPs

[a] Accessions (K analyses name) indicated in grey are in common between SSR and SNP markers, for both core collection construction methods

[b] Admixture threshold of 0.8 using SSRs, and 0.7 using SNPs

Consequently, SNP markers confirmed our results obtained using SSR markers (Bernard et al. 2018b). When considering the LD-pruned set of 6.7% of the entire set of SNPs, PCoA showed consistent clustering patterns and also consistency with those using the 13 SSRs.

These kinds of findings have previously been observed in other works, except that the number of markers needed to obtain a comparable resolution power would be different. For example, the broad patterns of PCoA were similar using 36 SSRs with 2.2 alleles per locus on average and 36 SNPs in 375 Indian accessions of rice (Singh et al. 2013). However, in local accessions of cowpea from East African countries, similar clustering patterns were found using more SNPs than SSRs (151 vs. 13) (Desalegne et al. 2017). In jujube, within a core-collection of 150 accessions, only 18 were classified into different groups based on the results of structure analysis using 24 SSRs and 4680 SNPs (Chen et al. 2017). Within various inbred maize lines, SSRs performed better at clustering accessions into groups using about 10 times more SNPs (Hamblin et al. 2007; Yang et al. 2011). In apple, another clonally propagated perennial, a comparison between 15 SSRs and more than 15,000 SNPs showed strong concordance between both markers for relatedness estimation and that about five to ten SNPs were equal to one SSR (Larsen et al. 2018). Other works suggest the use of three times (Müller et al. 2015) or seven to 11 times more SNPs to obtain comparable informative results (Van Inghelandt et al. 2010). In our study, we took care to choose highly polymorph and robust SSRs suitable for *J. regia* diversity studies, by reviewing the literature (Dangl et al. 2005; Woeste et al. 2002; Dang et al. 2016). When working with biallelic markers such as SNPs, it is known that the genetic distances can be equivalent to those calculated with SSRs using the formula $n(k^{-1})$, where k is the average number of alleles per locus, and n is the number of loci (Laval et al. 2002). With 13 loci and 8.1 alleles per locus on average (four times that for SNPs), we theoretically need $[13*(8.1^{-1})] = 93$ SNPs to obtain equivalent genetic distances in our panel. Our findings based on PCoA clustering using the random set of 100 SNPs are consistent with this number of SNPs.

Moreover, SNPs globally tend to give higher proportions of inferred admixture, as observed in sunflower (Filippi et al. 2015). Regarding the admixture thresholds chosen, the results found when comparing 0.7 or 0.8 for SNPs also show that 0.7 is clearly more suitable for obtaining comparable structure. In this respect, we found that the percentage of assignment decreases as the K increases, particularly for the SNPs. Such differences were also reported in maize (Yang et al. 2011) and grape (Emanuelli et al. 2013).

The choice of marker type will depend on the necessary task related to germplasm conservation or utilization

The management of PGR comprises the following main steps: their conservation, which consists in acquisition of plant material (through the collection or protection of reserves in situ, or by exchange of ex situ material), their maintenance (such as storing and propagation), their characterization (based on both genotype and phenotype), and their utilization for research, breeding programs, or production (Bretting and Widrlechner 1995). Due to the increasing availability of genomics tools, a cross-disciplinary field which aims to use genomics in germplasm management has grown (Jia et al. 2017). But undoubtedly, the choice of SSR or SNP markers will depend on the purposes. For obvious reasons, a first choice criterion could be the cost of genotyping. New genomics technologies have a cost that has decreased dramatically in recent years. Apart from the DNA extraction, the cost of SNP vs. SSR genotyping was about 2600 times less in our study. For guidance only, we paid USD$8.80 for 13 SSR loci per sample (0.7 USD$/locus/sample) and USD$98.90 for one array of 364,275 robust loci per sample (2.7E-4 USD$/locus/sample). However, SSR genotyping is often more "flexible," since we could choose precise numbers of loci and samples. In the case of a SNP array, all the loci available are assayed, for a 96-well DNA plate. In addition, only 59.8% of the available SNPs on the array were usable for the analyses after quality control. The cost comparison was also a topic of discussion in Larsen et al. (2018), in which SNP data was 10 times more expensive than SSR data. However, their SSR genotyping required much more manual preparation, such as PCR reactions.

A second choice criterion could be the nature of the plant material managed, particularly if the researcher works only on one crop, or with its wild species. In our case, the SSRs used were also highly transferable into wild species of the genus *Juglans* spp. (Bernard et al. 2018b). Conversely, the Axiom™ *J. regia* 700K SNP array used is valid on the cultivated species *J. regia* only and failed on our few wild species accessions tested. But this is not a general case. In Emanuelli et al.'s work (2013), the set of 384 SNPs used served to analyze 2273 accessions of grapes (*Vitis* spp.), including cultivated grapevines (*V. vinifera* ssp. *sativa*), wild grapevines (*V. vinifera* ssp. *sylvestris*), and non-*Vitis vinifera* species used as rootstocks.

A third reason to use either marker could be the main purpose of the genotyping. Well-chosen neutral SSRs would be sufficient for population structure and relationship determination, particularly in the first steps of germplasm management, since the computational time for analysis is lower. But SNPs, likely to be associated with functional variation, would be preferred for a genome-wide association study purpose (Vargas et al. 2016; Larsen et al. 2018).

Markers and construction methods must be compatible with the preservation of phenotypic variability

In light of our results, we found that SSR-based core collections could capture more than 99.5% of SNP total allelic diversity. SSR markers, when properly selected using the literature, may be the markers of choice to achieve this goal. Using SSRs, our analyses preferentially selected accessions from East Europe and Asia, as expected, because of their global higher genetic diversity, and helped us to understand that French landraces have a moderate level of genetic diversity. Given that French landraces represent 20% of the entire plant material panel, these findings confirm that their diversity is moderate. However, here it is important to note that the accessions selected in one core collection, even when slightly different, are not always the same in the repeated runs in Core Hunter 3, contrary to in Darwin 6.0.14. This feature, algorithm-dependent and more frequent when the panel has redundancy (Vargas et al. 2016), makes the comparison between construction methods more difficult.

In parallel with the preservation of allelic diversity, it is also necessary to take phenotypic variability into account. The INRAE walnut germplasm collection contains some accessions with unusual traits, such as weeping branches, laciniate leaves, or purple foliage, which may be used for ornamental purposes. Interestingly, the accession with weeping branches "Pleureur" is in the two core collections capturing 90% of SSR total allelic diversity, and the accession with laciniate leaves is in the two core collections constructed using Core Hunter 3. Based on chronological phenotypic data available (Bernard et al. 2019) and new data acquired, we also explored whether the core collections contain a high or low range of variability of some important traits. For the budbreak date for instance, crucial for climate change adaptation, the ten earlier accessions are "Early Ehrhardt," "Mire," "Payne," "Serr," "Kfar Hanania," "IR 60-1," "Sopore," "Z 53," "Ashley," and "Lu Guang," with a range of budbreak date from 65 to 75 Julian days. Five among the ten are found on the core collection "90% - Darwin." Conversely, none of the ten later accessions ("Fertignac," "Le Bordelais," "St Jean n°1," "Lalande," "Candelou," "Maribor," "Semence Comité Dordogne," "Ronde de Montignac," "Culplat," and "Romaine"), with a range of budbreak date from 110 to 122 Julian days, is found in the four core collections. Curators are well aware of the importance of phenotype knowledge for an effective PGR management. Currently, new ways to construct core collections are being evaluated to maximize the frequency of minor alleles and phenotypes (Vargas et al. 2016). An interesting application of such "custom methods" in carrot is well-described in Corak et al. (2019). The authors suggest that a core subset based on "custom methods," e.g., core balancing functional and genetic diversity based on high-density genomic data and estimated breeding values, would be directly useful for breeders and more efficient for breeding programs.

Conclusion

In our comparison using 150 *J. regia* accessions, both SSR and SNP markers were highly efficient for highlighting the INRAE walnut collection structure, even though the number of SSRs is much lower than the number of SNPs. Moreover, similar results were obtained for core collections, irrespective of the construction method and the marker type used. The strategy for core collection construction can be based on several criteria and must meet germplasm management requirements (genetic diversity) and constraints (cost and surface of conservation). The lower the number of selected accessions, the higher the probability of discarding a genotype with traits of interest. Thus, this highlights the importance of phenotypic evaluation in germplasm management.

It is therefore important to consider the task of germplasm management when choosing the most appropriate marker. In general, few SSR markers are suitable and sufficient for obtaining a global idea of the structure and the genetic diversity of a germplasm, and for constructing a core collection.

Acknowledgments We would like to thank the Fruit Tree Experimental Unit of the INRAE in Toulenne and the *Prunus/Juglans* Genetic Resources Center for the maintenance of the collection and for helping us to collect the samples. The CTIFL, as the INNOV'noyer project owner in partnership with INRAE Bordeaux, would like to thank the Région Nouvelle-Aquitaine, and the Cifre convention of ANRT (Agence Nationale de la Recherche et de la Technologie). It is also important to note that the project is supported by the Agri Sud-Ouest Innovation competitiveness cluster. Finally, we would like to thank the late Eric Germain, former head of the breeding program at INRAE Bordeaux from 1977 to 2007. His remarkable work, subsequently pursued by Francis Delort, has given us the opportunity to study a rich set of plant material.

Code availability Not applicable.

Data archiving statement The SSR genotyping dataset in "tab" format is freely and openly accessed in the Portail Data INRAE repository, via the identifier "INRAE's Walnut Genotyping Resources" and the following Digital Object Identifier (DOI): https://doi.org/10.15454/XPKII8. The dataset called "1_Diversity_SSR_data.tab" is related to a diversity analysis panel of 254 accessions, including the 150 accessions in this study. The additional file called "2_Diversity_SSR_info.txt" gives information concerning the SSR loci.

The SNP genotyping dataset in "hapmap" format is freely and openly accessed via the same DOI. The dataset called "3_GWAS_SNP_hapmap.txt" is related to a GWAS panel of 170 accessions, including the 150 accessions in this study. The additional file called "5_List_of_ID_SNP.tab" is used to link the array identifier name of the accessions with the identifier name used in this study.

Authors' contributions Anthony Bernard, Teresa Barreneche, and Elisabeth Dirlewanger conceived and supervised the study. Anthony Bernard and Armel Donkpegan implemented the analyses. Anthony

Bernard, Teresa Barreneche, and Elisabeth Dirlewanger interpreted the results. Anthony Bernard wrote the manuscript. Anthony Bernard, Teresa Barreneche, Armel Donkpegan, and Elisabeth Dirlewanger contributed in the discussion of the results. All authors reviewed the manuscript. All authors read and approved the final version of the manuscript.

Funding This work has been mainly funded by the "Région Nouvelle-Aquitaine" in the project "INNOV'noyer," coordinated by the CTIFL, and in partnership with the INRAE of Bordeaux. This work has been also funded by the "Cifre" convention number 2016/1558 of ANRT (Agence Nationale de la Recherche et de la Technologie). INRAE funded part of the PhD grant of Anthony Bernard.

Compliance with ethical standards

Competing interests The authors declare that they have no competing interests.

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Arab MM, Marrano A, Abdollahi-Arpanahi R et al (2019) Genome-wide patterns of population structure and association mapping of nut-related traits in Persian walnut populations from Iran using the Axiom *J. regia* 700K SNP array. Sci Rep 9:6376

Bernard A, Lheureux F, Dirlewanger E (2018a) Walnut: past and future of genetic improvement. Tree Genet Genomes 14:1

Bernard A, Barreneche T, Lheureux F, Dirlewanger E (2018b) Analysis of genetic diversity and structure in a worldwide walnut (*Juglans regia* L.) germplasm using SSR markers. PLoS One 13(11): e0208021

Bernard A, Barreneche T, Delmas M, Durand S, Pommier C, Lheureux F, Tranchand E, Naudin M, Dirlewanger E (2019) The walnut genetic resources of INRA: chronological phenotypic data and ontology. BMC Res Notes 12:662

Bernard A, Marrano A, Donkpegan A, Brown PJ, Leslie CA, Neale DB, Lheureux F, Dirlewanger E (2020) Association and linkage mapping to unravel genetic architecture of phenological traits and lateral bearing in Persian walnut (*Juglans regia* L.). BMC Genomics 21: 203

Bretting PK, Widrlechner MP (1995) Genetic markers and horticultural germplasm management. HortScience 30:1349–1356

Chen W, Hou L, Zhang Z, Pang X, Li Y (2017) Genetic diversity, population structure, and linkage disequilibrium of a core collection of *Ziziphus jujuba* assessed with genome-wide SNPs developed by genotyping-by-sequencing and SSR markers. Front Plant Sci 8:575

Corak KE, Ellison SL, Simon PW, Spooner DM, Dawson JC (2019) Comparison of representative and custom methods of generating core subsets of a carrot germplasm collection. Crop Sci 59:1107–1121

Dang M, Zhang T, Hu Y, Zhou H, Woeste K, Zhao P (2016) *De novo* assembly and characterization of bud, leaf and flowers transcriptome from *Juglans regia* L. for the identification and characterization of new EST-SSRs. Forests 7:247–263

Dangl GS, Woeste K, Aradhya MK, Koehmstedt A, Simon C, Potter D, Leslie CA, McGranahan G (2005) Characterization of 14 microsatellite markers for genetic analysis and cultivar identification of walnut. J Am Soc Hortic Sci 130:348–354

De Beukelaer H, Davenport GF, Fack V (2018) Core Hunter 3: flexible core subset selection. BMC Bioinformatics 19:203

Desalegne BA, Dagne K, Melaku G, Ousmane B, Fatokun CA (2017) Efficiency of SNP and SSR-based analysis of genetic diversity, population structure, and relationships among cowpea (*Vigna unguiculata* (L.) Walp.) germplasm from East Africa and IITA inbred lines. J Crop Sci Biotechnol 20:107–128

Earl DA, von Holdt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4:359–361

Ebrahimi A, Zarei A, Lawson S, Woeste KE, Smulders MJM (2016) Genetic diversity and genetic structure of Persian walnut (*Juglans regia*) accessions from 14 European, African, and Asian countries using SSR markers. Tree Genet Genomes 12:114

Emanuelli F, Lorenzi S, Grzeskowiak L, Catalano V, Stefanini M, Troggio M, Myles S, Martinez-Zapater JM, Zyprian E, Moreira FM, Grando MS (2013) Genetic diversity and population structure assessed by SSR and SNP markers in a large germplasm collection of grape. BMC Plant Biol 13:39

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. Mol Ecol 14:2611–2620

Filippi CV, Aguirre N, Rivas JG, Zubrzycki J, Puebla A, Cordes D, Moreno MV, Fusari CM, Alvarez D, Heinz RA, Hopp HE, Paniego NB, Lia VV (2015) Population structure and genetic diversity characterization of a sunflower association mapping population using SSR and SNP markers. BMC Plant Biol 15:52

Frichot E, François O (2015) LEA: An R package for landscape and ecological association studies. Methods Ecol Evol 6:925–929

Frichot E, Mathieu F, Trouillon T, Bouchard G, François O (2014) Fast and efficient estimation of individual ancestry coefficients. Genetics 196:973–983

Fu YB (2017) The vulnerability of plant genetic resources conserved ex situ. Crop Sci 57:2314–2328

Guichoux E, Lagache L, Wagner S et al (2011) Current trends in microsatellite genotyping. Mol Ecol Resour 11:591–611

Hamblin MT, Warburton ML, Buckler ES (2007) Empirical comparison of simple sequence repeats and single nucleotide polymorphisms in assessment of maize diversity and relatedness. PLoS One 2(12): e1367

Hammer K, Teklu Y (2008) Plant genetic resources: selected issues from genetic erosion to genetic engineering. J Agr Rural Dev Trop 109: 15–50

Jia J, Li H, Zhang X, Li Z, Qiu L (2017) Genomics-based plant germplasm research (GPGR). Crop J 5:166–174

Laval G, San Cristobal M, Chevalet C (2002) Measuring genetic distances between breeds: use of some distances in various short term evolution models. Genet Sel Evol 34:481–507

Larsen B, Gardner K, Pedersen C, Ørgaard M, Migicovsky Z, Myles S, Toldam-Andersen TB (2018) Population structure, relatedness and ploidy levels in an apple gene bank revealed through genotyping-by-sequencing. PLoS One 13(8):e0201889

Ljungqvist M, Åkesson M, Hansson B (2010) Do microsatellites reflect genome-wide genetic diversity in natural populations? A comment on. Mol Ecol 19:851–855

Marrano A, Martínez-García PJ, Bianco L, Sideli GM, di Pierro EA, Leslie CA, Stevens KA, Crepeau MW, Troggio M, Langley CH, Neale DB (2019a) A new genomic tool for walnut (*Juglans regia* L.): development and validation of the high-density Axiom™ *J. regia* 700 K SNP genotyping array. Plant Biotechnol J 17: 1027–1036

Marrano A, Sideli GM, Leslie CA, Cheng H, Neale DB (2019b) Deciphering of the genetic control of phenology, yield, and pellicle color in Persian walnut (*Juglans regia* L.). Front. Plant Sci 10:1140

Maxted N, Ford-Lloyd BV, Hawkes JG (1997) Complementary conservation strategies. In: Maxted N, Ford-Lloyd BV, Hawkes JG (eds) Plant genetic resources: the in situ approach. Chapman and Hall, London, pp 15–40

Müller BSF, Pappas GJ, Valdisser PAMR et al (2015) An operational SNP panel integrated to SSR marker for the assessment of genetic diversity and population structure of the common bean. Plant Mol Biol Report 33:1697–1711

Odong T, Jansen J, Van Eeuwijk F, van Hintum TJ (2013) Quality of core collections for effective utilisation of genetic resources review, discussion and interpretation. Theor Appl Genet 126:289–305

Ogwu MC, Osawaru ME, Ahana CM (2014) Challenges in conserving and utilizing plant genetic resources (PGR). Int J Genet Mol Biol 6: 16–22

Perrier X, Flori A, Bonnot F (2003) Data analysis methods. In: Hamon P, Seguin M, Perrier X, Glaszmann JC (eds) Genetic diversity of cultivated tropical plants. Science Publishers, Enfield, pp 43–76

Perrier X, Jacquemoud-Collet J (2006) DARwin software. Available from: http://darwin.cirad.fr/. Accessed 1 Nov 2019

Pollegioni P, Woeste KE, Chiocchini F, Olimpieri I, Tortolano V, Clark J, Hemery GE, Mapelli S, Malvolti ME (2014) Landscape genetics of Persian walnut (*Juglans regia* L.) across its Asian range. Tree Genet Genomes 10:1027–1043

Price AL, Weale ME, Patterson N, Myers SR, Need AC, Shianna KV, Ge D, Rotter JI, Torres E, Taylor KD, Goldstein DB, Reich D (2008) Long-range LD can confound genome scans in admixed populations. Am J Hum Genet 83:132–135

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Singh N, Choudhury DR, Singh AK, Kumar S, Srinivasan K, Tyagi RK, Singh NK, Singh R (2013) Comparison of SSR and SNP markers in estimation of genetic diversity and population structure of Indian rice varieties. PLoS One 8(12):e84136

Tsykun T, Rellstab C, Dutech C, Sipos G, Prospero S (2017) Comparative assessment of SSR and SNP markers for inferring the population genetic structure of the common fungus *Armillaria cepistipes*. Heredity 119:371–380

Ulukan H (2011) The use of plant genetic resources and biodiversity in classical plant breeding. Acta Agric Scand Sect B 61:97–104

Van Inghelandt D, Melchinger AE, Lebreton C, Stich B (2010) Population structure and genetic diversity in a commercial maize breeding program assessed with SSR and SNP markers. Theor Appl Genet 120:1289–1299

Vargas AM, de Andrés MT, Ibáñez J (2016) Maximization of minority classes in core collections designed for association studies. Tree Genet Genomes 12:28

Wambugu PW, Ndjiondjop MN, Henry RJ (2018) Role of genomics in promoting the utilization of plant genetic resources in genebanks. Brief Funct Genomics 17:198–206

Woeste K, Burns R, Rhodes O, Michler C (2002) Thirty polymorphic nuclear microsatellite loci from black walnut. J Hered 93:58–60

Yang X, Xu Y, Shah T, Li H, Han Z, Li J, Yan J (2011) Comparison of SSRs and SNPs in assessment of genetic relatedness in maize. Genetica 139:1045–1054

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Conclusion et Perspectives

Conclusion et Perspectives

L'amélioration des plantes cultivées est probablement un des domaines de la recherche agronomique faisant appel au plus large panel de compétences. C'est un domaine d'une grande complexité au carrefour des génétiques (population et quantitative), des « omics », de la biologie moléculaire, de la physiologie et de la phytopathologie. Mais le·la chercheur·euse en amélioration des plantes doit aussi connaître les techniques physiques et chimiques permettant de caractériser son matériel végétal ainsi que les interactions de ce matériel avec son environnement. Ce domaine revêt d'une importance cruciale pour l'autosuffisance alimentaire, à toutes échelles, car il vise à créer les variétés résilientes de demain, quel que soit le type d'agriculture (biologique, raisonnée ou conventionnelle) face au contexte global de changement climatique.

Synthèse des résultats obtenus

Cette présente thèse, portant sur l'étude des ressources génétiques du noyer en vue d'apporter les outils nécessaires à une sélection assistée par marqueurs, fait partie intégrante d'un travail crucial en amont de l'amélioration de cette espèce végétale. Elle permet d'abord d'apporter un grand nombre de nouvelles connaissances fondamentales sur le noyer concernant d'une part, l'analyse de la diversité génétique de la collection de ressources génétiques de l'INRAE de Nouvelle-Aquitaine-Bordaux et d'autre part, la caractérisation des déterminismes génétiques des principaux caractères d'intérêt agronomique de la fleur au fruit. Ensuite, elle apporte des connaissances complémentaires utiles pour une recherche plus appliquée, telles que la création d'une core collection et d'une riche base de données phénotypiques. Avant de conclure et d'apporter des perspectives pour une suite éventuelle à ce travail, voici un florilège des résultats obtenus :

1. Caractérisation phénotypique de la collection :
   1.1. Un panel de 170 accessions *J. regia* de la collection a été évalué pour la phénologie, durant les deux saisons 2018 et 2019 ; la saison 2020 ayant été annulée à cause de la pandémie de Covid-19. Différents stades phénologiques ont été observés, notamment le débourrement, le début/la pleine/la fin de floraison mâle, le début/la pleine/la fin de floraison femelle. Une grande variabilité phénotypique a été mise en évidence, par exemple pour la date de débourrement, avec 43 jours d'écart en 2018 entre l'accession la plus précoce et l'accession la plus tardive. Si la phénologie est connue

pour être sensible à l'environnement, l'héritabilité au sens large reste néanmoins très forte pour l'ensemble des stades. Une ACP a montré qu'il ne serait peut-être pas nécessaire de phénotyper tous les stades dans le cadre de l'évaluation en routine de la collection, mais seulement le début de chaque floraison, puisque les trois stades sont très fortement correlés.

1.2. Sur le même panel, 25 caractères liés au fruit ont été également observés. Pour n'en citer que quelques uns : le diamètre de face de la noix, le poids de 100 noix, le volume du cerneau, l'épaisseur de la coque, la teneur en acides gras unsaturés et la force nécessaire pour rompre la coque sur la face et la suture. Ces différents caractères ont été mesurés durant les saisons 2017 à 2019 à l'aide de nombreuses techniques physiques et chimiques. Ainsi, par exemple, les deux forces de cassage ont été mesurées à l'aide d'un texturomètre durant les trois saisons (grâce aux travaux de Nicolas Jammes, stagiaire IUT), alors que le volume du cerneau a été apprécié uniquement en 2018, via l'utilisation de la tomographie à rayons X. Comme pour la phénologie, la collection a montré une très grande variabilité phénotypique et de fortes valeurs d'héritabilité au sens large.

1.3. En parallèle, les données des archives d'Eric Germain viennent compléter ce phénotypage large en apportant des données plus anciennes sur certaines accessions. Elles ont notamment permis de mettre en avant une avancée de la date de débourrement au cours des 30 dernières années pour les deux variétés témoins 'Franquette' et 'Lara'.

- L'ensemble de ces données phénotypiques a permis, d'une part, de mener à bien des travaux de génétique d'association pour tous les caractères, puis d'autre part, de prodiguer au CTIFL une « carte d'identité » précise de chaque potentiel·le géniteur·trice, dans l'éventualité d'une mise en place d'un futur programme de création variétale.

2. Diversité génétique de la collection :

2.1. La diversité génétique de la collection *Juglans* a été étudiée et sa structure appréhendée via l'utilisation de 13 marqueurs microsatellites séléctionnés avec soin dans la littérature. Les 253 accessions étudiées comprennent 217 accessions *J. regia* et de nombreuses autres accessions des espèces apparentées au noyer cultivé. D'abord, une ACoP globale a bien séparé les espèces selon la classification

botanique. Ensuite, sur le groupe des 217 accessions *J. regia*, les marqueurs microsatellites ont montré un haut niveau de polymorphisme, avec 8.92 allèles en moyenne par locus. Ces données alléliques ont permis d'identifier deux grands groupes principaux de structure, en lien avec l'origine géographique des accessions. Ainsi, il est possible de distinguer les accessions d'Europe de l'Ouest et d'Amérique, de celles d'Europe de l'Est et d'Asie.

2.2. Les groupes de structure obtenus sont en accord avec les résultats de l'arbre phylogénétique construit et l'ACoP réalisée sur ces 217 accessions *J. regia*. En allant un peu plus loin, il a été possible de distinguer, chez les accessions françaises numériquement majoritaires, deux sous-groupes qui correspondent aux deux principaux bassins de production de la noix : Sud-Est et Sud-Ouest. Enfin, une première core collection a été proposée, basée sur la conservation des accessions les plus diverses génétiquement, et celles qui présentent un intérêt phénotypique, comme le cerneau rouge ou un débourrement extrêmement précoce.

- L'étude de la diversité génétique a permis d'une part de définir le « panel de phénotypage » plus restreint, décrit plus haut, pour mener à bien les travaux de génétique d'assocation, puis d'autre part de prodiguer au CTIFL les connaissances de la structure de la collection, facilitant ainsi le choix de potentiel·le·s géniteurs·trices dans l'éventualité d'une mise en place d'un futur programme de création variétale.

3. Déterminisme génétique des principaux caractères d'intérêt agronomique :

3.1. Concernant les caractères liés à la phénologie, le panel de GWAS a été utilisé conjointement avec une descendance de cartographie, issue du croisement entre 'Franquette' et 'UK 6-2', constituée de 78 individus. Cette descendance ségrège pour ces caractères et en parallèle, un génotypage dense des individus a été réalisé à l'aide de la puce de génotypage Axiom™ *J. regia* 700K SNP. Après réalisation du « filtre qualité », 364 275 SNP ont été retenus pour la GWAS, réalisée à l'aide des BLUPs calculés d'une part sur les deux saisons 2018 et 2019 nouvellement acquises, et calculés d'autre part sur les données des archives d'Eric Germain. Deux loci majeurs contrôlant la phénologie et le type de fructification ont été mis en évidence respectivement sur le début du chromosome 1 et le milieu du chromosome 11, confirmés par l'équipe de l'Université de Davis en Californie.

3.2. Concernant les caractères liés au fruit, seul le panel de GWAS a été utilisé et de nombreuses associations ont été obtenues. Un locus majeur a été identifié sur le chromosome 14 en lien avec la taille globale de la noix (diamètre de face, volume de la noix totale et volume du cerneau) et 5 associations significatives ont été détectées pour le poids de 100 noix (grâce aux travaux de Julie Crabier, stagiaire Master 2). Des gènes candidats ont été proposés, notamment un gène codant pour une beta-galactosidase, enzyme connue pour être impliquée dans le processus de développement du fruit chez d'autres espèces. Enfin, les zones du génome en lien avec la force nécessaire pour rompre la noix sur la suture identifiées par l'Université de Davis, sur les chromosomes 5 et 11, ont été confirmées sur d'autres matériels (Figure 25).

- Ce travail sur le déterminisme génétique des principaux caractères d'intérêt agronomique a permis d'une part de mettre au point pour la première fois sur noix, un marqueur de type KASP lié à la phénologie, puis d'autre part de prodiguer au CTIFL l'information nécessaire au développement de nouveaux marqueurs pour tout un éventail de caractères, de la fleur au fruit, dans l'éventualité d'une mise en place d'un futur programme de création variétale.

Cette sélection de résultats illustre bien l'intérêt de la complémentarité du laboratoire et de l'entreprise dans une thèse Cifre. Ces résultats ont été publiés (ou sont en cours) dans des journaux à comité de lecture. A ce jour, trois articles ont été publiés, deux autres ont été soumis et un sixième est en préparation. En parallèle, des outils concrets ont été apportés pour l'entreprise, comme une core collection largement caractérisée qui est en cours d'implantation (en tunnel et en pots pour l'instant au CTIFL), puis des informations pour l'obtention de marqueurs moléculaires pour une éventuelle future sélection assistée par marqueurs.

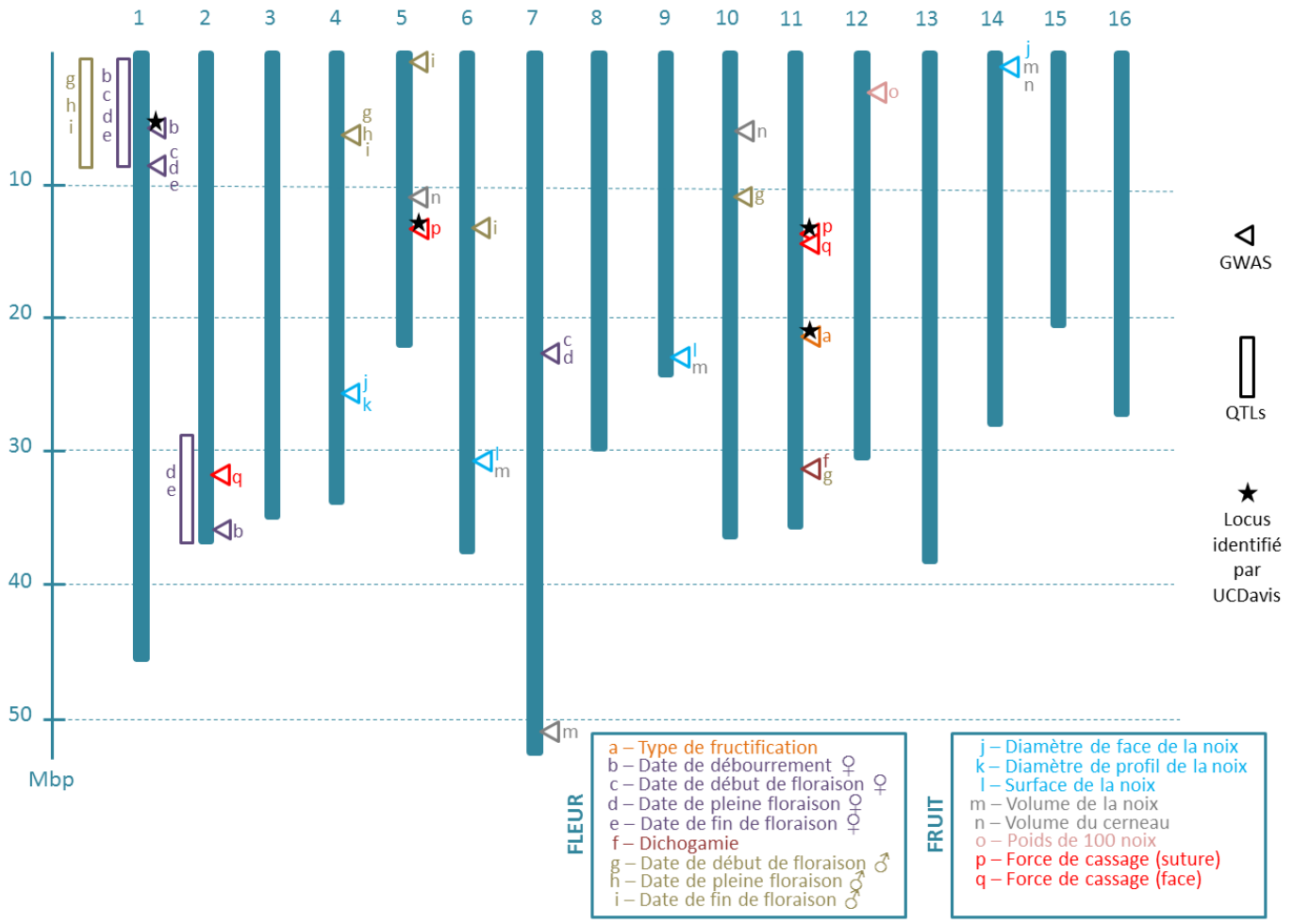


Figure 25 : Schéma de synthèse des loci majeurs identifiés dans l'architecture génétique des caractères d'intérêt agronomiques du noyer.

Discussion concernant les stratégies adoptées

Comme tout travail, cette thèse s'est concretisée par des choix stratégiques qui ont dû être fait tout au long des trois années. Certains se sont révélés être les bons puis d'autres sont plus discutables. Parmi ces choix discutables, il y a notamment a) la mise de côté du travail des questions « pathologie » et b) le choix du panel de GWAS :

a) La pathologie est un domaine large en soi et nécessite un travail de longue haleine pour trouver des sources de tolérances. Pour rappel, en dehors de la mouche du brou, trois pathogènes sont responsables de dégâts en vergers, qui peuvent être importants en fonction des années et des environnements : la bactériose due à *Xanthomonas arboricola* pv. *juglandis*, une forme d'anthracnose due à *Gnomonia leptostyla* et une nouvelle forme d'anthracnose due à *Colletotrichum acutatum* sp. La première année de thèse, en parallèle du phénotypage de la fleur au fruit et du génotypage microsatellites, a été consacrée à la recherche de sources de tolérances à *C. acutatum*

au sein de la collection, espèces apparentées incluses (grâce aux travaux de Joris Molina, stagiaire Licence Pro). Cette collection de ressources génétiques étant *in vivo* et *ex situ*, il était indispensable d'avoir recours à l'utilisation des organes détachés. Très vite, des difficultés sont apparues dans la mise en place de tels tests pathologiques sur organes détachés : répétabilité, fiabilité, temps de réalisation, état sanitaire des organes prélevés, etc. Malré tout, des accessions chez les espèces apparentées au noyer cultivé, telles que *J. cinerea* et *J. sieboldiana*, se sont avérées être plus tolérantes. En revanche, au sein de *J. regia*, il n'a pas été possible de déceler des différences dans un premier temps, avec un seul type de test. Les années suivantes, d'autres tests ont été mis en place, en multipliant les méthodes d'inoculation et les organes prélevés (grâce aux travaux de Morgane Vincent, stagiaire apprentie ingénieure). Si quelques résultats encourageants sont à signaler, il n'en demeure pas moins qu'une telle campagne de phénotypage suppose de disposer d'un test parfaitement fiable et d'une équipe importante d'expérimentateurs. Une solution résiderait en la reprise de la collection par greffage et en réalisant une infection en conditions contrôlées. Ainsi, les plants auraient le même âge et il serait possible d'observer l'interaction plante-pathogène à l'échelle de la plante entière. Une expérimentation en serre peut être imaginée, avec des génotypes identiques implantés en plein champ, pour comparer ces deux environnements, respectivement contrôlés ou non. Il serait également intéressant d'effectuer ces tests plusieurs années, pour vérifier les différences qui résident dans cette intéraction plant-pathogène entre des arbres juvéniles et des arbres adultes. La core collection étant actuellement en cours de reprise par greffage au CTIFL en plusieurs exemplaires par sûreté, ce type de test est tout à fait envisageable à l'avenir et pourrait faire l'objet d'un projet à part entière. Mais dans un premier temps, il serait intéressant de réfléchir à la mise en place de croisements interspécifiques entre *J. regia* et *J. cinerea*, puis entre *J. regia* et *J. sieboldiana* pour tenter d'apporter les sources de tolérance au sein du compartiment cultivé.

b) Durant cette thèse, les analyses de GWAS ont été réalisées uniquement sur la base du phénotypage et du génotypage de l'un des deux arbres disponibles pour chaque accession. Un choix contestable est celui de n'avoir phénotypé que l'arbre génotypé et non les deux. Pour information, le phénotypage de la phénologie est une tâche très chronophage puisqu'elle nécessite deux jours par semaine de notations, durant les trois mois de mars, avril et mai. Pour phénotyper un nombre d'arbres multiplié par deux,

cette tâche aurait obligatoirement nécéssité un troisième jour de notations. Mais la situation est bien pire encore pour le phénotypage des caractères liés au fruit. Le noyer cultivé présente une grande variabilité intra-individuelle, nécessitant donc le phénotypage d'un grand nombre de noix. Durant cette thèse, un nombre de 100 noix a notamment été retenu pour les poids et 50 pour les forces de cassage ou pour les caractères phénotypés à l'aide de la tomographie à rayons X. Il est évident que ces mesures ont demandé de nombreuses journées de travail, aussi bien au doctorant, aux stagiaires et même parfois aux encadrant·e·s. Phénotyper un nombre de lots de noix multiplié par deux aurait été trop chronophage. Avec du recul, une solution aurait pû résider dans le fait de diviser par deux le nombre de noix par accession, mais de récolter les noix des deux clones. Ceci aurait éventuellement permis d'engendrer un phénotypage plus précis et donc d'obtenir un effet génotypique plus proche de la réalité et essentiel en GWAS. En revanche, la logistique du ramassage lors de la récolte aurait été complexifiée et probablement source d'erreurs, puisque les canopées sont parfois chevauchantes. En ayant sélectionné un arbre sur deux (les clones étant côte à côte), ce risque est grandement diminué.

Des idées pour le pre-breeding

Le premier objectif consisterait à apporter une tardiveté dans la phénologie, en particulier pour le débourrement et la floraison. Il est primordial de conserver au maximum l'amélioration déjà réalisée sur l'espèce : 'Fernor' est une variété moyenne dans son calibre et sa phénologie, mais elle possède une qualité gustative appréciée du consommateur et fructifie latéralement. 'Fernor' pourrait être croisée avec les accessions les plus tardives de la collection comme 'St Martial', 'Le Bordelais', 'Lalande', 'Maribor' ou 'Culplat'. Les hybrides T en rouge comme « tardifs » seront obtenus (Figure 26).

Objectif 1 : apporter une tardiveté dans la phénologie

Exemples :
Fernor ✕ St Martial (24)
Fernor ✕ Le Bordelais (125)
Fernor ✕ Lalande (116)
Fernor ✕ Maribor (143)
Fernor ✕ Culplat (119)

→ **Hybrides T**

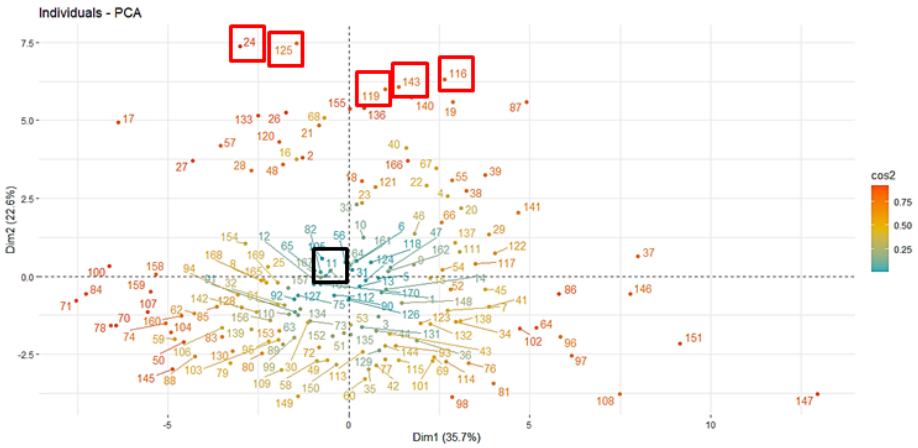


Figure 26 : Idées de croisements pour apporter une tardiveté dans la phénologie.

Le deuxième objectif, qui pourrait être mené en parallèle, consisterait à apporter un calibre plus important. 'Fernor' pourrait être croisée avec les accessions donnant les fruits de plus gros calibre comme 'Carmelo', 'Germisara', 'Sibisel 44', 'Sibisel 39' ou 'UK 239-10'. Les hybrides C en violet comme « calibre » seront obtenus (Figure 27).

Objectif 2 : apporter un calibre plus important

Exemples :
Fernor ✕ Carmelo (147)
Fernor ✕ Germisara (151)
Fernor ✕ Sibisel 44 (146)
Fernor ✕ Sibisel 39 (37)
Fernor ✕ UK 239-10 (108)

→ **Hybrides C**

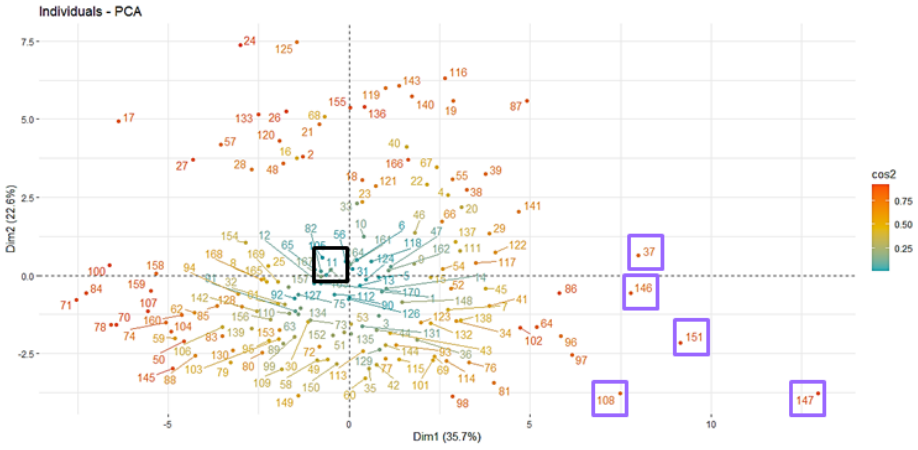


Figure 27 : Idées de croisements pour apporter un calibre plus important.

Attention toutefois, une dépression de consanguinité a pu déjà être observée lors du premier programme d'amélioration d'Éric Germain, entre les variétés françaises et américaines qui partagent un certain niveau de parenté, donc il est important de prendre en compte la structure. Ici, ces idées de croisements sont intéressantes puisque 'Fernor' appartient au groupe d'Europe de l'Ouest et d'Amérique, tandis que les autres appartiennent

soit au groupe d'Europe de l'Est et d'Asie, soit au groupe « admixed ». Ces croisements permettraient de surcroît de brasser la diversité génétique disponible.

Ensuite, les objectifs 1 et 2 pourraient être travaillés séparément et des rétrocroisements avec 'Fernor' pourraient être réalisés. 'Fernor' serait considérée comme la variété « élite ». Concernant l'objectif 1 de tardiveté, le marqueur KASP lié à la date de débourrement est disponible pour effectuer la sélection assistée par marqueur. Concernant l'objectif 2 de plus gros calibre, il faudrait rapidement valider le potentiel marqueur du chromosome 14. Néanmoins, il conviendra de garder à l'esprit le nombre de rétrocroisements, puisqu'avec une période juvénile de 6 ans, il faudra au minimum attendre 12 ans pour récupérer 75% de fond génétique « élite ». En parallèle, il pourrait être envisagé de croiser les hydrides des deux objectifs et ainsi espérer obtenir un nouvel hybride intéressant en utilisant les deux marqueurs qui sont par chance sur des chromosomes distincts (Figure 28).

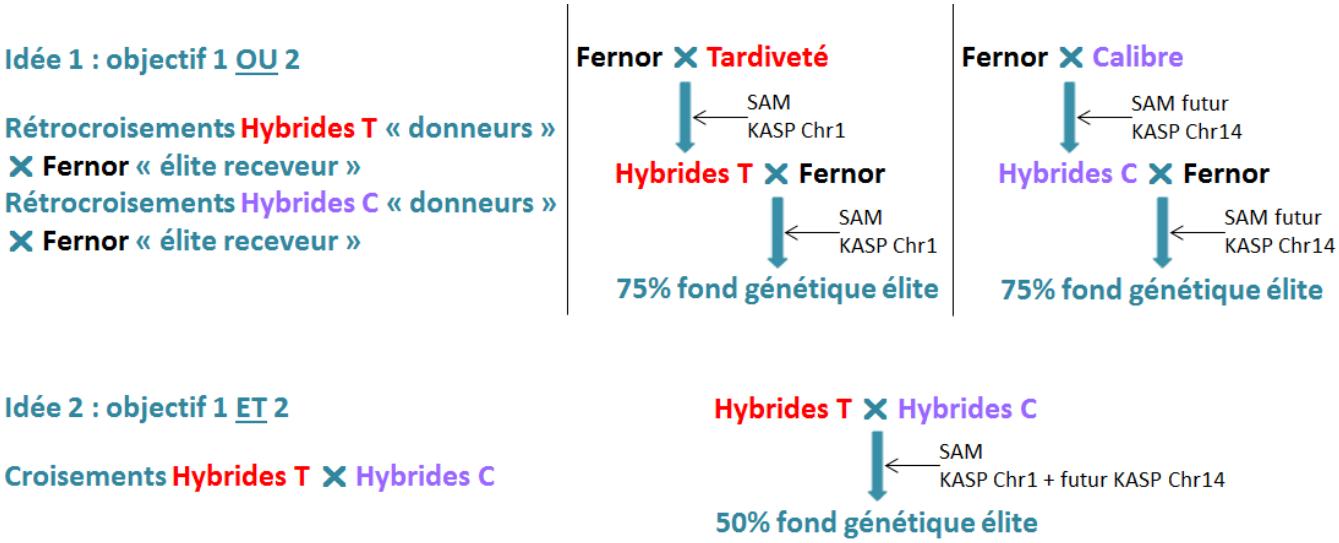


Figure 28 : Idées de rétrocroisements ou de croisements supplémentaires.

Des perspectives envisageables

Si réfléchir à ce qui aurait pû être amélioré durant la thèse représente la première des perspectives, il est tout autant intéressant de proposer de nouveaux travaux à engager suite à cette thèse. Parmi les résultats manquants, il y a l'obtention de plusieurs marqueurs pour la sélection. En effet, un seul marqueur de type KASP a été développé puis validé grâce à du matériel végétal en provenance de l'Université de Davis en Californie. Néanmoins, cette thèse a permis de mettre en évidence de nombreuses associations significatives qu'il faudrait de la même façon transformer en marqueurs validés pour le laboratoire. L'obtention de marqueurs pour l'ensemble des caractères agronomiques est une étape cruciale pour initier un nouveau

programme de création variétale utilisant la sélection assistée par marqueurs, mais cette étape est aussi très longue. Elle nécessite de phénotyper et de génotyper du nouveau matériel végétal. Par chance, il existe dans le monde notamment deux autres collections de ressources génétiques du noyer cultivé et phénotypées pour un grand nombre de caractères. Il y a celle de l'Université de Davis en Californie, principalement phénotypée comme déjà énoncé pour la phénologie, mais aussi pour le rendement, la force nécessaire pour rompre la noix et la couleur du cerneau. En parallèle, celle de l'Université de Téhéran en Iran a été caractérisée pour tous les caractères liés au fruit, en commun avec le travail de cette thèse, comme le poids de la noix, du cerneau, l'épaisseur de la coque, la forme, la rugosité, etc. Ainsi, il est tout à fait envisageable d'instaurer une coopération internationale entre Univerisité de Davis, Université de Téhéran et INRAE/CTIFL dans ce but, afin de partager des données de phénotypage puis de l'ADN. Une coopération entre les trois parties est d'ailleurs déjà en cours, puisqu'à l'initiative de l'Université de Téhéran, les trois collections seront génotypées par génotypage par séquençage, afin d'étudier plus en profondeur la diversité génétique de *J. regia*.

Pour les caractères liés à la teneur en acides gras insaturés (poly- et mono-) puis à l'activité vitaminique E, la quantification n'a été réalisée qu'une seule année. L'héritabilité au sens large n'a donc pas pu être appréciée et la qualité des fruits, en général, est connue pour être dépendante des conditions environnementales durant la période de maturité. Il serait donc essentiel de re-quantifier ces molécules, sur quelques variétés contrastées, pour savoir si ce caractère quantitatif est suffisamment héritable. Les associations significatives trouvées avec une seule année de quantification sont donc à considérer avec précaution, dans le cas éventuel de développement de marqueurs.

Après l'étape de création et de validation des marqueurs sur d'autres matériels végétaux, il serait tout à fait envisageable de passer à une échelle de validation fonctionnelle, en faisant donc appel à des travaux de génomique fonctionnelle. En effet, les nombreux gènes candidats proposés pourraient faire l'objet d'une étude spécifique, en transformant par exemple le noyer et étudier le phénotype des mutants. Contrairement à d'autres plantes ligneuses pour lesquelles la transformation génétique n'est pas aisée, le noyer se transforme avec succès via l'utilisation de l'embryogenèse somatique et *Agrobacterium tumefaciens* (McGranahan et al. 1988). Il serait alors possible de comprendre plus finement les mécanismes de régulation génétique. Création/validation de marqueurs (outils de recherche

plus appliquée) puis validation fonctionnelle des gènes (domaine plus fondamental) sont donc des travaux complémentaires, pouvant faire l'objet d'un projet de nouvelle thèse par exemple.

Toujours dans une démarche de recherche plus appliquée, la sélection génomique pourrait se mettre en place relativement rapidement, puisque le génotypage dense des potentiel·le·s géniteurs·trices est d'ores et déjà disponible. Si la sélection assistée par marqueurs semble être plus raisonnable dans un futur proche pour des caractères hautement héritables et dont les loci majeurs sont validés, la sélection génomique mérite toute notre attention. Il s'agit évidemment d'une méthode de sélection plus coûteuse puisqu'il faudra génotyper, avec la puce SNP par exemple, l'ensemble des hybrides obtenus. Une idée serait de conduire les deux types de sélection en parallèle. La sélection assistée par marqueurs peut s'envisager pour des caractères simples, comme le type de fructification, ou pour d'autres caractères dont des loci majeurs ont été détectés, comme la date de débourrement. La sélection génomique pourrait être pratiquée pour des caractères hautement quantitatifs pour lesquels aucun locus n'expliquant une très grande variation n'a été détecté (un seuil de 10% pourrait être choisi) ; cette sélection basée sur la valeur génétique plus globale des hybrides serait beaucoup plus coûteuse mais réalisée sur un nombre restreint d'hybrides. Un travail important réside dans la constitution de la population d'entraînement nécessaire au développement de la sélection génomique. Différentes lignées pourraient être développées pour ces caractères complexes en utilisant des géniteurs·trices du panel GWAS.

Les travaux réalisés durant cette thèse sur la phénologie restent assez sommaires puisqu'ils ont consisté en l'observation des dates de débourrements et des floraisons mâle et femelle. L'adaptation au changement climatique passera certainement par la proposition de variétés ayant de plus faibles besoins en froid durant nos hivers qui ont tendance à se réchauffer, puis ayant de plus forts besoins en chaud pour éviter les gelées printanières. En parallèle, ces variétés devront posséder de plus grandes aptitudes à la sécheresse, durant les étés qui deviennent moins pluvieux. Il serait donc intéressant d'étudier les besoins en froid et en chaud de la collection, puis l'aptitude à la résistance à la sécheresse des différents porte-greffes disponibles. L'équipe Micro-Environnement et Arbre de l'UMR PIAF (INRAE de Clermont-Auvergne-Rhône-Alpes) a longtemps travaillé sur les questions de l'adaptation hivernale des noyers, comme expliqué dans le chapitre 1. En parallèle, l'équipe A3C applique couramment des protocoles d'estimation des besoins en froid et en chaud du cerisier, à l'échelle de plusieurs cultivars. Les connaissances sont donc disponibles pour mettre en place, facilement et rapidement, de telles expériences sur la collection de noyers. Du côté des porte-

greffes, le matériel végétal est présent au CTIFL et il pourrait être envisagé d'étudier le potentiel hydrique de ce matériel, avec les phénomènes de transpiration et de cavitation.

Le choix des accessions pour de futurs croisements pourra être guidé par les résultats de l'ACP globale réalisée sur l'ensemble des données de phénotypage. En effet, cette ACP a permis de mettre en évidence deux dimensions majeures : la première expliquée principalement par les caractères liés à la noix sèche et la seconde par ceux liés à la phénologie. Les résultats de cette ACP permettront très rapidement de sélectionner des accessions contrastées, comme par exemple, un parent donnant des noix de beau calibre, puis un autre parent possédant une phénologie tardive.

Si la collection de ressources génétiques de l'INRAE de Nouvelle-Aquitaine-Bordeaux est d'une grande richesse, il ne faut pas s'arrêter à ce matériel. Eric Germain l'avait compris et a rapporté de nombreuses accessions durant ses longues années dédiées à la prospection de matériel végétal intéressant. Ce travail nécessiterait d'être poursuivi d'une part, par un échange plus important des variétés traditionnelles et/ou développées par les différents pays européens partageant des similarités de climat et d'autre part, en allant prospecter du matériel dans les régions de domestication de *J. regia*, soit l'Asie Centrale.

A ce sujet, une étude toute récente parue le 23 juin 2020 s'est concentrée sur l'évaluation de la diversité génétique des populations de noyer cultivées et sauvages de l'Ouzbékistan, le Tadjikistan et le Kirghizistan, pays actuels du centre d'origine. En parallèle, une évaluation des menaces qui pèsent sur ces ressources génétiques a été conduite (Gaisberger et al. 2020). Les auteurs·trices montrent que les menaces à court terme, comme la surexploitation et le surpâturage, sont les plus alarmantes et touchent les zones qui possèdent également une grande richesse allélique. Par exemple, les populations de la région de Sariosiyo, au sud-est de l'Ouzbékistan, sont considérées comme zones prioritaires de conservation. L'équipe préconise alors une conservation *ex situ*, pour tout simplement sauver le matériel végétal.

Cet article tout juste publié vient, encore une fois, appuyer l'importance de protéger nos espèces et nos espaces. L'importance des ressouces génétiques, une des formes de la biodiversité, est incontestable pour le futur de nos systèmes de production agricole.

Références bibliographiques

Abedi B, Parvaneh T (2016) Study of Correlations between Horticultural Traits and Variables Affecting Kernel Percentage of Walnut (*Juglans regia* L.). J Nuts 7:35–44.

Adaskaveg JE, Förster H (2000) Occurrence and management of anthracnose epidemics caused by *Colletotrichum* species on tree fruit crops in California. In: Prusky D, Freeman S, Dickman MB (eds) *Colletotrichum*: Host Specificity, Pathology, and Host-Pathogen Interactions. American Phytopathological Society, St. Paul, MN. pp 317–336.

Ahandani EA, Ramandi HD, Sarmad J, et al (2014) Evaluation of Morphological Diversity among Some Persian Walnut Accessions (*Juglans regia* L.) in Guilan, Northern Iran. Int J Plant Biol Res 2:1015–1022.

Akça Y (2014) 'Şebin' Persian Walnut. Acta Hortic 1050:139–142.

Akça Y, Ozongun S (2004) Selection of late leafing, late flowering, laterally fruitful walnut (*Juglans regia*) types in Turkey. N Z J Crop Hortic Sci 32:337–342.

Akça Y, Polat AA (2007) Present Status and Future of Walnut Production in Turkey. Eur J Plant Sci Biotechnol 1:57–64.

Akça Y, Sen SM (1997) The relationship between dichogamy and yield-nut characteristics in *Juglans regia* L. Acta Hortic 442:215–216.

Akça Y, Sütyemez M, Yilmaz S, Karadag H (2016) The new walnut variety breeding program in Turkey. VIIth International Scientific Agricultural Symposium. Jahorina, Bosnia and Herzegovina, pp 461–466.

Aletà N, Olarte C, Truco MJ, Arus P (1990) Identification of walnut cultivars by isozyme analysis. Acta Hortic 284:91–96.

Alexandri V (1853) Les Doïnas : poésies moldaves. Impr. de De Soye et Bouchet. Paris.

Ali AM, Zubair SJ, Abbas AM, Jubrael JMS (2016) Genetic diversity among Walnuts (*Juglans regia*) population in Kurdistan Region–Iraq using AFLP-PCR. ZANCO J Pure Appl Sci 28:50–55.

Alonso-Blanco C, El-Assal SE, Coupland G, Koornneef M (1998) Analysis of natural allelic variation at flowering time loci in the Landsberg erecta and Cape Verde Islands ecotypes of *Arabidopsis thaliana*. Genetics 149:749–764.

Aly MA, Fjellstrom RG, McGranahan GH, Parfitt DE (1992) Origin of Walnut Somatic Embryos Determined by RFLP and Isozyme Analysis. HortScience 27:61–63.

Amiri R, Vahdati K, Mohsenipoor S, et al (2010) Correlations between Some Horticultural Traits in Walnut. HortScience 45:1690–1694.

Ampatzidis Y, Partel V (2019) UAV-Based High Throughput Phenotyping in Citrus Utilizing Multispectral Imaging and Artificial Intelligence. Remote Sensing 11:410.

Andersen JR, Lübberstedt T (2003) Functional markers in plants. Trends Plant Sci 8:554–560.

Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT (2010) Data quality control in genetic case-control association studies. Nat Protoc 5:1564–1573.

Arab MM, Marrano A, Abdollahi-Arpanahi R, et al (2019) Genome-wide patterns of population structure and association mapping of nut-related traits in Persian walnut populations from Iran using the Axiom *J. regia* 700K SNP array. Sci Rep 9:6376.

Arab MM, Marrano A, Abdollahi-Arpanahi R, Leslie CA, Cheng H, et al (2020) Combining phenotype, genotype and environment to uncover genetic components underlying water use efficiency in Persian walnut. J Exp Bot 71:1107–1127.

Aradhya M, Woeste K, Velasco D (2009) Genetic Diversity, Structure and Differentiation in Cultivated Walnut (*Juglans regia* L.). Acta Hortic 861:127–132.

Aradhya MK, Potter D, Gao F, Simon CJ (2007) Molecular phylogeny of *Juglans* (*Juglandaceae*) : a biogeographic perspective. Tree Genet Genomes 3:363–378.

Aradhya MK, Potter D, Simon CJ (2006) Cladistic Biogeography of *Juglans* (*Juglandaceae*) Based on Chloroplast DNA Intergenic Spacer Sequences. In: Motley TJ, Zerega N, Cross H (eds) Darwin's Harvest: New Approaches to the Origins, Evolution, and Conservation of Crops, Columbia University Press. pp 143–170.

Aradhya MK, Velasco D, Wang JR, Ramasamy R, You FM, et al (2019) A fine-scale genetic linkage map reveals genomic regions associated with economic traits in walnut (*Juglans regia*). Plant Breed 00:1–12.

Araji S, Grammer TA, Gertzen R, et al (2014) Novel roles for the polyphenol oxidase enzyme in secondary metabolism and the regulation of cell death in walnut. Plant Physiol 164:1191–1203.

Aranzana MJ, Kim S, Zhao K, Bakker E, Horton M, et al (2005) Genome-wide association mapping in *Arabidopsis* identifies previously known flowering time and pathogen resistance genes. PLoS Genet 1:e60.

Arnaudov V, Gandev S, Dimova M (2015) Susceptibility of Some Walnut Cultivars to *Gnomonia leptostyla* and *Xanthomonas arboricola* pv. *juglandis* in Bulgaria. Agroznanje 15:41–54.

Arulsekar S, McGranahan GH, Parfitt DE (1986) Inheritance of phosphoglucomutase and esterase isozymes in Persian walnut. J Hered 77:220–221.

Arzani K, Mansouri-Ardakan H, Vezvaei A, Roozban MR (2008) Morphological variation among Persian walnut (*Juglans regia*) genotypes from central Iran. N Z J Crop Hortic Sci 36:159–168.

Asadian G, Pieber K (2005) Morphological Variations in Walnut Varieties of the Mediterranean Regions. Int J Agric Biol 7:71–73.
Aslamarz AA, Vahdati K, Rahemi M, Hasani D (2009) Estimation of chilling and heat requirements of some Persian walnut cultivars and genotypes. HortScience 44(3):697–701.
Aslantaş R (2006) Identification of superior walnut (*Juglans regia*) genotypes in north-eastern Anatolia, Turkey. N Z J Crop Hortic Sci 34:231–237.
Asma BM (2012) Pomological and phenological characterization of promising walnut (*Juglans regia* L.) genotypes from Malatya, Turkey. Acta Sci Pol Horturum Cultus 11:169–178.
Astle W, Balding DJ (2009) Population Structure and Cryptic Relatedness in Genetic Association Studies. Statistical Science 24:451–471.
Atefi J (1990) Preliminary research of Persian walnut and correlation between pair characters. Acta Hortic 284:97–104.
Austin L, Kumar R, Kousar B, Lampadaris CH, Lucas MM (2018) Tomography of a walnut. Technical Report. DOI:10.13140/RG.2.2.25130.52161.
Bai W-N, Liao W-J, Zhang D-Y (2010) Nuclear and chloroplast DNA phylogeography reveal two refuge areas with asymmetrical gene flow in a temperate walnut tree from East Asia. New Phytol 188:892–901.
Bandi A, Hevesi M, Szani Z, Tóth M (2015) Assessment of Bacterial Blight Tolerance of Persian Walnut Based on Immature Nut Test. Not Sci Biol 7:253–257.
Bandillo N, Raghavan C, Muyco PA, et al (2013) Multi-parent advanced generation inter-cross (MAGIC) populations in rice: progress and potential for genetics research and breeding. Rice 6:11.
Banel DK, Hu FB (2009) Effects of walnut consumption on blood lipids and other cardiovascular risk factors: a meta-analysis and systematic review. Am J Clin Nutr 90:56–63.
Banta JA, Richards CL (2018) Quantitative epigenetics and evolution. Heredity 121:210–224.
Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21:263–265.
Bates D, Maechler M, Bolker B, Walker S (2015) Fitting Linear Mixed-Effects Models Using lme4. J Stat Softw 67:1–48.
Bayazit S (2012) Determination of relationships among kernel percentage and yield characteristics in some Turkish walnut genotypes by correlation and path analysis. J Anim Plant Sci 22:513–517.
Bayazit S, Kazan K, Gülbitti S, et al (2007) AFLP analysis of genetic diversity in low chill requiring walnut (*Juglans regia* L.) genotypes from Hatay, Turkey. Sci Hortic 111:394–398.
Becraft PW (2002) Receptor Kinase Signaling in Plant Development. Annu Rev Cell Dev Biol 18:163–192.
Behnke HD (1984) Plant trichomes – Structure and ultrastructure: General terminology, taxonomic applications, and aspects of trichome-bacteria interaction in leaf tips of Dioscorea. In: Rodriguez E, Healey PL, Mehta I (eds) Biology and chemistry of plant trichomes. Plenum, New York, pp 1–21.
Belisario A, Maccaroni M, Corazza L (2002) Occurrence and Etiology of Brown Apical Necrosis on Persian (English) Walnut Fruit. Plant Disease 86:599–602.
Belisario A, Scotton M, Santori A, Onofri S (2008) Variability in the Italian population of *Gnomonia leptostyla*, homothallism and resistance of *Juglans* species to anthracnose. For Pathol 38:129–145.
Belisario A, Zoina A, Pezza L, Luongo L (1999) Susceptibility of species of *Juglans* to pathovars of *Xanthomonas campestris*. Eur J Plant Pathol 29:75–80.
Beló A, Zheng P, Luck S, Shen B, Meyer D, et al (2008) Whole genome scan detects an allelic variant of *fad2* associated with increased oleic acid levels in maize. Mol Genet Genomics 279:1–10.
Ben Bâaziz K, Lopez D, Bouzid S, et al (2012) Early gene expression in the walnut tree occurring during stimulation of leaf hydraulic conductance by irradiance. Biol Plant 56:657–666.
Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: A practical and powerful approach to multiple testing. J R Statist Soc B 57:289–300.
Bernard A, Barreneche T, Lheureux F, Dirlewanger E (2018a) Analysis of genetic diversity and structure in a worldwide walnut (*Juglans regia* L.) germplasm using SSR markers. PLoS ONE 13(11):e0208021.
Bernard A, Lheureux F, Dirlewanger E (2018b) Walnut: Past and Future of Genetic Improvement. Tree Genet Genomes 14:1.
Bernard A, Barreneche T, Delmas M, Durand S, Pommier C, et al (2019) The walnut genetic resources of INRA: chronological phenotypic data and ontology. BMC Res Notes 12:662.
Bernard A, Marrano A, Donkpegan A, Brown PJ, Leslie CA, et al (2020) Association and linkage mapping to unravel genetic architecture of phenological traits and lateral bearing in Persian walnut (*Juglans regia* L.). BMC Genomics 21:203.
Beucher S, Meyer F (1993) The morphological approach to segmentation: the watershed transformation. In: Dougherty ER (ed). Mathematical Morphology in Image Processing. pp 433–482.
Bijelic A, Pretzler M, Molitor C, Zekiri, F, Rompel A (2015) The Structure of a Plant Tyrosinase from Walnut Leaves Reveals the Importance of "Substrate-Guiding Residues" for Enzymatic Specificity. Angew Chem Int Ed Engl. 54:14677–14680.
Biskup B, Scharr H, Schurr U, Rascher U (2007) A stereo imaging system for measuring structural parameters of plant canopies. Plant Cell Environ 30:1299–1308.

Bliss FA (2010) Marker-assisted breeding in horticultural crops. Acta Hortic 859:339–350.
Bolaños-Villegas P, Yang X, Wang HJ, Juan CT, Chuang MH, et al. (2013) Arabidopsis *CHROMOSOME TRANSMISSION FIDELITY 7 (AtCTF7/ECO1)* is required for DNA repair, mitosis and meiosis. Plant J 75:927–940.
Bolling BW, McKay DL, Blumberg JB (2010) The phytochemical composition and antioxidant actions of tree nuts. Asia Pac J Clin Nutr 19:117–123.
Botstein DR, White RL, Skolnick MH, Davis RW (1980) Construction of a Genetic Linkage Map in Man Using Restriction Fragment Length Polymorphisms. Am J Hum Genet 32:314–331.
Brachi B, Faure N, Horton M, Flahauw E, Vazquez A, Nordborg N, et al (2010) Linkage and association mapping of *Arabidopsis thaliana* flowering time in nature. PLoS Genet 6(5):e1000940.
Brachi B, Morris GP, Borevitz O (2011) Genome-wide association studies in plants: the missing heritability is in the field. Genome Biol 12:232.
Bretting PK, Widrlechner MP (1995) Genetic markers and horticultural germplasm management. HortScience 30:1349–1356.
Britton M, Leslie C, McGranahan G, Dandekar A (2007) Analysis of genes expressed in nematode-infected walnut plants. Unpublished raw data.
Brown AHD (1989) Core collections: a practical approach to genetic resources management. Genome 31:818–824.
Brush SB (1995) In situ conservation of landraces in centers of crop diversity. Crop Sci 35:346–354.
Burghardt LT, Young ND, Tiffin P (2017) A Guide to Genome-Wide Association Mapping in Plants. Curr Protoc Plant Biol. 2:22–38.
Bush WS, Moore JH (2012) Chapter 11: Genome-Wide Association Studies. PLoS Comput Biol 8(12):e1002822.
Busov VB, Rink G, Woeste K (2002) Allozyme variation and mating system of black walnut (*Juglans nigra* L.) in the central hardwood region of the United States. For Genet 9:315–322.
Butterfield T, Phu M, Chakraborty S, et al (2015) Identifying the genetic determinant(s) of pellicle coloration in walnut. Walnut Research Reports.
Cabrera-Bosquet L, Crossa J, von Zitzewitz J, Serret MD, Araus JL (2012) High-throughput phenotyping and genomic selection: the frontiers of crop breeding converge. J Integr Plant Biol 54:312–320.
Camerarius RJ (1684) *De sexu plantarum epistola*.
Campoy JA, Lerigoleur-Balsemin E, Christmann H, Beauvieux R, Girollet N, Quero-García J, et al (2016) Genetic diversity, linkage disequilibrium, population structure and construction of a core collection of *Prunus avium* L. landraces and bred cultivars. BMC Plant Biol 16:1–15.
Campoy JA, Ruiz D, Egea J (2011) Dormancy in temperate fruit trees in a global warming context: A review. Sci Hort 130:357–372.
Carrion JS, Sanchez-Gomez P (1992) Palynological Data in Support of the Survival of Walnut (*Juglans regia* L.) in the Western Mediterranean Area During Last Glacial Times. J Biogeogr 19:623–630.
Castède S, Campoy JA, Le Dantec L, Quero-García J, Barreneche T, Wenden B, et al (2015) Mapping of Candidate Genes Involved in Bud Dormancy and Flowering Time in Sweet Cherry (*Prunus avium*). PLoS ONE 10(11): e0143250.
Cavanagh C, Morell M, Mackay I, Powell W (2008) From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. Curr Opin Plant Biol 11:215–221.
Ceballos G, Ehrlich PR, Dirzo R (2017) Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. Proc Natl Acad Sci USA 14:6089–6096.
Ćelepirović N, Karija Vlahović M, Dounavi A, Ivanković M (2016). Optimizations of high throughput multiplex polymerase chain reaction with simple sequence repeat markers for genotyping of common walnut populations (*Juglans regia* L.). Silva Fenn 50 article id 1674.
Chakraborty S (2016) YeATSAM analysis of the chloroplast genome of walnut reveals several putative un-annotated genes and mis-annotation of the trans-spliced rps12 gene in other organisms. Preprint bioRxiv 094441.
Chakraborty S, Britton M, Martínez-García PJ, Dandekar AM (2016) Deep RNA-Seq profile reveals biodiversity, plant–microbe interactions and a large family of NBS-LRR resistance genes in walnut (*Juglans regia*) tissues. AMB Express 6:12.
Chakraborty S, Britton M, Wegrzyn J, et al (2015) YeATS-a tool suite for analyzing RNA-seq derived transcriptome identifies a highly transcribed putative extensin in heartwood/sapwood transition zone in black walnut [version 2; referees: 3 approved]. F1000Research 4:155.
Charmet G (2011) De la détection de QTL à la génétique d'association. Le Sélectionneur Français 62:17–29.
Charrier G, Améglio T. (2011a) The timing of leaf fall affects cold acclimation by interactions with air temperature through water and carbohydrate contents. Environ Exp Bot 72:351–357.
Charrier G, Bonhomme M, Lacointe A, Améglio T (2011b) Are budburst dates, dormancy and cold acclimation in walnut trees (*Juglans regia* L.) under mainly genotypic or environmental control? Int J Biometeorol 55:763–774.
Charrier G, Chuine I, Bonhomme M, Améglio T (2017) Assessing frost damages using dynamic models in walnut trees: exposure rather than vulnerability controls frost risks. Plant Cell Environ DOI: 10.1111/pce.12935.
Charrier G, Ngao J, Saudreau M, Améglio T (2015) Effects of environmental factors and management practices on microclimate, winter physiology and frost resistance in trees. Front Plant Sci 6:259.

Charrier G, Poirier M, Bonhomme M, et al (2013) Frost acclimation in different organs of walnut trees *Juglans regia* L.: How to link physiology and modelling? Tree Physiol 33:1229–1241.

Chen AH, Ge W, Metcalf W, Jakobsson E, Mainzer LS, Lipka AE (2019) An assessment of true and false positive detection rates of stepwise epistatic model selection as a function of sample size and number of markers. Heredity 122:660–671.

Chen C-M, Han S-J, Yuan S-S, et al (2013) Isolation and characterization of 20 polymorphic microsatellite markers for *Juglans mandshurica* (*Juglandaceae*). Appl Plant Sci 1:1–4.

Chen L-H, Hu T-X, Zhang F (2009) AFLP analysis on genetic diversity of *Juglans* populations in dry and dry-hot valleys of Sichuan province. J Fruit Sci 26:48–54.

Chen L-H, Hu T-X, Zhang F, Li G-H (2008) Genetic diversities of four *Juglans* populations revealed by AFLP in Sichuan province, China. J Plant Ecol 32:1362–1372.

Chen W, Hou L, Zhang Z, Pang X, Li Y (2017) Genetic diversity, population structure, and linkage disequilibrium of a core collection of *Ziziphus jujuba* assessed with genome-wide SNPs developed by genotyping-by-sequencing and SSR markers. Front Plant Sci 8:575.

Chen X, Xu L, Zhang SL, Liu ZQ (2014) Walnut Genebank in China National Clonal Plant Germplasm Repository. Acta Hortic 1050:89–94.

Chen X, Zhou DX (2013) Rice epigenomics and epigenetics: challenges and opportunities. Curr Opin Plant Biol 16:164–169.

Cheng S, Yang W (1987) Taxonomic studies of ten species of the genus *Juglans* based on isozymic zymograms. Acta Hortic Sin 14:90–96.

Chmielewski FM, Rötzer T (2001) Response of tree phenology to climate change across Europe. Agric For Meteorol 108:101–112.

Cho LH, Pasriga R, Yoon J, Jeon JS, An G (2018) Roles of Sugars in Controlling Flowering Time. J Plant Biol 61:121–130.

Choi S, Wing RA (2000) The Construction of Bacterial Artificial Chromosome (BAC) Libraries. Plant Mol Biol Man H5:1–28.

Christopoulos MV, Rouskas D, Tsantili E, Bebeli PJ (2010) Germplasm diversity and genetic relationships among walnut (*Juglans regia* L.) cultivars and Greek local selections revealed by Inter-Simple Sequence Repeat (ISSR) markers. Sci Hortic 125:584–592.

Ciarmiello LF, Piccirillo P, Pontecorvo G, et al (2011) A PCR based SNPs marker for specific characterization of English walnut (*Juglans regia* L.) cultivars. Mol Biol Rep 38:1237–1249.

Ciska M, Moreno Díaz de la Espina S (2014) The intriguing plant nuclear lamina. Front Plant Sci 5:166.

Clarke GM, Anderson CA, Pettersson FH, Cardon LR, Morris AP, Zondervan KT (2011) Basic statistical analysis in genetic case-control studies. Nat Protoc 6:121–133.

Clarke JH, Mithen R, Brown JK, Dean C (1995) QTL analysis of flowering time in *Arabidopsis thaliana*. Mol Gen Genet 248:278–286.

Colaric M, Veberic R, Solar A, Hudina M, Stampar F (2005) Phenolic acids, syringaldehyde, and juglone in fruits of different cultivars of *Juglans regia* L. J Agric Food Chem 53:6390–6396.

Cook J, Oreskes N, Doran PT, et al (2016) Consensus on consensus: a synthesis of consensus estimates on human-caused global warming. Environ Res Lett 11:1–8.

Cooke JEK, Eriksson ME, Juntilla O (2012) The dynamic nature of bud dormancy in trees: environmental control and molecular mechanisms. Plant Cell Environ 35:1707–1728.

Cortijo S, Wardenaar R, Colome-Tatche M, et al (2014) Mapping the epigenetic basis of complex traits. Science 343:1145–1148.

Cosgrove DJ (2000) Loosening of plant cell walls by expansins. Nature 407:321–326.

Cosmulescu S, Botu M (2012) Walnut biodiversity in south-western Romania-resource for perspective cultivars. Pak J Bot 44:307–311.

Cosmulescu S, Ionescu MB (2018) Phenological calendar in some walnut genotypes grown in Romania and its correlation with air temperature. Int J Biometeorol 62:2007–2013.

Costa J, Carrapatoso I, Oliveira MBPP, Mafra I (2014) Walnut allergens: molecular characterization, detection and clinical relevance. Clin Exp Allergy 44:319–341.

Coupel-Ledru A, Pallas B, Delalande M, Boudon F, Carrié E, et al (2019) Multi-scale high-throughput phenotyping of apple architectural and functional traits in orchard reveals genotypic variability under contrasted watering regimes. Horticulture Research 6:52.

Črepinšek Z, Solar M, Štampar F, Solar A (2009) Shifts in walnut (*Juglans regia* L.) phenology due to increasing temperatures in Slovenia. J Hortic Sci Biotechnol 84:59–64.

Cubry P, De Bellis F, Pot D, Musoli P, Leroy T (2013) Global analysis of *Coffea canephora* Pierre ex Froehner (*Rubiaceae*) from the Guineo-Congolese region reveals impacts from climatic refuges and migration effects. Genet Resour Crop Evol 60(2):483–501.

Curry KJ, Abril M, Avant JB, Smith BJ (2002) Strawberry anthracnose: Histopathology of *Colletotrichum acutatum* and *C. fragariae*. Phytopathology 92:1055–1063.

Damm U, Cannon PF, Woudenberg JHC, Crous PW (2012) The *Colletotrichum acutatum* species complex. Stud Mycol 73: 37–113.

Dang M, Liu Z-X, Chen X, et al (2015) Identification, development, and application of 12 polymorphic EST-SSR markers for an endemic Chinese walnut (*Juglans cathayensis* L.) using next-generation sequencing technology. Biochem Syst Ecol 60:74–80.

Dang M, Zhang T, Hu Y, Zhou H, Woeste K, et al (2016) *De Novo* Assembly and Characterization of Bud, Leaf and Flowers Transcriptome from *Juglans Regia* L. for the Identification and Characterization of New EST-SSRs. Forests 7:247–263.

Dang PM, Chen CY (2013) Modified method for combined DNA and RNA isolation from peanut and other oil seeds. Mol Biol Rep 40:1563–1568.

Dangl GS, Woeste K, Aradhya MK, Koehmstedt A, Simon C, et al (2005) Characterization of 14 Microsatellite Markers for Genetic Analysis and Cultivar Identification of Walnut. J Am Soc Hortic Sci 130:348–354.

Dean R, Van Kan JL, Pretorius ZA, Hammond-Kosack KE, Di Pietro A, et al (2012) The Top 10 fungal pathogens in molecular plant pathology. Mol Plant Pathol 13:414–430.

Debette S (2012) How to interpret a genome-wide association study (GWAS)? Sang Thrombose Vaisseaux 24:240–247.

De Beukelaer H, Davenport GF, Fack V (2018) Core Hunter 3: flexible core subset selection. BMC Bioinformatics 19:203.

De Candolle A (1885) Origin of Cultivated Plants. University of Michigan. D. Appleton, New York, NY, p 18.

De la Fuente L, Conesa A, Lloret A, et al (2015) Genome-wide changes in histone H3 lysine 27 trimethylation associated with bud dormancy release in peach. Tree Genet Genomes 11:3.

Denis M, Bouvet JM (2013) Efficiency of genomic selection with models including dominance effect in the context of *Eucalyptus* breeding. Tree Genet Genomes 9:37–51.

Der Sarkissian H, Lucka F, van Eijnatten M, et al (2019) A cone-beam X-ray computed tomography data collection designed for machine learning. Scientific Data 6:215.

Desalegne BA, Dagne K, Melaku G, Ousmane B, Fatokun CA (2017) Efficiency of SNP and SSR-based analysis of genetic diversity, population structure, and relationships among cowpea (*Vigna unguiculata* (L.) Walp.) germplasm from East Africa and IITA inbred lines. J Crop Sci Biotechnol 20:107–128.

Diallo A, Prigent C (2011) The serine/threonine kinases that control cell cycle progression as therapeutic targets. Bulletin du Cancer 98:1335–1345.

Diéguez-Uribeondo J, Förster H, Soto-Estrada A, Adaskaveg JE (2005) Subcuticular-intracellular hemitrophic and intercellular necrotrophic development of *Colletotrichum acutatum* on almond. Phytopathology 95:751–758.

Dirlewanger E, Quero-García J, Le Dantec L, et al (2012) Comparison of the genetic determinism of two key phenological traits, flowering and maturity dates, in three *Prunus* species: peach, apricot and sweet cherry. Heredity 109:280–292.

Dode LA (1909) Contribution to the study of the genus *Juglans* (English translation by R. E. Cuendett). Bull Soc Dendrol Fr 12:165–215.

Dong W, Xu C, Li W, et al (2017) Phylogenetic Resolution in *Juglans* Based on Complete Chloroplast Genomes and Nuclear DNA Sequences. Front Plant Sci 8:1148.

Donis-González IR, Guyer DE, Fulbright DW, Pease A (2014) Postharvest noninvasive assessment of fresh chestnut (*Castanea* spp.) internal decay using computer tomography images. Postharvest Biol Tec 94:14–25.

Downs ML, Baumert JL, Taylor SL, Mills ENC (2016) Mass spectrometric analysis of allergens in roasted walnuts. J Proteomics 142:62–69.

Ducci F, De Rogatis A, Proietti R (2010) *Juglans regia* L., phenotypic selection and assessment of genetic variation within a simulated seed orchard. Ann Silvic Res 36:139–150.

Dvorak J, Aradhaya M, Leslie C, Luo M-C (2015) Discovery of the causative mutation of the lateral bearing phenotype in walnut. Walnut Research Reports.

Earl DA, VonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4:359–361.

Eastburn DM, Gubler WD (1990) Strawberry anthracnose: Detection and survival of *Colletotrichum acutatum* in soil. Plant Dis 74:161–163.

Ebrahimi A, Fatahi R, Zamani Z (2011) Analysis of genetic diversity among some Persian walnut genotypes (*Juglans regia* L.) using morphological traits and SSRs markers. Sci Hortic 130:146–151.

Ebrahimi A, Khadivi-Khub A, Nosrati Z, Karimi R (2015) Identification of superior walnut (*Juglans regia*) genotypes with late leafing and high kernel quality in Iran. Sci Hortic 193:195–201.

Ebrahimi A, Zarei A, McKenna JR, Bujdoso G, Woeste KE (2017) Genetic diversity of Persian walnut (*Juglans regia*) in the cold-temperate zone of the United States and Europe. Sci Hortic 220:36–41.

Elsadr H, Sherif S, Banks T, Somers D, Jayasankar S (2019) Refining the Genomic Region Containing a Major Locus Controlling Fruit Maturity in Peach. Sci Rep 9:7522.

Emanuelli F, Lorenzi S, Grzeskowiak L, Catalano V, Stefanini M, et al (2013) Genetic diversity and population structure assessed by SSR and SNP markers in a large germplasm collection of grape. BMC Plant Biol 13:39.

Emilia MM, Spada M, Beritognolo I, Cannata F (1997) Differentiation of walnut hybrids (*Juglans Nigra* L. x *Juglans Regia* L.) through RAPD markers. Acta Hortic 442:43–52.

Erturk U, Dalkilic Z (2011) Determination of genetic relationship among some walnut (*Juglans regia* L.) genotypes and their early-bearing progenies using RAPD markers. Romanian Biotechnol Lett 16:5944–5952.

Escobar MA (2013) RNA-seq and Metabolite Profiling Reveal Novel Functions for the Polyphenol Oxidase Enzyme in Walnut (*Juglans regia*). Plant and Animal Genome XXI Conference, San Diego, CA, January 14.

Escobar MA, Leslie CA, McGranahan GH, Dandekar AM (2002) Silencing crown gall disease in walnut (*Juglans regia* L.). Plant Sci 163:591–597.

Escobar MA, Park J, Polito VS, et al (2000) Using GFP as a Scorable Marker in Walnut Somatic Embryo Transformation. Ann Bot 85:831–835.

Eskandari S, Hassani D, Abdi A (2005) Investigation on genetic diversity of Persian walnut and evaluation of promising genotypes. Acta Hortic 705:159–166.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. Mol Ecol 14:2611–2620.

Famula RA, Richards JH, Famula TR, Neale DB (2019) Association Genetics of Carbon Isotope Discrimination in the Founding Individuals of a Breeding Population of *Juglans regia* L. Tree Genet Genomes 15:6.

Farooqui A, Khan A, Borghetto I, et al (2015) Synergistic Antimicrobial Activity of *Camellia sinensis* and *Juglans regia* against Multidrug-Resistant Bacteria. PLoS ONE 10(2):e0118431.

Fatahi R, Ebrahimi A, Zamani Z (2010) Characterization of Some Iranians and Foreign Walnut Genotypes Using Morphological Traits and RAPDs Markers. Hortic Environ Biotechnol 51:51–60.

Feldman MD, Elaine B (2002) The scientific evidence for a beneficial health relationship between walnuts and coronary heart disease. J Nutr 132:1062S–1101S.

Feng JY, Wen YJ, Zhang J, Zhang YM (2016) Advances on Methodologies for Genome-wide Association Studies in Plants. Acta Agronomica Sinica 42(7):945–956.

Fernández-Cruz J, Fernández-López J (2016) Genetic structure of wild sweet chestnut (*Castanea sativa* Mill.) populations in northwest of Spain and their differences with other European stands. Conserv Genet 17(4):949–967.

Fikret Balta M, Dogan A, Kazankaya A, et al (2007) Pomological Definition of Native Walnuts (*Juglans regia* L.) Grown in Central Bitlis. J Biol Sci 7:442–444.

Filippi CV, Aguirre N, Rivas JG, Zubrzycki J, Puebla A, et al (2015) Population structure and genetic diversity characterization of a sunflower association mapping population using SSR and SNP markers. BMC Plant Biol 15:52.

Fiorani F, Schurr U (2013) Future Scenarios for Plant Phenotyping. Ann Rev Plant Biol 64:267–291.

Fisher RA (1918) The correlation between relatives on the supposition of Mendelian inheritance. Trans. R. Soc. Edinburgh 52:399–433.

Fitchett JM, Grab SW, Thompson DI (2015) Plant phenology and climate change: progress in methodological approaches and application. Progr Phys Geogr Earth Environ 39:460–482.

Fjellstrom RG, Parfitt DE (1994a) RFLP inheritance and linkage in walnut. Theor Appl Genet 89:665–670.

Fjellstrom RG, Parfitt DE (1994b) Walnut (*Juglans* spp.) genetic diversity determined by restriction fragment length polymorphism. Genome 37:690–700.

Fjellstrom RG, Parfitt DE, McGranahan GH (1994) Genetic Relationships and Characterization of Persian Walnut (*Juglans regia* L.) Cultivars Using Restriction Fragment Length Polymorphisms (RFLPs). J Am Soc Hortic Sci 119:833–839.

Flint-Garcia SA, Thornsberry JM, Buckler ES 4th (2003) Structure of linkage disequilibrium in plants. Annu Rev Plant Biol 54:357–374.

Flint-Garcia SA, Thuillet AC, Yu J, Pressoir G, Romero SM et al (2005) Maize association population: a high-resolution platform for quantitative trait locus dissection. Plant J 44:1054–1064.

Fodor A, Segura V, Denis M, et al (2014) Genome-wide prediction methods in highly diverse and heterozygous species: Proof-of-concept through simulation in grapevine. PLoS ONE 9:e110436.

Forde HI (1975) Walnuts. In: Janick J, Moore JN (eds) Advances in Fruit Breeding, Purdue University Press. pp 439–455.

Forde HI, McGranahan GH (1996) Walnuts. In: Janick J, Moore JN (eds) Fruit breeding, vol. 3, Nuts, John Wiley and Sons, Inc. New York, NY. pp 241–274.

Fornari B, Malvolti ME, Taurchini D, et al (2001) Isozyme and organellar DNA analysis of genetic diversity in natural/naturalised European and Asiatic walnut (*Juglans regia* L.) populations. Acta Hortic 544:167–178.

Foroni I, Rao R, Woeste K, Gallitelli M (2005) Characterisation of *Juglans regia* L. with SSR markers and evaluation of genetic relationships among cultivars and the 'Sorrento' landrace. J Hortic Sci Biotechnol 80:49–53.

Foroni I, Woeste K, Monti LM, Rao R (2007) Identification of "Sorrento" walnut using simple sequence repeats (SSRs). Genet Resour Crop Evol 54:1081–1094.

Förster H, Adaskaveg JE (1999). Identification of Subpopulations of *Colletotrichum acutatum* and Epidemiology of Almond Anthracnose in California. Phytopathology 89:1056–1065.

Fournier-Level A, Korte A, Cooper MD, Nordbord M, Schmitt J, Wilczek AM (2011) A map of local adaptation in *Arabidopsis thaliana*. Science 334:86–89.
Fraley RT, Rogers SG, Horsch RB, et al (1983) Expression of bacterial genes in plant cells. Proc Natl Acad Sci U S A 80:4803–4807.
Freeman S, Katan T, Shabi E (1998) Characterization of *Colletotrichum* species responsible for anthracnose diseases of various fruits. Plant Dis 82:596–605.
Freeman S, Shabi E, Katan T (2000) Characterization of *Colletotrichum acutatum* causing anthracnose of anemone (*Anemone coronaria* L.). Appl Environ Microbiol 66:5267–5272.
Frichot E, François O (2015) LEA: An R package for landscape and ecological association studies. Methods Ecol Evol 6:925–929.
Frichot E, Mathieu F, Trouillon T, Bouchard G, François O (2014) Fast and efficient estimation of individual ancestry coefficients. Genetics 196:973–983.
Fu YB (2017) The Vulnerability of Plant Genetic Resources Conserved Ex Situ. Crop Sci 57:2314–2328.
Furbank RT, Tester M (2011) Phenomics – technologies to relieve the phenotyping bottleneck. Trends Plant Sci 16:635–644.
Gaisberger H, Legay S, Andre C, Loo J, Azimov R, et al (2020) Diversity Under Threat: Connecting Genetic Diversity and Threat Mapping to Set Conservation Priorities for *Juglans regia* L. Populations in Central Asia. Front Ecol Evol 8:171.
Gallagher MD, Chen-Plotkin AS (2018) The Post-GWAS Era: From Association to Function. Am J Hum Genet 102:717–730.
Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, et al (2002) The structure of haplotype blocks in the human genome. Science 296:2225–2229.
Gallais A (2011) Méthodes de creation de variétés en amélioration des plantes. Quæ.
Gallais A (2018) Histoire de la génétique et de l'amélioration des plantes. Quæ.
Gallusci P, Dai Z, Génard M, et al (2017) Epigenetics for Plant Improvement: Current Knowledge and Modeling Avenues. Trends Plant Sci 22:610–623.
Gandev S, Dzhuvinov V (2015) Evaluation of Some Walnut Cultivars under the Climatic Conditions of South Bulgaria. Agroznanje 15:5–16.
Garcia-Aloy M, Llorach R, Urpi-Sarda M, et al (2014) Novel multimetabolite prediction of walnut consumption by a urinary biomarker model in a free-living population: the PREDIMED study. J Proteome Res 13:3476–3483.
Garris AJ, Tai TH, Coburn J, Kresovich S, McCough S (2005) Genetic structure and diversity in *Oryza sativa* L. Genetics 169:1631–1638.
Germain E (1990) Inheritance of late leafing and lateral bud fruitfulness in walnut (*Juglans regia* L.). Phenotypic correlations among some traits of the trees. Acta Hortic 284:125–134.
Germain E (1997) Genetic improvement of the Persian walnut (*Juglans regia* L.). Acta Hort 442:21–32.
Germain E, Prunet J-P, Garcin A (1999) Le noyer, monographie, Ctifl.
Ghasemi M, Arzani K, Hassani D (2012) Evaluation and identification of walnut (*Juglans regia* L.) genotypes in Markazi province of Iran. Crop Breed J 2:119–124.
Ghimiray M, Vernooy R (2017) The importance and challenges of crop germplasm interdependence: the case of Bhutan. Food Sec 9:301–310.
Giraud M (2012) Observation sur la flore fongique du noyer: cas particulier du *Colletotrichum*. InfosCtifl, 285. pp 4–5.
GISTEMP Team (2018) GISS Surface Temperature Analysis (GISTEMP). NASA Goddard Institute for Space Studies. Dataset accessed 2018-08-22 at https://data.giss.nasa.gov/gistemp/.
Goh L, Yap VB (2009) Effects of normalization on quantitative traits in association test. BMC Bioinformatics 10:415.
Golzarian MR, Frick RA, Rajendran K, et al (2011) Accurate inference of shoot biomass from high-throughput images of cereal plants. Plant Methods 7:1–11.
Gordo O, Sanz JJ (2005) Phenology and climate change: a long-term study in a Mediterranean locality. Oecologia 146:484–495.
Grattapaglia D, Sederoff R (1994) Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a Pseudo-Testcross: Mapping strategy and RAPD markers. Genetics 137:1121–1137.
Grauke LJ, Thompson TE, Mendoza-Herrera MA (2012) Native Walnuts of Texas. Acta Hortic 948:199–210.
Guàrdia M, Charrier G, Vilanova A, et al (2016) Genetics of frost hardiness in *Juglans regia* L. and relationship with growth and phenology. Tree Genet Genomes 12:83.
Guichoux E, Lagache L, Wagner S, Chaumeil P, Léger P, et al (2011) Current trends in microsatellite genotyping. Mol Ecol Resour 11:591–611.
Gunn BF, Aradhya M, Salick JM, et al (2010) Genetic variation in walnuts (*Juglans regia* and *J. sigillata* ; *Juglandaceae*) : Species distinctions, human impacts, and the conservation of agrobiodiversity in Yunnan, China. Am J Bot 97:660–671.
Haga H, Yamada R, Ohnishi Y, Nakamura Y, Tanaka T (2002) Gene-based SNP discovery as part of the Japanese Millenium Genome Project: identification of 190,562 genetic variations in the human genome. J Hum Genet 47:605–610.

Haghjooyan R, Ghareyazi B, Sanei Shariat-Panahi M, Khalighi A (2005) Investigation of genetic variation of walnut in some region of Iran by using quantitative morphological characters. Pajouhesh Sazandegi 69:22–30.

Hajri A, Meyer D, Delort F, Guillaumès J, Brin C, Manceau C (2010) Identification of a genetic lineage within *Xanthomonas arboricola* pv. *juglandis* as the causal agent of vertical oozing canker of Persian (English) walnut in France. Plant Pathology 59:1014–1022.

Halewood M, Chiurugwi T, Hamilton RS, et al (2018) Plant genetic resources for food and agriculture: opportunities and challenges emerging from the science and information technology revolution. New Phytol 217:1407–1419.

Hamblin MT, Warburton ML, Buckler ES (2007) Empirical comparison of simple sequence repeats and single nucleotide polymorphisms in assessment of maize diversity and relatedness. PLoS ONE 2(12):e1367.

Hammer K, Teklu Y (2008) Plant Genetic Resources: Selected Issues from Genetic Erosion to Genetic Engineering. J Agr Rural Dev Trop 109:15–50.

Han H, Woeste KE, Hu Y, et al (2016) Genetic diversity and population structure of common walnut (*Juglans regia*) in China based on EST-SSRs and the nuclear gene phenylalanine ammonia-lyase (PAL). Tree Genet Genomes 12:111.

Hansche PE, Beres V, Forde HI (1972) Estimates of quantitative genetic properties of walnut and their implications for cultivar improvement. J Am Soc Hortic Sci 97:279–285.

Hansen J, Ruedy R, Sato M, Lo K (2010) Global surface temperature change. Rev Geophys 48:RG4004.

Harrison RD, Gardner WA, Tollner WE, Kinard DJ (1993) X-ray computed tomography studies of the burrowing behaviour of fourth-instar pecan weevil (*Coleoptera*: *Curculionidae*). J Econ Entomol 86:1714–1719.

Hassani D, Dastjerdi R, Haghjooyan R, et al (2014) Genetic Improvement of Persian Walnut (*Juglans regia* L.) in Iran. Acta Hortic 1050:95–102.

Hassankhah A, Vahdati K, Rahemi M, Sarikhani S (2017) Persian Walnut Phenology: Effect of Chilling and Heat Requirements on Budbreak and Flowering Date. Int J Hort Sci Tech 4:259–271.

He F, Wang H, Zhang Z, et al (2010) Identification of walnut cultivars with AFLP fingerprinting. Acta Hortic 861:151–154.

Heffner EL, Jannink JL, Sorrells ME (2011) Genomic selection accuracy using multifamily prediction models in a wheat breeding program. Plant Genome 4:65–75.

Henry RJ (2012) Evolution of DNA marker technology in plants. In: Robert H (ed) Molecular Markers in Plants. Blackwell Publishing Ltd, United Kingdom. pp 1–19.

Heuertz M, De Paoli E, Kallman T, Larsson H, Jurman I et al (2006) Multilocus patterns of nucleotide diversity, linkage disequilibrium and demographic history of Norway spruce *Picea abies* (L.) Karst. Genetics 174:2095–2105.

Hiersche M, Rühle F, Stoll M (2013) Postgwas: Advanced GWAS Interpretation in R. PLoS ONE 8(8):e71775.

Hillis DM (1987) Molecular versus morphological approaches to systematics. Ann Rev Ecol Syst 18:23–42.

Hintun TJL, Brown AHD, Spillane C, Hodgkin T (2000) Core collections of plant genetic resources. IPGRI Technical Bulletin 3:48.

Hoban S, Anderson R, McCleary T, et al (2008) Thirteen nuclear microsatellite loci for butternut (*Juglans cinerea* L.). Mol Ecol Resour 8:643–646.

Horowitz S, Freeman S, Sharon A (2002) Use of green fluorescent protein-transgenic strains to study pathogenic and non-pathogenic lifestyles in *Colletotrichum acutatum*. Phytopathology 92:743–749.

Horvath D (2009) Common mechanisms regulate flowering and dormancy. Plant Sci 177:523–531.

Hu J, Wang L, Li J (2011) Comparison of genomic SSR and EST-SSR markers for estimating genetic diversity in cucumber. Biol Plant 55:577–580.

Hu Y, Woeste KE, Dang M, et al (2016) The complete chloroplast genome of common walnut (*Juglans regia*). Mitochondrial DNA Part B 1:189–190.

Hu Y, Woeste KE, Zhao P (2017) Completion of the Chloroplast Genomes of Five Chinese *Juglans* and Their Contribution to Chloroplast Phylogeny. Front Plant Sci 7:1955.

Hu Y-H, Zhao P, Zhang Q, et al (2015) *De novo* assembly and characterization of transcriptome using Illumina sequencing and development of twenty five microsatellite markers for an endemic tree *Juglans hopeiensis* Hu in China. Biochem Syst Ecol 63:201–211.

Huang J, Liao H, Zhu Y, et al (2012) Hyperspectral detection of rice damaged by rice leaf folder (*Cnaphalocrocis medinalis*). Comput Electr Agric 82:100–107.

Huang M, Liu X, Zhou Y, Summers RM, Zhang Z (2018) BLINK: a package for the next level of genome-wide association studies with both individuals and markers in the millions. GigaScience 8:1–12.

Huang Z, Meilan R, Woeste K (2009) A KNAT3-like homeobox gene from *Juglans nigra* L., JnKNAT3-like, highly expressed during heartwood formation. Plant Cell Rep 28:1717–1724.

Hunter RL, Markert CL (1957) Histochemical demonstration of enzymes separated by zone electrophoresis in starch gels. Science 125:1294–1295.

IHGSC (2004) Finishing the euchromatic sequence of the human genome. Nature 431:931–945.

Ikegawa S (2012) A Short History of the Genome-Wide Association Study: Where We Were and Where We Are Going. Genomics Inform 10:220–225.

Ikhsan AS, Topçu H, Sütyemez M, Kafkas S (2016) Novel 307 polymorphic SSR markers from BAC-end sequences in walnut (*Juglans regia* L.): Effects of motif types and repeat lengths on polymorphism and genetic diversity. Sci Hortic 213:1–4.

Ince AG, Karaca M, Uygur Göçer E, Aydın A (2017) Descriptive Statistics and PIC Values of Genomic-and Transcriptomic-Microsatellites in Several Plant Species. J Sci Eng Res 2017:236–246.

IPCC (2014) Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Core writing team, Pachauri RK, Meyer LA (eds) IPCC, Geneva, Switzerland. pp 151.

IPGRI (1994) Descriptors for walnut (*Juglans* spp.). International Plant Genetic Resources Institute, Rome, Italy.

Iwata H, Minamikawa MF, Kajiya-Kanegae H, Ishimori M, Hayashi T (2016) Genomics-assisted breeding in fruit trees. Breeding Science 66:100–115.

Jacquenet S, Moneret-Vautrin DA (2007) Classification botanique des fruits à coque. Rev Fr Allergol 47:487–491.

Jahanbani R, Ghaffari SM, Salami M, et al (2016) Antioxidant and Anticancer Activities of Walnut (*Juglans regia* L.) Protein Hydrolysates Using Different Proteases. Plant Foods Hum Nutr 71:402–409.

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23:1801–1806.

Jarolmasjed S, Sankaran S, Marzougui A, et al (2019) High-Throughput Phenotyping of Fire Blight Disease Symptoms Using Sensing Techniques in Apple. Front Plant Sci 10:576.

Jaynes RA (1969) Handbook of North American Nut Trees, Northern Nut Growers Association.

Jenner RA (2004) Accepting Partnership by Submission? Morphological Phylogenetics in a Molecular Millennium. Syst Biol 53:333–342.

Ji A, Wang Y, Wu G, et al (2014) Genetic Diversity and Population Structure of North China Mountain Walnut Revealed by ISSR. Am J Plant Sci 5:3194–3202.

Jia J, Li H, Zhang X, Li Z, Qiu L (2017) Genomics-based plant germplasm research (GPGR). Crop J 5:166–174.

Jiang S, Balan B, Assis RAB (2020) Genome-Wide Profiling and Phylogenetic Analysis of the SWEET Sugar Transporter Gene Family in Walnut and Their Lack of Responsiveness to *Xanthomonas arboricola* pv. *juglandis* Infection. Int J Mol Sci 21:1251.

Jochner S, Menzel A (2015) Does flower phenology mirror the slowdown of global warming? Ecol Evol 5:2284–2295.

Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403–1405.

Jonas E, de Koning DJ (2013) Does genomic selection have a future in plant breeding. Trends Biotechnol 31:497–504.

Jones AL, Ehret GR, Meyer MP, Shane WW (1996) Occurrence of bitter rot on apple in Michigan. Plant Dis 80:1294–1297.

Jones HG, Serraj R, Loveys BR, et al (2009) Thermal infrared imaging of crop canopies for the remote diagnosis and quantification of plant responses to water stress in the field. Funct Plant Biol 36:978–989.

Kafkas S, Ozkan H, Sütyemez M (2005) DNA Polymorphism and Assessment of Genetic Relationships in Walnut Genotypes Based on AFLP and SAMPL Markers. J Am Soc Hortic Sci 130:585–590.

Kaler AS, Gilman JD, Beissinger T, Purcell LC (2020) Comparing Different Statistical Models and Multiple Testing Corrections for Association Mapping in Soybean and Maize. Front Plant Sci 10:1794.

Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Molecular Ecology 16:1099–1106.

Kang HM, Sul JH, Service SK, et al (2010) Variance component model to account for sample structure in genome-wide association studies. Nature Genetics 42:348–354.

Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, et al (2008) Efficient control of population structure in model organism association mapping. Genetics 178:1709–1723.

Kao CH, Zeng ZB, Teasdale RD (1999) Multiple interval mapping for quantitative trait loci. Genetics 152:1203–1216.

Karadag H, Akça Y (2011) Phenological and pomological properties of promising walnut (*Juglans regia* L.) genotypes from selected native population in Amasya Province. Afr J Biotechnol 10:16763–16768.

Karimi R, Ershadi A, Vahdati K, Woeste K (2010) Molecular Characterization of Persian Walnut Populations in Iran with Microsatellite Markers. HortScience 45:1403–1406.

Keenan K, McGinnity P, Cross TF, Crozier WW, Prodöhl PA (2013) diveRsity: An R package for the estimation of population genetics parameters and their associated errors. Methods in Ecology and Evolution 4:782–788.

Keneni G, Bekele E, Imtiaz M, Dagne K (2012) Genetic Vulnerability of Modern Crop Cultivars: Causes, Mechanism and Remedies. Int J Plant Res 2:69–79.

Keqiang Y, Yuejin W, Yindong Z, Xueqin Z (2002) RAPD Analysis for the Identification of the Precocious Trait in Walnuts. Acta Hortic Sin 2002:06.

Khadivi-Khub A, Ebrahimi A, Sheibani F, Esmaeili A (2015) Phenological and pomological characterization of Persian walnut to select promising trees. Euphytica 205:557–567.

Khan MW, Khan IA, Ahmad H, et al (2010) Estimation of genetic diversity in walnut. Pak J Bot 42:1791–1796.

Khanduri VP, Sharma CM, Singh SP (2008) The effects of climate change on plant phenology. The Environmentalist 28:143–147.

Khosa I, Pasero E (2014) Feature extraction in X-ray images for hazelnuts classification. Proc Int Jt Conf Neural Netw 2354–2360.

Kim S, Schatzki T (2001) Detection of pinholes in almonds through X-ray imaging. Trans ASAE 44:997–1003.

Kluepfel D, Leslie C, Aradhya M, Browne G (2015) Development of disease-resistant walnut rootstocks: integration of conventional and genomic approaches (SCRI-match Year 3). Walnut Research Reports.

Knipfer T, Barrios-Masias FH, Cuneo IF, et al (2018) Variations in xylem embolism susceptibility under drought between intact saplings of three walnut species. Tree Physiol 38:1180–1192.

Kolbert E (2014) The Sixth Extinction: an Unnatural History. Henry Holt and Company, New York.

Komanich IG (1980) Correlations between walnut fruit characters in a seed population. Bull Acad Sci RSS Mold 2:27–30.

Kornienko NA (1974) Types of dichogamy in walnut. Plant Breed Abstr 44:452.

Korte A, Farlow A (2013) The advantages and limitations of trait analysis with GWAS: a review. Plant Methods 9:29.

Korte A, Vilhjálmsson BJ, Segura V, Platt A, Long Q, Nordborg M (2012) A mixed-model approach for genome-wide association studies of correlated traits in structured populations. Nat Genet 44:1066–1071.

Kosambi D (1944) The estimation of map distances from recombination values. Annals of Eugenics 12:172–175.

Kotwaliwale N, Singh K, Kalne A, et al (2014) X-ray imaging methods for internal quality evaluation of agricultural produce. J Food Sci Technol 51:1–15.

Kotwaliwale N, Weckler PR, Brusewitz GH (2006) X-ray attenuation coefficients using polychromatic X-ray imaging of pecan components. Biosyst Eng 94:199–206.

Kowalski SP, Lan TH, Feldmann KA, Paterson AH (1994) QTL mapping of naturally-occuring variation in flowering time of *Arabidopsis thaliana*. Mol Gen Genet 245:548–555.

Lamarck J-B de M de, Poiret J-L-M (1783) Encyclopédie méthodique: botanique. Paris.

Lamari L (2008) Assess 2.0: image analysis software for plant disease quantification. St. Paul, MN: American Phytopathological Society.

Lambert P, Dirlewanger E, Laurens F (2009) La sélection assistée par marqueurs (SAM) chez les arbres fruitiers: une approche prometteuse au service de l'innovation variétale. Innovations Agronomiques 7:139–152.

Lamichhane JR (2014) *Xanthomonas arboricola* diseases of stone fruit, almond, and walnut trees: progress toward understanding and management. Plant Disease 98:1600–1610.

Lammertyn J, Jancsok P, Dresselaers T, et al (2003) Analysis of the time course of core breakdown in 'Conference' pears by means of MRI and X-ray CT. Postharvest Biol Technol 29:19–28.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al (2001) Initial sequencing and analysis of the human genome. Nature 409:860–921.

Laval G, San Cristobal M, Chevalet C (2002) Measuring genetic distances between breeds: use of some distances in various short term evolution models. Genet Sel Evol 34:481–507.

Lê S, Josse J, Husson F (2008) FactoMineR: An R Package for Multivariate Analysis. J Stat Softw 25:1–18.

Leslie C, McGranahan G, Hackett W, et al (2015) Walnut improvement program 2015. Walnut Research Reports.

Leslie CA, McGranahan GH, Dandekar AM, et al (2001) Development and field-testing of walnuts expressing the cry1A(c) gene for lepidopteran insect resistance. Acta Hortic 544:195–199.

Leslie CE, McGranahan GH (1988) Native populations of *Juglans regia* : a draft. In: International Conference on Walnuts. Hort. Research Inst., ed. 124, Yalova, Turkey, Ataturk Central.

Li L, Zhang Q, Huang D (2014) A review of imaging techniques for plant phenotyping. Sensors 14:20078–20111.

Li Z, Lanying Z, Qianwen X (2007) Identification of RAPD Markers Linked to Thickness Gene of Shuck in Walnut. Adv Biol Res 5–6:137–140.

Liang W, Dondini L, De Franceschi P, Paris R, Sansavini S, Tartarini S (2015) Genetic Diversity, Population Structure and Construction of a Core Collection of Apple Cultivars from Italian Germplasm. Plant Mol Biol Report 33(3):458–473.

Liao Z, Feng K, Chen Y, et al (2014) Genome-wide discovery and analysis of single nucleotide polymorphisms and insertions/deletions in *Juglans regia* L. by high-throughput pyrosequencing. Plant Omics J 7:445–449.

Lin RZ, Li RQ, Lu AM, Zhu JY, Chen ZD (2016) Comparative flower development of *Juglans regia*, *Cyclocarya paliurus* and *Engelhardia spicata*: homology of floral envelopes in *Juglandaceae*. Bot J Linn Soc 181:279–293.

Lindow S, Olson W, Buchner R (2014) Colonization of Dormant Walnut Buds by *Xanthomonas arboricola* pv. *juglandis* Is Predictive of Subsequent Disease. Phytopathology 104:1163–1174.

Lipka AE, Tian F, Wang Q, Peiffer J, Li M, et al (2012) GAPIT: genome association and prediction integrated tool. Bioinformatics 28:2397–2399.

Lippert C, Listgarten J, Liu Y, et al (2011) FaST linear mixed models for genome-wide association studies. Nat Methods 8:833–835.

Listgarten J, Lippert C, Heckerman D, et al (2013) FaST-LMM-Select for addressing confounding from spatial structure and rare variants. Nat Genet 45:470–471.

Liu B, Louws FJ, Sutton TB, Correll JC (2012) A rapid qualitative molecular method for the identification of *Colletotrichum acutatum* and *C. gloeosporioides*. Eur J Plant Pathol 132:593–607.

Liu D, Hu R, Palla KJ, Tuskan GA, Yang X (2016) Advances and perspectives on the use of CRISPR/Cas9 systems in plant genomics research. Curr Opin Plant Biol 30:70–77.

Liu X, Huang M, Fan B, et al (2016) Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. PLoS Genet 12:e1005767.

Liu X, Walawage SL, Leslie CA, et al (2017) In vitro gene expression and mRNA translocation from transformed walnut (*Juglans regia*) rootstocks expressing DsRED fluorescent protein to wild-type scions. Plant Cell Rep 36:877–885.

Ljungqvist M, Åkesson M, Hansson B (2010) Do microsatellites reflect genome- wide genetic diversity in natural populations? A comment on. Mol Ecol 19:851–855.

Long Y, Xia W, Li R, Wang J, Shao M, et al (2011) Epigenetic QTL mapping in Brassica napus. Genetics 189:1093–1102.

Lorenz A, Chao S, Asoro F, Heffner EL, Hayashi T, Iwata H, et al. (2011). Genomic selection in plant breeding: knowledge and prospects. Adv Agron 110:77–123.

Loreti S, Gallelli A, Belisario A, et al (2001) Investigation of genomic variability of *Xanthomonas arboricola* pv. *juglandis* by AFLP analysis. Eur J Plant Pathol 107:583–591.

Luedeling E, Gassner A (2012) Partial Least Squares Regression for analyzing walnut phenology in California. Agric For Meteorol 158-159:43–52.

Luedeling E, Steinmann K, Zhang M, et al (2010) Climate change effects on walnut pests in Californie. Glob Change Biol 17:228–238.

Lukk T, Sakai A, Kalyanaraman C, Brown SD, Imker HJ, Song L, et al (2012) Homology models guide discovery of diverse enzyme specificities among dipeptide epimerases in the enolase superfamily. Proc Natl Acad Sci USA 109:4122–4127.

Luo C, Li X, Dai Q (2014) Biology's drones: New and improved. Science 344:1351.

Luo M-C, You FM, Li P, et al (2015) Synteny analysis in Rosids with a walnut physical map reveals slow genome evolution in long-lived woody perennials. BMC Genomics 16:707.

Ma Q, Zhang J, Pei D (2011) Genetic Analysis of Walnut Cultivars in China Using Fluorescent Amplified Fragment Length Polymorphism. J Am Soc Hortic Sci 136:422–428.

Maguire LS, O'Sullivan SM, Galvin K, et al (2004) Fatty acid profile, tocopherol, squalene and phytosterol content of walnuts, almonds, peanuts, hazelnuts and the macadamia nut. Int J Food Sci Nutr 55:171–178.

Mahmoodi R, Dadpour MR, Hassani D, et al (2019) Development of a core collection in Iranian walnut (*Juglans regia* L.) germplasm using the phenotypic diversity. Sci Hortic 249:439–448.

Mahmoodi R, Hassani D, Amiri ME, Jaffaraghaei M (2016) Phenological and Pomological Characteristics of Five Promised Walnut Genotypes in Karaj, Iran. J Nuts 7:1–8.

Mahmoodi R, Rahmani F, Rezaee R (2013) Genetic diversity among *Juglans regia* L. genotypes assessed by morphological traits and microsatellite markers. Span J Agric Res 11:431–437.

Majacka AD (1971) Dichogamy and fertility in walnut. Plant Breed Abstr 41:186.

Malvolti ME, Fineschi S, Pigliucci M (1994) Morphological Integration and Genetic Variability in *Juglans regia* L. J Hered 85:389–394.

Malvolti ME, Fornari B, Maccaglia E, Cannata F (2001) Genetic linkage mapping in an intraspecific cross of walnut (*Juglans regia* L.) using molecular markers. Acta Hortic 544:179–185.

Malvolti ME, Paciucci M, Cannata F, Fineschi S (1993) Genetic variation in Italian populations of *Juglans regia* L. Acta Hortic 311:86–94.

Malvolti ME, Pollegioni P, Bertani A, et al (2010) *Juglans regia* Provenance Research by Molecular, Morphological and Biochemical Markers : A Case Study in Italy. Bioremediation Biodivers Bioavailab 4:84–92.

Manchester SR (1987) The fossil history of *Juglandaceae*. Missouri Botanical Garden.

Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen W-M (2010) Robust relationship inference in genome-wide association studies. Bioinformatics 26:2867–2873.

Manning WE (1978) The classification within the *Juglandaceae*. Ann Mo Bot Gard 65:1058–1087.

Mao X, Hua Y, Chen G (2014) Amino Acid Composition, Molecular Weight Distribution and Gel Electrophoresis of Walnut (*Juglans regia* L.) Proteins and Protein Fractionations. Int J Mol Sci 15:2003–2014.

Marees AT, De Kluiver H, Stringer S, Vorspan F, Curis E, et al (2017) A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. Int J Methods Psychiatr Res 27:e1608.

Mariette S, Wong Jun Tai F, Roch G, Barre A, Chague A, et al (2016) Genome-wide association links candidate genes to resistance to *Plum Pox Virus* in apricot (*Prunus armeniaca*). New Phytol 209:773–784.

Marrano A, Britton M, Zaini PA, Zimin AV, Workman RE, et al (2020) High-quality chromosome-scale assembly of the walnut (*Juglans regia* L) reference genome. bioRxiv 809798; doi:https://doi.org/10.1101/809798.

Marrano A, Martínez-García PJ, Bianco L (2019a) A new genomic tool for walnut (*Juglans regia* L.): development and validation of the high-density Axiom™ *J. regia* 700K SNP genotyping array. Plant Biotechnol J 17:1027–1036.

Marrano A, Sideli GM, Leslie CA, Cheng H, Neale DB (2019b) Deciphering of the Genetic Control of Phenology, Yield and Pellicle Color in Persian Walnut (*Juglans regia* L.). Front Plant Sci 10:1140.

Marris E (2013) Drones in science: Fly, and bring me data. Nature 498:156–158.

Martínez ML, Labuckas DO, Lamarque AL, Maestri DM (2010) Walnut (*Juglans regia* L.) : genetic resources, chemistry, by-products. J Sci Food Agric 90:1959–1967.
Martínez-García PJ, Crepeau MW, Puiu D, et al (2016) The walnut (*Juglans regia*) genome sequence reveals diversity in genes coding for the biosynthesis of nonstructural polyphenols. Plant J 87:507–532.
Marulanda JJ, Melchinger AE, Würschum T (2015) Genomic selection in biparental populations: assessment of parameters for optimum estimation set design. Plant Breeding 134:623–630.
Mason AS, Zhang J, Tollenaere R, et al (2015) High-throughput genotyping for species identification and diversity assessment in germplasm collections. Mol Ecol Resour 15:1091–1101.
Mathworks T (2013) MATLAB. https://www.mathworks.com/products/matlab.html.
Maxted N, Ford-Lloyd BV, Hawkes JG (1997) Complementary conservation strategies. In: Maxted N, Ford-Lloyd BV, Hawkes JG (eds) Plant Genetic Resources: The In situ Approach, Chapman and Hall, London. pp 15–40.
McGranahan G, Leslie C (1991) Walnuts (*Juglans*). In: Moore JN, Ballington JR. Jr (eds) Genetic resources of temperate fruit and nut crops, part 2, Wageningen: The Netherlands; International Society of Horticultural Sciences, pp 907–951.
McGranahan G, Leslie C (2004) 'Robert Livermore', a Persian walnut cultivar with a red seedcoat. HortScience 39:1772.
McGranahan G, Leslie C (2009) Breeding Walnuts (*Juglans regia*). In: Jain SM, Priyadarshan PM (eds) Breeding Plantation Tree Crops: Temperate Species, Springer Science, pp 249–273.
McGranahan G, Leslie C (2012) Walnut. In: Badenes ML, Byrne DH (eds) Fruit Breeding. Springer US, Boston, MA. pp 827–846.
McGranahan GH, Leslie CA, Uratsu SL, et al (1988) Agrobacterium-Mediated Transformation of Walnut Somatic Embryos and Regeneration of Transgenic Plants. Nat Biotechnol 6:800–804.
McGranahan GH, Leslie CA, Uratsu SL, Dandekar AM (1990) Improved efficiency of the walnut somatic embryo gene transfer system. Plant Cell Rep 8:512–516.
McGranahan GH, Tulecke W, Arulsekar S, Hansen JJ (1986) Intergeneric Hybridization in the *Juglandaceae* : *Pterocarya* sp. x *Juglans regia*. J Am Soc Hortic Sci 111:627–630.
McGrown J (2006) Out of Africa: Mysteries of Access and Benefit Sharing. In: Burrows B (ed) Edmonds Institute and the Africa Centre for Biosafety, Washington, USA. p 54.
Menzel A, Fabian P (1999) Growing season extended in Europe. Nature 397:659.
Menzel A, Sparks TH, Estrella N, et al (2006) European phenological response to climate change matches the warming pattern. Glob Change Biol 12:1969–1976.
Meuwissen T (2003) Genomic selection: the future of marker-assisted selection and animal breeding. In: Marker-assisted selection: a fast track to increase genetic gain in plant and animal breeding? Session II: MAS in animals. p 54-59.
Meuwissen THE, Karlsen A, Lien S, Olsaker I, Goddard ME (2002) Fine mapping of a quantitative trait locus for twinning rate using combined linkage and linkage disequilibrium mapping. Genetics 161:373–379.
Meyer P (2015) Epigenetic variation and environmental change. J Exp Bot 66:3541–3548.
Michaels SD, Amasino RM (1999) FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering. Plant Cell 11:949–956.
Miletic R, Zikic M, Mitic N, Nikolic R (2003) Biological and pomological characteristics of superior walnut selections. Genetika 35:123–130.
Minamikawa MF, Takada N, Terakami S, et al (2018) Genome-wide association study and genomic prediction using parental and breeding populations of Japanese pear (*Pyrus pyrifolia* Nakai). Sci Rep 8:11994.
Mir RR, Reynolds M, Pinto F, Khan MA, Bhat MA (2019) High-throughput phenotyping for crop improvement in the genomics era. Plant Sci 282:60–72.
Mo JY, Navi SS, Li X, Guo T, Yang XB (2007) Detached Leaf Assay: A Rapid Screening Technique to Study Foliar Diseases of Soybean. Plant Pathology Presentations and Posters.
Modrusan Z, Reiser L, Feldmann KA, Fischer RL, Haughn GW (1994) Homeotic transformation of ovules into carpel-like structures in Arabidopsis. Plant Cell 6:333–349.
Mohammadi M, Xavier A, Beckett T, Beyer S, Chen L, Chikssa H, et al (2020) Identification, deployment, and transferability of quantitative trait loci from genome-wide association studies in plants. Curr Plant Biol In Press, available online 27 March 2020.
Mohsenipoor S, Vahdati K, Amiri R, Mozaffari MR (2010) Study of the Genetic Structure and Gene Flow in Persian Walnut (*Juglans regia* L.) Using SSR Markers. Acta Hortic 861:133–142.
Moldenke HN, Moldenke AL (1952) Plants of the Bible, Chronica Botanica Co. Waltham, Mass.
Moragrega C, Matias J, Aletà N, Montesinos E, Rovira M (2011) Apical necrosis and premature drop of Persian (English) walnut fruit caused by *Xanthomonas arboricola* pv. *juglandis*. Plant Disease 95:1565–1570.
Moragrega C, Özaktan H (2010) Minireview: Apical necrosis of Persian (English) walnut (*Juglans regia*): an update. Journal of Plant Pathology 92:S1.67–S1.71.
Morgante M, Olivieri AM (1993) PCR-amplified microsatellites as markers in plant genetics. Plant J 3:175–182.

Mozaffarian F, Mardi M, Sarafrazi A, Ganbalani GN (2008) Assessment of geographic and host-associated population variations of the carob moth, *Ectomyelois ceratoniae*, on pomegranate, fig, pistachio and walnut, using AFLP markers. J Insect Sci 8:6.

Muir R, Baek J, Leslie A, et al (2004) Analysis of genes expressed in walnut seed coat tissue. Unpublished raw data.

Müller BSF, Pappas GJ, Valdisser PAMR, Coelho GRC, de Menezes IPP, et al (2015) An operational SNP panel integrated to SSR marker for the assessment of genetic diversity and population structure of the common bean. Plant Mol Biol Rep 33:1697–1711.

Munns R, James RA, Sirault XR, Furbank RT, Jones HG (2010) New phenotyping methods for screening wheat and barley for beneficial responses to water deficit. J Exp Bot 61:3499–3507.

Muto H, Yabe N, Asami T, Hasunuma K, Yamamoto KT (2004) Overexpression of constitutive differential growth 1 gene, which encodes a RLCK VII- subfamily protein kinase, causes abnormal differential and elongation growth after organ differentiation in Arabidopsis. Plant Physiol 136:3124–3133.

Myers N (1994) Protected Areas – protected for a Greater What? Biodivers Conser 3:411–418.

Myles S, Peiffer J, Brown PJ, Ersoz ES, Zhang Z, Costich DE, Buckler ES (2009) Association Mapping: Critical Considerations Shift from Genotyping to Experimental Design. Plant Cell 21:2194–2202.

Nagaraj SH, Gasser RB, Ranganathan S (2006) A hitchhiker's guide to expressed sequence tag (EST) analysis. Brief Bioinform 8:6–21.

Najafi F, Mardi M, Fakheri B, et al (2014) Isolation and Characterization of Novel Microsatellite Markers in Walnut (*Juglans regia* L.). Am J Plant Sci 5:409–415.

National Research Council (1993) Managing global genetic resources: agricultural crop issues and policies. The National Academies Press, Washington.

Neale D, Martínez-García PJ, Crepeau MW, et al (2015) Application of marker breeding in the walnut improvement program. Walnut Research Reports.

Nei M (1972) Genetic Distance between Populations. Am Nat 106:283–292.

Newhouse AE, Spitzer JE, Maynard CA, Powell WA (2014) Chestnut Leaf Inoculation Assay as a Rapid Predictor of Blight Susceptibility. Plant Disease 98:4–9.

Nicese FP, Hormaza JI, McGranahan GH (1998) Molecular characterization and genetic relatedness among walnut (*Juglans regia* L.) genotypes based on RAPD markers. Euphytica 101:199–206.

Ninot A, Aletà N (2003) Identification and genetic relationship of Persian walnut genotypes using isozyme markers. J Am Pomol Soc 57:106–114.

Noor Shah U, Mir JI, Ahmed N, Fazili KM (2016) Assessment of germplasm diversity and genetic relationships among walnut (*Juglans regia* L.) genotypes through microsatellite markers. J Saudi Soc Agric Sci. DOI: 10.1016/j.jssas.2016.07.005.

Odong TL, Jansen J, van Eeuwijk FA, van Hintum TJL (2013) Quality of core collections for effective utilisation of genetic resources review, discussion and interpretation. Theor Appl Genet 126:289–305.

Odong TL, van Heerwaarden J, Jansen J, van Hintum TJL, van Eeuwijk FA (2011) Determination of genetic structure of germplasm collections: are traditional hierarchical clustering methods appropriate for molecular marker data? Theor Appl Genet 123:195–205.

Ogwu MC, Osawaru ME, Ahana CM (2014) Challenges in conserving and utilizing plant genetic resources (PGR). Int J Genet Mol Biol 6:16–22.

Olukolu BA, Trainin T, Fan S, et al (2009) Genetic linkage mapping for molecular dissection of chilling requirement and budbreak in apricot (*Prunus armeniaca* L.). Genome 52:819–828.

Omrani M, Roth M, Roch G, Blanc A, Morris CE, Audergon JM (2019) Genome-wide association multi-locus and multi-variate linear mixed models reveal two linked loci with major effects on partial resistance of apricot to bacterial canker. BMC Plant Biol 19:31.

Ozaki K, Ohnishi Y, Iida A, Sekine A, Yamada R, Tsunoda T, et al (2002) Functional SNPs in the lymphotoxin-alpha gene that are associated with susceptibility to myocardial infarction. Nat Genet 32:650–654.

Özaktan H, Erdal M, Akkopru A, Aslan E (2008) Evaluation of susceptibility of some walnut cultivars to *Xanthomonas arboricola* pv. *juglandis* by immature nut test. In: Cost 873, WG and Management Committee Meeting, Athens, Greece.

Pan C, Ye L, Qin L, et al (2016) CRISPR/Cas9-mediated efficient and heritable targeted mutagenesis in tomato plants in the first and later generations. Sci Rep 6:24765.

Panth N, Paudel KR, Karki R (2016) Phytochemical profile and biological activity of *Juglans regia*. J Integr Med 14:359–373.

Paterson AH (1996) Making genetic maps. In: Paterson AH (ed) Genome mapping in plants, Academic Press. Austin, Texas.

Paterson AH, Beavis WD (1998) QTL analyses: power, precision, and accuracy. In: Paterson AH (ed) Molecular Dissection of Complex Traits, CRC Press. New York, pp 145–162.

Payghamzadeh K, Kazemitabar SK (2011) *In vitro* propagation of walnut – A review. Afr J Biotechnol 10:290–311.

Peace CP (2017) DNA-informed breeding of rosaceous crops: promises, progress and prospects. Horticulture Research 4:17006.

Peakall R, Smouse PE (2006) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes 6:288–295.

Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics 28:2537–2539.

Peiffer JA, Romay MC, Gore MA, Flint-Garcia SA, Zhang Z, Millard MJ, et al (2014) The genetic architecture of maize height. Genetics 196:1337–1356.

Pereira-Lorenzo S, Ramos-Cabrer AM, Barreneche T, Mattioni C, Villani F, Díaz-Hernández MB, et al (2017) Database of European chestnut cultivars and definition of a core collection using simple sequence repeats. Tree Genet Genomes 13(5).

Peres NA, Timmer LW, Adaskaveg JE, Correll JC (2005) Lifestyles of *Colletotrichum acutatum*. Plant Dis 89:784–796.

Pereira-Lorenzo S, Costa RML, Ramos-Cabrer AM, Ciordia-Ara M, Ribeiro CAM, Borges O, et al (2011) Chestnut cultivar diversification process in the Iberian Peninsula, Canary Islands, and Azores. Genome 54(4):301–315.

Perrier X, Flori A, Bonnot F (2003) Data analysis methods. In: Hamon P, Seguin M, Perrier X, Glaszmann JC (eds) Genetic Diversity of Cultivated Tropical Plants, Science Publishers, Enfield, pp 43-76.

Perrier X, Jacquemoud-Collet J (2006) DARwin software. Available from: http://darwin.cirad.fr/.

Pikkuhookana P, Sillanpää MJ (2014) Combined linkage disequilibrium and linkage mapping: Bayesian multilocus approach. Heredity 112:351–360.

Platt A, Vilhjalmsson BJ, Nordborg M (2010) Conditions under which genome-wide association studies will be positively misleading. Genetics 186:1045–1052.

Poggetti L, Ermacora P, Cipriani G, et al (2017) Morphological and carpological variability of walnut germplasm (*Juglans regia* L.) collected in North-Eastern Italy and selection of superior genotypes. Sci Hortic 225:615–619.

Pollegioni P, Bartoli S, Cannata F, Malvolti ME (2003) Genetic differentiation of four Italian walnut (*Juglans regia* L.) varieties by inter simple sequence repeat (ISSR). J Genet Breed 57:231–240.

Pollegioni P, Van der Linden G, Belisario A, Gras M, Anselmi N, Olimpieri I, et al (2012) Mechanisms governing the responses to anthracnose pathogen in *Juglans* spp. J Biotechnol 159:251–264.

Pollegioni P, Woeste K, Chiocchini F, Del Lungo S, Ciolfi M, Olimpieri I, et al (2017) Rethinking the history of common walnut (*Juglans regia* L.) in Europe: its origins and human interactions. PLoS ONE 12(3):e0172541.

Pollegioni P, Woeste K, Major A, et al (2008) Characterization of *Juglans nigra* (L.), *Juglans regia* (L.) and *Juglans* x *intermedia* (Carr.) by SSR markers : a case study in Italy. Silvae Genet 57:68–78.

Pollegioni P, Woeste K, Mugnozza GS, Malvolti ME (2009) Retrospective identification of hybridogenic walnut plants by SSR fingerprinting and parentage analysis. Mol Breed 24:321–335.

Pollegioni P, Woeste K, Olimpieri I, et al (2011) Long-term human impacts on genetic structure of Italian walnut inferred by SSR markers. Tree Genet Genomes 7:707–723.

Ponnu J, Wahl V, Schmid M (2011) Trehalose-6-phosphate: connecting plant metabolism and development. Front Plant Sci 2:70.

Pop IF, Pamfil D, Raica P, et al (2010) Assessment of the genetic variability among some *Juglans* cultivars from the Romanian National Collection at S.C.D.P. Vâlcea using RAPD markers. Romanian Biotechnol Lett 15:41–49.

Pop IF, Vicol AC, Botu M, et al (2013) Relationships of walnut cultivars in a germplasm collection: Comparative analysis of phenotypic and molecular data. Sci Hortic 153:124–135.

Potter D, Gao F, Aiello G, et al (2002) Intersimple Sequence Repeat Markers for Fingerprinting and Determining Genetic Relationships of Walnut (*Juglans regia*) Cultivars. J Am Soc Hortic Sci 127:75–81.

Powell JE, Visscher PM, Goddard ME (2010) Reconciling the analysis of IBD and IBS in complex trait studies. Nat Rev Genet 11:800–805.

Pretzsch H, Biber P, Schütze G, Kemmerer J, Uhl E (2018) Wood density reduced while wood volume growth accelerated in Central European forests since 1870. For Ecol Manag 429:589–616.

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38:904–909.

Price AL, Weale ME, Patterson N, Myers SR, Need AC, et al (2008) Long-range LD can confound genome scans in admixed populations. Am J Hum Genet 83:132–135.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, et al (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81:559–575.

Qi J, Hao Y, Zhu Y, et al (2011) Studies on Germplasm of *Juglans* by EST-SSR Markers. Acta Hortic Sin 38:441–448.

Qianwen X, Kaizhi W, Lanying Z, et al (2010) RAPD Markers and Heterotic Effect of Walnut Quality in Sichuan of China. Adv Biol Res 4:81–85.

QingGuo M, Jing Q, Dong P (2010) FISH-AFLP analysis of genetic diversity of early-fruiting walnut cultivars. For Res 23:631–636.

Quan S, Niu J, Zhou L, Xu H, Ma L, Qin Y (2019) Genome-wide Identification, Classification, Expression and Duplication Analysis of GRAS Family Genes in *Juglans regia* L. Sci Rep 9:11643.

R Development Core Team (2008) R: A language and environment for statistical computing. Vienna: Austria; R Foundation for Statistical Computing. Available online at: http://www.R-project.org/.

Rahaman MM, Chen D, Gillani Z, Klukas C, Chen M (2015) Advanced phenotyping and phenotype data analysis for the study of plant growth and development. Front Plant Sci 6:619.

Raj A, Stephens M, Pritchard JK (2014) fastSTRUCTURE: Variational Inference of Population Structure in Large SNP Data Sets. Genetics 197:573–598.

Ramain S (2012) Noix noires: la conséquence d'une année au climat atypique. Info Noix 12.

Rao G, Sui J, Zhang J (2016) Metabolomics reveals significant variations in metabolites and correlations regarding the maturation of walnuts (*Juglans regia* L.). Biol Open 5:829–836.

Rascher U, Blossfeld S, Fiorani F, et al (2011) Non-Invasive approaches for phenotyping of enhanced performance traits in bean. Funct Plant Biol 38:968–983.

Rehder A (1947) Manual of cultivated trees and shrubs. In: The MacMillan Co. ed. pp 115–124.

Resende MF Jr, Muñoz P, Acosta JJ, Peter GF, Davis JM, et al (2012) Accelerating the domestication of trees using genomic selection: accuracy of prediction models across ages and environments. New Phytol 193:617–624.

Rezaee R, Vahdati K, Valizadeh M (2009) Variability of seedling vigour in Persian walnut as influenced by the vigour and bearing habit of the mother tree. J Hortic Sci Biotechnol 84:228–232.

Risch N, Merikangas K (1996) The Future of Genetic Studies of Complex Human Diseases. Science 273:1516-1517.

Robichaud RL, Glaubitz JC, Rhodes OE, Woeste KE (2006) A robust set of black walnut microsatellites for parentage and clonal identification. New For 32:179–196.

Rondeau A-S (1997) Le Noyer, Actes Sud.

Roor W, Konrad H, Mamadjanov D, Geburek T (2017) Population Differenciation in Common Walnut (*Juglans regia* L.) across Major Parts of Its Native Range – Insights from Molecular and Morphometric Data. J Hered 108:391–404.

Ros E (2010) Health Benefits of Nut Consumption. Nutrients 2:652–682.

Rosenberg NA (2004) Distruct: a program for the graphical display of population structure. Mol Ecol Notes 4:137–138.

Ross-Davis A, Huang Z, McKenna J, et al (2008) Morphological and molecular methods to identify butternut (*Juglans cinerea*) and butternut hybrids : relevance to butternut conservation. Tree Physiol 28:1127–1133.

Ruiz-Garcia L, Lopez-Ortega G, Denia AF, Tomas DF (2011) Identification of a walnut (*Juglans regia* L.) germplasm collection and evaluation of their genetic variability by microsatellite markers. Span J Agric Res 9:179–192.

Russ JC, Neal FB (2017) The Image Processing Handbook. Raleigh, Seventh Edition. CRC Press.

Saintagne C, Bodénès C, Barreneche T, Pot D, Plomion C, Kremer A (2004) Distribution of genomic regions differentiating oak species assessed by QTL detection. Heredity 92:20–30.

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425.

Salles C (2003) La mythologie grecque et romaine. Hachette Pluriel.

Santamaría M, Hasbún R, Valera M, et al (2009) Acetylated H4 histone and genomic DNA methylation patterns during bud set and bud burst in Castanea sativa. J Plant Physiol 166:1360–1369.

Schell J (1975) The role of plasmids in crown gall formation by *A. tumefaciens*. In: Ledoux L (ed) Genetic manipulation with Plant Materials, Plenum Press, New York, pp. 163–181.

Schlötterer C (2004) The evolution of molecular markers—just a matter of fashion? Nat Rev Genet 5:63–69.

Schnable P, Ware D, Fulton R, Stein J, Wei F, et al (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326:1112–1115.

Schoeman L, Williams P, du Plessis A, Manley M (2016) X-ray micro-computed tomography (μCT) for non-destructive characterisation of food. Trends Food Sci Technol 47:10–24.

Scott KD, Eggler P, Seaton G, Rossetto M, Ablett EM, Lee LS, et al (2000) Analysis of SSRs derived from grape ESTs. Theor Appl Genet 100:723–726.

Sébillot P (1984) Le Folklore de France - La Flore. Editions Imago.

Segura V, Vilhjálmsson BJ, Platt A, Korte A, Seren Ü, et al (2012) An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. Nat Genet 44:825–830.

Semagn K, Bjørnstad A, Ndjiondjop MN (2006) Principles, requirements and prospects of genetic mapping in plants. Afr J Biotechnol 5:2569–2587.

Sen SM (1985) Correlations between shell thickness, shell cracking resistance, shell seal and shell upright cracking resistance and some other fruit quality characters in walnut (*Juglans regia* L.). Doga Bilim Derg D2 Tarm Ve Orman 9:10–24.

Sharma OC (1996) Variability in seedlings trees of Persian walnut in Solan area of Himachal Pardesh. Thesis, Solan University of Horticulture & Forestry, India.

Sharma OC, Sharma SD (2001) Genetic divergence in seedling trees of Persian walnut (*Juglans regia* L.) for various metric nut and kernel characters in Himachal Pradesh. Sci Hortic 88:163–171.

Sharma P, Lin T, Grandellis C, Yu M, Hannapel DJ (2014) The BEL1-like family of transcription factors in potato. J Exp Bot 65:709–723.

Sheikh Beig Goharrizi MA, Dejahang A, Tohidfar M, et al (2016) Agrobacterium Mediated Transformation of Somatic Embryos of Persian Walnut using *fld* Gene for Osmotic Stress Tolerance. J Agr Sci Tech 18:423–435.

Shibayama M, Sakamoto T, Takada E, et al (2011) Estimating paddy rice leaf area index with fixed point continuous observation of near infrared reflectance using a calibrated digital camera. Plant Product Sci 14:30–46.

Sholokhov LV (1974) Quality characters of fruit in walnut and correlation between them. Tr Novocherkas Inzha Melior Inta 13:106–112.

Sideli GM, Marrano A, Montanari S, Leslie CA, Allen BJ, et al (2020) Quantitative phenotyping of shell suture strength in walnut (*Juglans regia* L.) enhances precision for detection of QTL and genome-wide association mapping. PLoS ONE 15(4):e0231144.

Singh N, Choudhury DR, Singh AK, Kumar S, Srinivasan K, et al (2013) Comparison of SSR and SNP markers in estimation of genetic diversity and population structure of Indian rice varieties. PLoS ONE 8(12):e84136.

Smith DL, Gross KC (2000) A family of at least seven β-galactosidase genes is expressed during tomato fruit development. Plant Physiol 123:1173–1183.

Singh RK, Mishra GP, Thakur AK, Singh SB (2008) Molecular markers in plants. In: Singh RK, Singh R, Ye G, et al (eds) Molecular Plant Breeding: Principle, Method and Application, Studium Press LLC. Texas, USA. pp 35–78.

Smith JSC, Smith OS (1989) The description and assessment of distance between inbred lines of maize. 2: The utility of morphological - biochemical - and genetic descriptors and a scheme for the testing of distinctiveness between inbred lines. Maydica 34:151–161.

Solar A, Colaric M, Usenik V, Stampar F (2006) Seasonal variations of selected flavonoids, phenolic acids and quinones in annual shoots of common walnut (*Juglans regia* L.). Plant Sci 170:453–461.

Solar A, Hudina M, Stampar F (2001) Relationship between tree architecture, phenological data and generative development in walnut (*Juglans regia* L.). Acta Hortic 544:275–285.

Solar A, Smole J, Stampar F, Viršček-Marn M (1994) Characterization of isozyme variation in walnut (*Juglans regia* L.). Euphytica 77:105–112.

Solar A, Stampar F (2011) Characterisation of selected hazelnut cultivars: phenology, growing and yielding capacity, market quality and nutraceutical value. J Sci Food Agric 91:1205–1212.

Soltani J, Aliabadi AA (2010) Genetic resistance to bacterial blight disease in Persian walnut. Eur J Plant Pathol 128:65–70.

Stanford AM, Harden R, Parks CR (2000) Phylogeny and biogeography of *Juglans* (*Juglandaceae*) based on matK and ITS sequence data. Am J Bot 87:872–882.

Stamp P, Visser R (2012) The twenty-first century, the century of plant breeding. Euphytica 186:585–591.

Steinwand BJ, Kieber JJ (2010) The Role of Receptor-Like Kinases in Regulating Cell Wall Function. Plant Physiol 153:479–484.

Sun X, Liu D, Zhang X, et al (2013) SLAF-seq: An Efficient Method of Large-Scale *De Novo* SNP Discovery and Genotyping Using High-Throughput Sequencing. PLoS ONE 8:e58700.

Sütyemez M (2006) Comparison of AFLP Polymorphism in Progeny Derived from Dichogamous and Homogamous Walnut Genotypes. Pak J Biol Sci 9:2303–2307.

Sütyemez M (2016) New Walnut Cultivars: Maras 18, Sütyemez 1, and Kaman 1. HortScience 51:1301–1303.

Taha NA, Al-Wadaan MA (2011) Utility and importance of walnut, *Juglans regia* Linn: A review. Afr J Microbiol Res 5:5796–5805.

Tai L, Li BB, Nie XM, Zhang PP, Hu CH, Zhang L, et al (2019) Calmodulin Is the Fundamental Regulator of NADK-Mediated NAD Signaling in Plants. Front Plant Sci 10:681.

Tal O, Kisdi E, Jablonka E (2010) Epigenetic Contribution to Covariance Between Relatives. Genetics 184:1037–1050.

Tapsell LC, Batterham MJ, Teuss G, et al (2009) Long-term effects of increased dietary polyunsaturated fat from walnuts on metabolic parameters in type II diabetes. Eur J Clin Nutr 63:1008–1015.

Tester M, Langridge P (2010) Breeding Technologies to Increase Crop Production in a Changing World. Science 327:818–822.

Thornton T, Tang H, Hoffmann TJ, Ochs-Balcom HM, Caan BJ, Risch N (2012) Estimating Kinship in Admixed Populations. Am J Hum Genet 91:122–138.

Tian C, Gregersen PK, Seldin MF (2008) Accounting for ancestry: population substructure and genome-wide association studies. Hum Mol Genet 17(R2):R143–R150.

Tian J (ed) (2014) Proceedings of the VIIth International Walnut Symposium: Fenyang, China, July 20-23, 2013. Fenyang, China. Acta Hort 1050.

Tollner EW, Hung YC, Upchurch BL, Prussia SE (1992) Relating X-ray absorption to density and water content in apples. Trans ASAE 35:1921–1928.

Topçu H, Ikhsan AS, Sütyemez M, et al (2015) Development of 185 polymorphic simple sequence repeat (SSR) markers from walnut (*Juglans regia* L.). Sci Hortic 194:160–167.

Torres AR, Grunvald AK, Martins TB, dos Santos MA, Lemos NG, Silva LAS, et al (2015) Genetic structure and diversity of a soybean germplasm considering biological nitrogen fixation and protein content. Sci Agric 72:47–52.

Torres-Sánchez J, López-Granados F, Serrano N, Arquero O, Peña JM (2015) High-Throughput 3-D Monitoring of Agricultural-Tree Plantations with Unmanned Aerial Vehicle (UAV) Technology. PLoS ONE 10(6):e0130479.

Tsiantos J, Vagelas IK, Rumbos CI, Chatzaki A, Rouskas D, Gravanis FT (2008) Evaluation of resistance of cultivated walnut varieties and selections to *Xanthomonas arboricola* pv. *juglandis* in Greece. In: Cost 873, WG and Management Committee Meeting, Athens, Greece.

Tsykun T, Rellstab C, Dutech C, Sipos G, Prospero S (2017) Comparative assessment of SSR and SNP markers for inferring the population genetic structure of the common fungus *Armillaria cepistipes*. Heredity 119:371–380.

Tulecke W, McGranahan G (1985) Somatic embryogenesis and plant regeneration from cotyledons of walnut, *Juglans regia* L. Plant Sci 40:57–63.

Tulecke W, McGranahan G (1994) The Walnut Germplasm Collection of the University of California, Davis. A Description of the Collection and a History of the Breeding Program of Eugene F. Serr and Harold I. Forde. Report No. 13. DANR, UC Davis, CA.

Turner S, Armstrong LL, Bradford Y, Carlson CS, Crawford DC, et al (2011) Quality Control Procedures for Genome Wide Association Studies. Curr Protoc Hum Genet Chapter 1:Unit1.19.

Ulukan H (2011) The use of plant genetic resources and biodiversity in classical plant breeding. Acta Agric Scand B 61:97–104.

Urrestarazu J, Muranty H, Denancé C, et al (2017) Genome-Wide Association Mapping of Flowering and Ripening Periods in Apple. Front Plant Sci 8:1923.

US Department of Agriculture, Agricultural Research Service, Nutrient Data Laboratory (2015) USDA National Nutrient Database for Standard Reference, Release 28. https://ndb.nal.usda.gov/ndb/. Accessed 9 Dec 2016.

Vahdati K, Arab MM, Sarikhani S, Sadat-Hosseini M, Leslie CA, Brown PJ (2019) Advances in Persian Walnut (*Juglans regia* L.) Breeding Strategies. In: Al-Khayri J, Jain SM, Johnson DV (eds) Advances in Plant Breeding Strategies: Nut and Beverage Crops, Springer, pp 401–472.

Vahdati K, Massah Bavani AR, Khosh-Khui M, Fakour P, Sarikhani S (2019) Applying the AOGCM-AR5 models to the assessments of land suitability for walnut cultivation in response to climate change: A case study of Iran. PLoS ONE 14(6): e0218725.

Vahdati K, McKenna J, Dandekar AM, et al (2002) Rooting and Other Characteristics of a Transgenic Walnut Hybrid (*Juglans hindsii* x *J. regia*) Rootstock Expressing rolABC. J Am Soc Hortic Sci 127:724–728.

Van Inghelandt D, Melchinger AE, Lebreton C, Stich B (2010) Population structure and genetic diversity in a commercial maize breeding program assessed with SSR and SNP markers. Theor Appl Genet 120:1289–99.

Van Ooijen JW (2006) JoinMap® 4, Software for the calculation of genetic linkage maps in experimental populations. Wageningen: Netherlands; Kyazma B.V.

Vargas AM, de Andrés MT, Ibáñez J (2016) Maximization of minority classes in core collections designed for association studies. Tree Genet Genomes 12:28.

Verhaeghe A, Giraud M (2015) Anthracnose à *Colletotrichum* sp. en verger de noyers. Arboriculture Fruitière 695.

Verrier E, Brabant P, Gallais A (2001) Faits et concepts de base en génétique quantitative. Polycopié AgroParisTech, p 134.

Victory ER, Glaubitz JC, Rhodes Jr OE, Woeste KE (2006) Genetic homogeneity in *Juglans nigra* (*Juglandaceae*) at nuclear microsatellites. Am J Bot 93:118–126.

Vidhya M, Varadharaju N, John Kennedy Z, Amirtham D, Manohar Jesudas D (2017) Applications of X-Ray Computed Tomography in Food Processing. J Food Process Technol 8:673.

Vigouroux Y, Glaubitz JC, Matsuoka Y, Goodman MM, Sánchez G J, Doebley J (2008) Population structure and genetic diversity of New World maize races assessed by DNA microsatellites. Am J Bot 95(10):1240–1253.

Vimont N, Beauvieux R, Arkoun M, et al (2019) The presence of H3K4me3 histone mark is positively correlated with the expression at the DAM loci in sweet cherry during dormancy. Acta Hortic 1235:413–420.

Vischi M, Chiabà C, Raranciuc S, et al (2017) Genetic Diversity of Walnut (*Juglans regia* L.) in the Eastern Italian Alps. Forests 8:81–94.

Vitasse Y, François C, Delpierre N, et al (2011) Assessing the effects of climate change on the phenology of European temperate trees. Agric For Meteorol 151:969–980.

Voorrips RE (2002) MapChart: Software for the graphical presentation of linkage maps and QTLs. J Hered 93:77–78.

Vos P, Hogers R, Bleeker M, et al (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23:4407–4414.

Vyas D, Sharma SK, Sharma DR (2003) Genetic structure of walnut genotype using leaf isozymes as variability measure. Sci Hortic 97:141–152.

Wainberg M, Sinnott-Armstrong N, Mancuso N, Barbeira AN, Knowles DA, et al (2019) Opportunities and challenges for transcriptome-wide association studies. Nat Genet 51:592–599.

Walawage SL, Britton MT, Leslie CA, et al (2013) Stacking resistance to crown gall and nematodes in walnut rootstocks. BMC Genomics 14:668.

Wambugu PW, Ndjiondjop MN, Henry RJ (2018) Role of genomics in promoting the utilization of plant genetic resources in genebanks. Brief Funct Genomics 17:198–206.

Wang DG, Fan JB, Siao CJ, et al (1998) Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science 280:1077–1082.
Wang G, Wang JY, He Q, et al (2014a) Agronomic Evaluation and Heritability of Jin RS-2 and Jin RS-3 Walnut Rootstocks. Acta Hortic 1050:113–121.
Wang GA, Zhang Q, Huang MM, Yakup A (2014b) The Breeding of Six Xinjiang Dwarf Walnut Cultivars. Acta Hortic 1050:151–159.
Wang H, Pei D, Gu R, Wang B (2008) Genetic Diversity and Structure of Walnut Populations in Central and Southwestern China Revealed by Microsatellite Markers. J Am Soc Hortic Sci 133:197–203.
Wang H, Wu W, Pan G, Pei D (2016) Analysis of genetic diversity and relationships among 86 Persian walnut (*Juglans regia* L.) genotypes in Tibet using morphological traits and SSR markers. J Hortic Sci Biotechnol 90:563–570.
Wang H, Zhao S, Zhang Z, et al (2010) Genetic relationship and diversity of eight *Juglans* species in China estimated through AFLP analysis. Acta Hortic 861:143–150.
Wang J, Zhang Z (2019) GAPIT Version 3: An Interactive Analytical Tool for Genomic Association and Prediction. Bioinformatics, preprint.
Wang SB, Feng JY, Ren WL, Huang B, Zhou L, et al (2016) Improving power and accuracy of genome-wide association studies via multi-locus mixed linear model methodology. Sci Rep 6:19444.
Wani N, Bhat MA, Ahmad MF, et al (2010) Molecular markers and their application in walnut improvement. Int J Curr Res 3:6–11.
Warren R, Price J, VanDerWal J, et al (2018) The implications of the United Nations Paris Agreement on climate change for globally significant biodiversity areas. Climatic Change 147:395–409.
Way DA (2011) Tree phenology responses to warming: spring forward, fall back? Tree Physiol 31:469–471.
Wei T, Simko VR (2017) Package "corrplot": Visualization of a Correlation Matrix version 0.84. https://github.com/taiyun/corrplot.
Wharton PS, Diéguez-Uribeondo J (2004) The biology of *Colletotrichum acutatum*. 4 An Jardín Botánico Madr 61:3–22.
Wickham H (2017) Tidyverse: Easily Install and Load the 'Tidyverse'. R package version 1.2.1. https://CRAN.R-project.org/package=tidyverse.
Williams JG, Kubelik AR, Livak KJ, et al (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res 18:6531–6535.
Wilson EO (1988) Biodiversity. National Academy Press, Washington, D.C.
Woeste K, Burns R, Rhodes O, Michler C (2002) Thirty Polymorphic Nuclear Microsatellite Loci From Black Walnut. J Hered 93:58–60.
Woeste K, McGranahan G, Bernatzky R (1996a) The identification and characterization of a genetic marker linked to hypersensitivity to the cherry leafroll virus in walnut. Mol Breed 2:261–266.
Woeste K, McGranahan G, Bernatzky R (1998) Low Correlation Between Genomic and Morphological Introgression Estimates in a Walnut Backcross. J Am Soc Hortic Sci 123:258–263.
Woeste K, McGranahan GH, Bernatzky R (1996b) Randomly Amplified Polymorphic DNA Loci from a Walnut Backcross [(*Juglans hindsii* x *J. regia*) x *J. regia*]. J Am Soc Hortic Sci 121:358–361.
Wolfe JA (1978) A Paleobotanical Interpretation of Tertiary Climates in the Northern Hemisphere: Data from fossil plants make it possible to reconstruct Tertiary climatic changes, which may be correlated with changes in the inclination of the earth's rotational axis. Am Sci 66:694–703.
Wong CK, Bernardo R (2008) Genome wide selection in oil palm: increasing selection gain per unit time and cost with small populations. Theor Appl Genet 116:815–824.
Woodroof JG (1967) Tree nuts: production, processing, products. AVI Pub. Co.
Woodworth RH (1930) Meiosis of micro-sporogenesis in the *Juglandaceae*. Am J Bot 17:863–869.
Wu J, Gu YQ, Hu Y, et al (2012) Characterizing the walnut genome through analyses of BAC end sequences. Plant Mol Biol 78:95–107.
Xiao Y, Liu H, Wu L, Warburton M, Yan J (2017) Genome-wide Association Studies in Maize: Praise and Stargaze. Mol Plant 10:359–374.
Xie L, Zhang JZ, Wan Y, Hu DW (2010) Identification of *Colletotrichum* spp. isolated from strawberry in Zhejiang Province and Shanghai City, China. J Zhejiang Univ Sci B 11(1):61–70.
Xu F, Deng G, Cheng S, et al (2012a) Molecular cloning, characterization and expression of the phenylalanine ammonia-lyase gene from *Juglans regia*. Molecules 17:7810–7823.
Xu L, Chen X, Zhang LS, Liu QZ (2014) Molecular Cloning and Expression Analysis of the Transcription Factor Gene *JrCBF* from *Juglans regia* L. Acta Hortic 1050:41–47.
Xu Y, Yang T, Zhou Y, et al (2018) Genome-Wide Association Mapping of Starch Pasting Properties in Maize Using Single-Locus and Multi-Locus Models. Front Plant Sci 9:1311.
Xu Z, Hu T, Zhang F (2012b) Genetic Diversity of Walnut Revealed by AFLP and RAPD Markers. J Agric Sci 4:271–276.
Yang H, Liu J, Dang M, Zhang B, Li H, Meng R, et al (2018) Analysis of β-Galactosidase During Fruit Development and Ripening in Two Different Texture Types of Apple Cultivars. Front Plant Sci 9:539.

Yang X, Xu Y, Shah T, Li H, Han Z, et al (2011) Comparison of SSRs and SNPs in assessment of genetic relatedness in maize. Genetica 139:1045–1054.

Yang Z, Wang Y, Zhou Y, Gao Q, Zhang E, Zhu L, et al (2013) Evolution of land plant genes encoding L-Ala-D/L-Glu epimerases (AEEs) via horizontal gene transfer and positive selection. BMC Plant Biol 13:34.

Yarilgac T, Koyuncu F, Koyuncu MA, et al (2001) Some promising walnut selections (*Juglans regia* L.). Acta Hortic 544:93–96.

Yeung KF, Yang Y, Yang C, Liu J (2019) CoMM: A Collaborative Mixed Model That Integrates GWAS and eQTL Data Sets to Investigate the Genetic Architecture of Complex Traits. Bioinform Biol Insights 13:1–6.

Yi F, Zhijun Z, Shelong Z (2011) Development of walnut EST-SSR markers and primer design. Agric Sci Technol 12:1810–1813.

You FM, Deal KR, Wang J, et al (2012) Genome-wide SNP discovery in walnut with an AGSNP pipeline updated for SNP discovery in allogamous organisms. BMC Genomics 13:354.

Yu J, Holland JB, McMullen MD, Buckler ES (2008) Genetic Design and Statistical Power of Nested Association Mapping in Maize. Genetics 178: 539–551.

Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, et al (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38:203–208.

Zekiri F, Molitor C, Mauracher SG, et al (2014) Purification and characterization of tyrosinase from walnut leaves (*Juglans regia*). Phytochemistry 101:5–15.

Zeneli G, Kola H, Dida M (2005) Phenotypic variation in native walnut populations of Northern Albania. Sci Hortic 105:91–100.

Zerillo MM, Ibarra Caballero J, Woeste K, et al (2014) Population Structure of *Geosmithia morbida*, the Causal Agent of Thousand Cankers Disease of Walnut Trees in the United States. PLoS ONE 9:e112847.

Zeven AC, Zhukovsky PM (1975) Dictionary of cultivated plants and their centres of diversity excluding ornamentals, forest trees, and lower plants. Centre for Agricultural Publishing and Documentation, Wageningen.

Zhang MY, Xu Y, Xiang K, Wang XL (2014) Review of Walnut Breeding Research at the Shandong Institute of Pomology. Acta Hortic 1050:55–60.

Zhang Q, Walawage SL, Tricoli DM, et al (2015) A red fluorescent protein (DsRED) from *Discosoma* sp. as a reporter for gene expression in walnut somatic embryos. Plant Cell Rep 34:861–869.

Zhang R, Zhu A, Wang X, et al (2010) Development of *Juglans regia* SSR Markers by Data Mining of the EST Database. Plant Mol Biol Report 28:646–653.

Zhang YM, Jia Z, Dunwell JM (2019) Editorial: The Applications of New Multi-Locus GWAS Methodologies in the Genetic Dissection of Complex Traits. Front Plant Sci 10:100.

Zhang ZY, Han JW, Jin Q, et al (2013) Development and characterization of new microsatellites for walnut (*Juglans regia*). Genet Mol Res 12:4723–4734.

Zhang Z, Ersoz E, Lai CQ, Todhunter RJ, Tiwari HK, Gore MA, et al (2010) Mixed linear model approach adapted for genome-wide association studies. Nat Genet 42:355–360.

Zhao K, Aranzana MJ, Kim S, Lister C, Shindo C, et al (2007) An *Arabidopsis* example of association mapping in structured samples. PLoS Genet 3(1):e4.

Zhao MH, Jiang ZT, Liu T, Li R (2014) Flavonoids in *Juglans regia* L. Leaves and Evaluation of In Vitro Antioxidant Activity via Intracellular and Chemical Methods. ScientificWorldJournal 2014 article id 303878.

Zhao ML, Ni J, Chen MS, Xu ZF (2019) Ectopic Expression of *Jatropha curcas TREHALOSE-6-PHOSPHATE PHOSPHATASE J* Causes Late-Flowering and Heterostylous Phenotypes in *Arabidopsis* but not in *Jatropha*. Int J Mol Sci 20:2165.

Zhao P, Zhang T, Zhou HJ (2015) Identification, development, and application of 40 polymorphic EST-SSR markers for common walnut (*Juglans regia* L.). NCBI, EST unpublished raw data.

Zheng X, Levine D, Shen J, Gogarten S, Laurie C, Weir B (2012) A High-performance Computing Toolset for Relatedness and Principal Component Analysis of SNP Data. Bioinformatics 28:3326–3328.

Zhou X, Stephens M (2013) Genome-wide Efficient Mixed Model Analysis for Association Studies. Nat Genet 44:821–824.

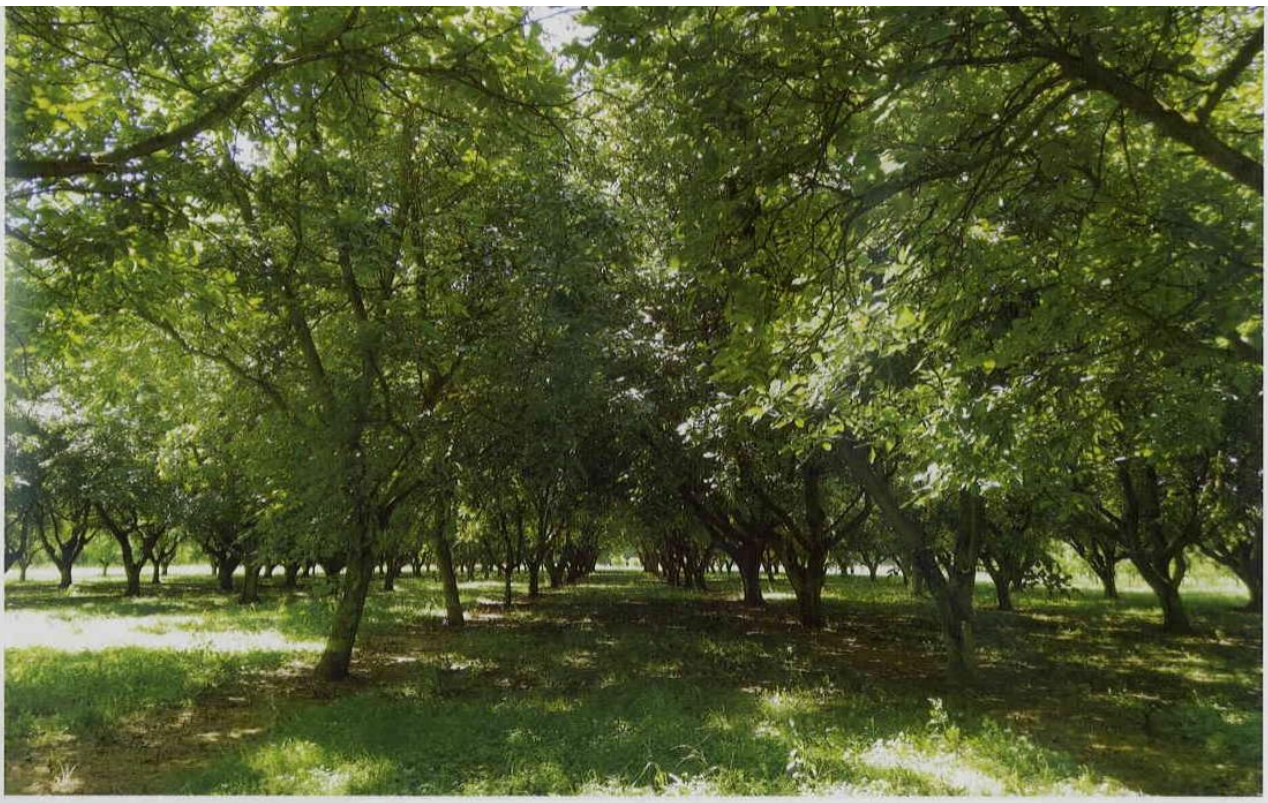
Zhu C, Gore M, Buckler ES, Yu J (2009) Status and prospects of association mapping in plants. Plant Genome 1:5–20.

Zhu Y, Yin Y, Yang K, et al (2015) Construction of a high-density genetic map using specific length amplified fragment markers and identification of a quantitative trait locus for anthracnose resistance in walnut (*Juglans regia* L.). BMC Genomics 16:614.

Zietkiewicz E, Rafalski A, Labuda D (1994) Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. Genomics 20:176–183.

Annexes

Annexe 1 : Article « La collection de noyers de l'INRA, un réservoir de diversité essentiel pour l'avenir de la filière nucicole » dans la revue trimestrielle « Hommes et Plantes ».



La collection de noyers de l'INRA
un réservoir de diversité essentiel pour l'avenir de la filière nucicole

Arbre de croyances et de légendes, le noyer méritait bien ce petit tour d'horizon de sa riche diversité, de la situation de la filière nucicole en France et de l'apport actuel et à venir des biotechnologies…

Par Anthony Bernard

ANTHONY BERNARD
Après un Master à l'université de Bordeaux en Biologie et Biotechnologies des plantes, et quelques travaux en biologie moléculaire et génétique sur la laitue et le cerisier, il entame un doctorat portant sur l'étude des ressources génétiques du noyer.

Preuve de sa robustesse, il n'est pas rare de rencontrer un vénérable noyer au détour d'un chemin ou au beau milieu d'un champ. Il symbolise à lui seul le mois d'octobre, quand son fruit nous enchante par son parfum et sa texture uniques. Avant de décrire quelques mythes et légendes autour de cet arbre, intéressons-nous à l'origine des mots « noix » et *Juglans*. La noix vient du mot latin *nux*, qui désigne tout fruit à coque et à amande, ce qui explique le x muet dont nous avons hérité. Le noyer commun est également appelé en vieux français goguier, noguier ou gojeutier. Le nom latin du genre, *Juglans*, proviendrait de la contraction de *Jovis glans*, le gland de Jupiter. Dès l'antiquité, le noyer a fait l'objet de nombreuses légendes. La mythologie grecque nous conte que Dionysos s'est épris de la plus jeune fille du roi Dion, Carya, dont les deux sœurs sont transformées en rochers en raison de leur jalousie. Plus tard, Carya meurt de tristesse et est immortalisée par Dionysos en un noyer fécond. Son sanctuaire se trouvait à Karyai, près de Sparte. Elle cède ensuite sa place à Artémis qui, lorsqu'elle n'est pas la déesse de la chasse ou de la fécondité, est la maîtresse du noyer et des arbres portant des fruits, sous le nom d'Artémis *Karyatis*. Un temple lui est dédié et des colonnes en noyer, taillées aux formes de la déesse, sont érigées pour soutenir le chapiteau : ce sont les fameuses caryatides ou cariatides, que l'on retrouve également en pierre à l'entrée de l'Erechthéion de l'Acro-

JUGLANS, UN GENRE PLUTÔT HOMOGÈNE

Des preuves fossiles font remonter l'origine du genre *Juglans* à l'Éocène, il y a environ 40 millions d'années. L'expansion et la diversification du genre *Juglans* depuis l'Amérique du Nord se sont réalisées lorsque le détroit de Béring était encore un pont terrestre entre les deux continents, la Béringie. Le genre s'est adapté au climat froid de la période Miocène, il y a 10 millions d'années, avant de continuer son expansion en Asie et en Amérique centrale, ce qui explique que l'on trouve des noyers aussi bien dans le Yunnan qu'au Mexique.

Le genre *Juglans* abrite plus de vingt espèces réparties en trois sections botaniques. C'est la section *Dioscaryon* qui nous intéresse principalement puisqu'elle comprend le noyer qui produit la noix telle qu'on la connaît sur les étals des marchés à l'automne, *Juglans regia*. Il possède généralement des feuilles constituées de sept à neuf folioles lancéolées, très peu ou pas dentées, glabres, avec un fruit possédant des cloisons minces et un brou déhiscent. Largement propagé par les hommes au cours de leurs migrations, le noyer se retrouve dans une grande partie de l'hémisphère nord depuis le continent américain jusqu'en Asie, mais aussi en Afrique du Sud, en Australie et en Nouvelle-Zélande où il a été introduit plus récemment. Cet arbre pouvant atteindre vingt à vingt-cinq mètres de haut se reconnaît à son écorce gris clair et à ses feuilles composées et imparipennées. Les fleurs mâles sont regroupées en chatons tandis que les fleurs femelles sont souvent par paires à l'extrémité des rameaux. Son fruit est une drupe contenant un noyau qui renferme une amande, le cerneau. Les autres noyers appartenant à cette section se répartissent dans une vaste zone allant de l'Iran jusqu'au Vietnam en passant par la Chine, et certains y sont récoltés ou cultivés pour leurs fruits comestibles.

Juglans regia, in Michaux, The north american sylva, 1817-1819, DR).

La section botanique *Rhysocaryon* comporte environ seize espèces dont *Juglans nigra*, qui vit spontanément à l'est des États-Unis, et a d'abord été utilisé en France pour l'ornement ou l'ombrage dans les jardins où il s'est montré plus résistant aux froids extrêmes que le noyer commun. Il peut s'hybrider avec ce dernier, et le premier hybride spontané a été signalé en France dans les pépinières du jardin de Trianon, vers 1830. Le nom de *nigra* vient de son écorce et de son bois à la couleur foncée. Ce noyer et ses hybrides sont de nos jours principalement utilisés comme bois d'ameublement mais leurs noix sont également appréciées pour la confection de glaces et de cookies aux Etats-Unis... à condition de les récolter avant les écureuils. Autre membre de cette section, *Juglans mollis*, natif du Mexique, est mal connu. Il possède en moyenne entre 14 et 23 folioles lancéolées. *Juglans microcarpa*, originaire du Texas et du Nord-Est du Mexique, est un arbre buissonnant. Il dépasse rarement les dix mètres de hauteur, là où *J. nigra* peut atteindre quarante mètres. Ses noix possèdent un diamètre moyen d'un centimètre et demi. Il peut s'hybrider facilement avec *J. nigra*.

La section asiatique *Cardiocaryon* contient quant à elle notamment *Juglans mandshurica*, originaire de Mandchourie et du Corée, qui est utilisé comme porte-greffe en Chine et se remarque au premier coup d'œil à ses feuilles qui peuvent atteindre jusqu'à 90 centimètres de long ; et *Juglans sieboldiana*, natif du Japon, dont les noix se présentent en grappes de quinze à vingt fruits, ovoïdes, pointus et lisses. Il est très précoce. Les noyers de cette section présentent des noix en forme de cœur et des feuilles à la pilosité marquée.

Juglans cathayensis

Juglans major

Juglans nigra

Juglans cinerea

UN APERÇU SUR D'AUTRES ESPÈCES APPARTENANT AU GENRE JUGLANS

Ces vues de fruits de plusieurs espèces apparentées au noyer cultivé et présentes dans la collections de l'INRA permettent de juger de la diversité du genre Juglans.

Juglans mandshurica

Juglans hindsii

Juglans mollis

Juglans sieboldiana var. *cordiformis*

Le feuillage du noyer cultivé présente en règle générale une feuille composée, imparipennée avec sept folioles lancéolées. La collection de l'INRA possède quelques accessions aux folioles laciniées irrégulièrement découpées, ainsi que des noyers au feuillage pourpre, tous deux à fort intérêt ornemental (à droite, ph. Anthony Bernard ; ci-dessus, *Juglans squamosa*, Michaux, The north american sylva).

pole d'Athènes mais aussi dans beaucoup de bâtiments classiques comme au Louvre.

Le noyer est tantôt nommé « arbre de vie » et tantôt « arbre de mort », une ambivalence qui se retrouve dans les superstitions qu'il suscite. Au IV[e] siècle, saint Augustin fait de la noix un symbole du Christ : le brou serait l'enveloppe charnelle de Jésus, la coque évoquerait la croix et le cerneau représenterait la divinité cachée du Christ. Selon une légende slave, les noix sauvent du déluge les personnes vertueuses qui se sauvent sur une coquille de noix. Mais il est dit également que dans l'Italie du Moyen Age, les sorcières du monde entier se réunissaient à Bénévent sous un noyer séculaire pour y célébrer le Sabbat. De nos jours, on entend encore couramment que « rien ne pousse sous un noyer » ou que « dormir sous un noyer peut entraîner un mal de tête ». Il est donc clair que les croyances autour du noyer ne semblent pas prêtes de s'essouffler… Ces superstitions n'ont pas empêché d'utiliser tous les produits du noyer : outre l'amande précieuse pour son huile, le bois permettait de fabriquer les meilleurs sabots et on se servait du brou de noix comme colorant pour fabriquer des encres ou comme teinture du bois et des cuirs. Encore aujourd'hui, le bois du noyer est très apprécié en ébénisterie et pour confectionner les crosses de fusil de luxe, et les fruits encore verts sont la base de vins de noix réputés.

Le noyer figure sur le livre d'heures d'Anne de Bretagne (ci-dessous) ainsi que sur le très ancien *Tacuinum sanitatis*, dérivé d'un traité médical arabe du XI[e] siècle (doc. DR).

Le noyer cultivé, *Juglans regia*, serait originaire d'Asie centrale, de la zone des contreforts de l'Himalaya depuis la région du Cachemire jusqu'à l'actuel Kirghizistan. Il s'est d'abord disséminé en Ouzbékistan, au nord de l'Iran, dans la région du Caucase et à l'est de la Turquie, pour ensuite atteindre le nord de l'Inde, le Népal et l'ouest de la Chine. Son introduction en Europe remonterait à la dernière période glaciaire, il y a 50 000 ans environ. Des preuves archéologiques retrouvées au Périgord laissent penser que l'homme de Cro-Magnon aurait pu consommer des formes ancestrales du fruit de *Juglans regia*, il y a 17 000 ans. La dernière grande glaciation a raison du noyer en Europe, hormis dans le sud du continent, en Grèce, Italie et Espagne. Ailleurs, il ne réapparaît que vers 4 000 ans avant notre ère, apporté par les hommes depuis l'Asie occidentale. Sous la Grèce Antique, la noix est devenue plus abondante grâce à Alexandre le Grand et sa dispersion s'est accentuée ensuite notamment en Espagne, au Portugal et en Afrique du Nord, car il semble bien que les Romains appréciaient la noix et plantaient le noyer de façon assez systématique autour de leurs villas. Au XVI[e] siècle, la conquête du Nouveau Monde a permis son arrivée en Amérique du Sud, notamment au Chili, avant son introduction en Californie au XIX[e] siècle.

La production mondiale de noix ne cesse de

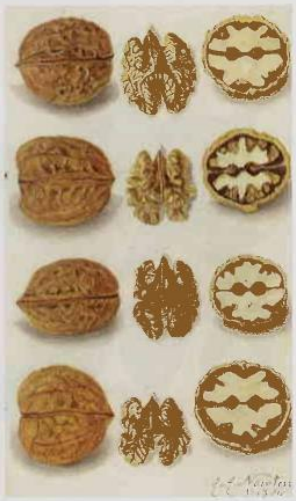
croître depuis le début des années 1970 et a dépassé les 3,5 millions de tonnes en 2016. La France a produit plus de 40 000 tonnes en 2016, ce qui la place en septième position au niveau mondial, et en première place en Europe devant la Roumanie. Si le marché mondial est dominé par la Chine, les États-Unis et l'Iran, la noix française résiste néanmoins à la concurrence grâce à ses qualités organoleptiques reconnues. Il existe deux appellations d'origine contrôlée qui correspondent aux deux principales régions de production, les AOC « Noix du Périgord » et « Noix de Grenoble ». La France exporte 80 % de sa production, principalement en Europe et majoritairement destinée au marché de la noix de table. Environ 70 % de la production est apportée par la variété ancienne 'Franquette' et les surfaces allouées au noyer ont progressé de 19 % entre 2000 et 2010, pour atteindre environ 20 000 hectares. Ce développement en fait la deuxième culture fruitière juste derrière la pomme, avec plus de 1 800 000 arbres.

Le premier programme d'amélioration en France a été mené par Éric Germain, ingénieur à l'INRA de Bordeaux de 1977 à 1995. Il avait alors réalisé 28 croisements, principalement entre des variétés anciennes françaises comme 'Franquette', 'Grandjean' et 'Marbot', appréciées pour leur débourrement tardif et leur qualité, et des variétés californiennes comme 'Pedro' et 'Chandler', connues pour leur capacité à fructifier tout le long du rameau. Il existe en effet deux grands types de fructification chez le noyer : en position terminale, à l'extrémité du rameau de l'année, comme c'est le cas des variétés anciennes françaises ; et en position latérale, tout le long du rameau d'un an. C'est le cas de la variété américaine 'Payne', présente dans le pedigree de la majorité des variétés américaines. Deux cultivars à fructification latérale ont été inscrits en 1995 au catalogue officiel des espèces et variétés de plantes cultivées : 'Fernette' et 'Fernor', issus d'un croisement entre 'Franquette' et 'Lara'®, ce dernier possédant 'Payne' dans son pedigree. Un troisième cultivar à fructification latérale et à cerneaux assez petits, 'Ferjean', issu du

Chez *Juglans cinerea*, les feuilles peuvent atteindre 70 centimètres de long et comporter jusqu'à 17 folioles. Atteint par un chancre, ce noyer fait désormais partie des espèces en voie de disparition au Canada (Michaux, The north american sylva, DR).

DES NOIX VENUES DU MONDE ENTIER

Ces échantillons d'accessions permettent de constater une grande diversité de forme et de taille de noix.



Sur cette aquarelle, la variété 'Mayette' apparaît en deuxième ligne (coll. USDA, DR).

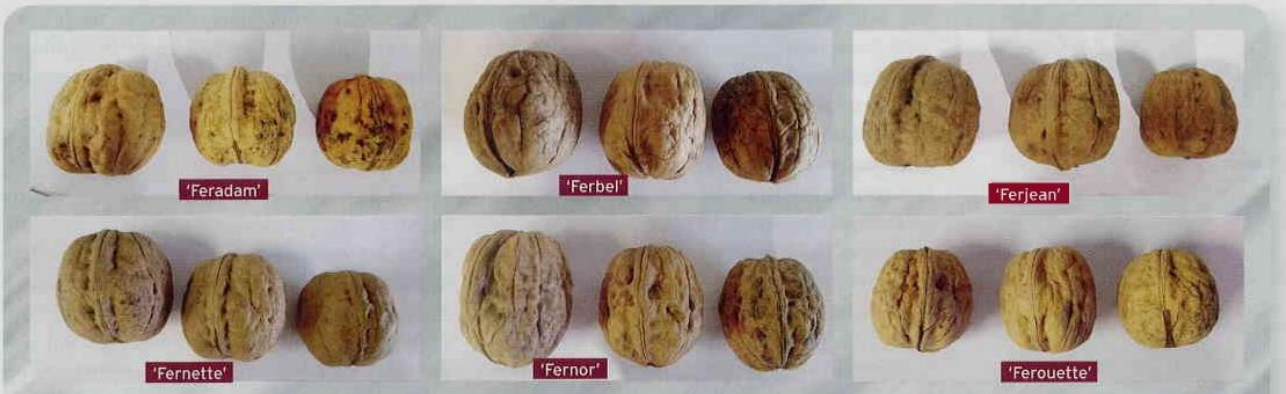
croisement entre 'Grosvert' et 'Lara'®, a été inscrit en 1999. Il est destiné à remplacer 'Grandjean' pour produire du cerneau. Enfin, les quatre derniers cultivars sont sortis en 2010 : 'Feradam', 'Ferbel', 'Ferouette' et 'Fertignac'.

Le second programme d'amélioration, entre 1996 et 2007, a été conduit par le même Éric Germain accompagné de Francis Delort, en utilisant cette fois la collection de ressources génétiques dont il est question dans cet article. Les hybrides les plus intéressants ont été croisés avec du matériel végétal provenant de différentes régions du monde, comme le pourtour méditerranéen, le Moyen-Orient, l'Asie centrale et le Japon. Cette diversification avait pour but de réduire les problèmes de consanguinité précédemment observés au sein des croisements entre matériels français et américain. Si quelques hybrides prometteurs sont à ce jour caractérisés et évalués en stations expérimentales, le programme d'amélioration a pris fin en 2007 suite à la décision de l'INRA de restreindre globalement le nombre d'espèces sur lesquelles l'institut développe des programmes de recherche de génétique et de sélection variétale.

La collection française de noyers est cependant maintenue par l'INRA de Nouvelle-Aquitaine-Bordeaux, plus précisément au Centre de Ressources Génétiques *Prunus-Juglans*, au sein de l'Unité Expérimentale Arboricole, dans les sites de Toulenne et de Bourran, respectivement en Gironde et en Lot-et-Garonne. Cette collection résulte d'un important travail de prospection réalisé par Éric Germain entre 1988 et 2000 dans de nombreux pays du monde. Conscient de l'importance de la sauvegarde de la biodiversité, ce chercheur savait que la création d'un réservoir de diversité génétique constitue une base essentielle dans l'initiation d'un programme d'amélioration. Ainsi, cette collection d'une grande richesse rassemble une majeure partie des espèces du genre *Juglans*. Elle possède plus de 400 individus au sein de douze espèces du genre *Juglans*. Pour le seul noyer cultivé, *Juglans regia*, la collection possède près de deux cents variétés provenant du monde entier.

UN BEL ENSEMBLE DE VARIÉTÉS FRANÇAISES

Cette petite partie de la riche collection INRA montre l'importance du noyer dans l'économie agricole d'autrefois, pourvoyeur de bois mais aussi de lumière grâce à son huile.



LES VARIÉTÉS DE L'INRA

Créées grâce au travail assidu d'Eric Germain et de son équipe, ces variétés ont représenté une amélioration par rapport aux variétés traditionnelles. La connaissance des marqueurs moléculaires va permettre de nouvelles avancées.

Visuellement, cette collection présente une large diversité, dite phénotypique. Au sein de l'espèce *Juglans regia*, on remarque des arbres au feuillage pourpre ou d'autres au feuillage lacinié. Certains donnent des noix au cerneau rouge, le tégument recouvrant les cotylédons étant également pourpre. Les noix les plus grosses, comme celles de la variété 'Bijou', peuvent atteindre 10 centimètres de diamètre, et d'autres ne dépassent pas la taille d'une noisette. Les arbres les plus précoces débourrent la première semaine de mars alors que les plus tardifs attendent la dernière semaine d'avril. Cette grande variabilité phénotypique est importante pour le sélectionneur, puisqu'il possède alors un large éventail de géniteurs possibles pour créer une nouvelle variété en fonction des besoins.

Les autres collections mondiales possèdent *a priori* moins de diversité rassemblée sur un même site. En Iran et en Turquie, pays où l'amélioration de la noix s'est plutôt pratiquée par sélection de matériel au sein de populations naturelles, les collections de ressources génétiques semblent restreintes à quelques zones géographiques précises. C'est le même constat pour la Chine.

En Californie, pays leader en création variétale pour le noyer, le matériel disponible pour de potentiels géniteurs est également plus restreint et semble constitué en majorité de variétés cultivées. Par comparaison, la collection française de l'INRA s'avère donc une des plus importantes au monde.

Comment aider le noyer à résister au changement climatique demain

Aujourd'hui, la filière nucicole française est confrontée à une concurrence mondiale accrue. Elle doit aussi faire face aux effets déjà visibles du changement climatique, avec une augmentation des températures moyennes et l'apparition de gelées tardives printanières, qui peuvent engendrer des pertes de production comme en 2017 où la production du Périgord a plafonné à 12 000 tonnes au lieu des 20 000 tonnes habituelles. Par ailleurs, les pathogènes futurs et déjà existants doivent être pris en compte dans un éventuel programme de création variétale. On peut citer par exemple les dégâts causés par la bactériose due à *Xanthomonas arboricola* pv. *juglandis*, et ceux provoqués par deux formes d'anthracnose, liés aux champignons *Gnomonia leptostyla* et *Colletotrichum acutatum*. Il est donc dans l'intérêt de la filière de pouvoir disposer de variétés à cycle court, capables d'éviter les gelées printanières tout en proposant les fruits sur les étals relativement tôt dans la saison pour résister à la concurrence américaine. Si ces critères sont primordiaux, il ne faut pas oublier la base requise pour une variété française : une forte productivité, une grande vigueur de l'arbre et une qualité organoleptique toujours au rendez-vous. Le levier génétique semble donc être une stratégie de choix pour aider l'améliorateur à parvenir à ces variétés, même si les pratiques culturales devront également être améliorées en parallèle.

La noix 'Bijou', dite aussi à cavernes, est particulièrement grosse. Autrefois on sertissait ses bords en argent et on l'équipait d'une charnière pour en faire un petit coffre à bijoux... ou pour abriter de minuscules couteaux, une spécialité de Nontron, dans le nord de la Dordogne (Doc. La Belgique horticole, 1853).

Le pinceau du grand Audubon a immortalisé le noyer noir d'Amérique, *Juglans nigra* (Doc. DR).

Au sein de ce levier génétique, la connaissance des marqueurs moléculaires peut faciliter la tâche du sélectionneur. Elle permet de dépasser la simple observation des individus pour s'intéresser à leur capital génétique. Les marqueurs moléculaires peuvent être utilisés au cours des processus de sélection à n'importe quel stade de développement de l'individu et sur n'importe quel organe, puisque toute cellule possède la même information génétique. Plus besoin d'attendre six à sept ans pour juger des caractères de l'arbre et de ses fruits. Prenons un exemple : un sélectionneur souhaite obtenir une variété à gros fruit et à débourrement tardif. Pour cela, il décide de croiser un parent à gros fruits mais à débourrement intermédiaire avec un autre parent à débourrement tardif mais donnant des fruits moyens. Dans la première génération, le sélectionneur cherchera les individus qui possèdent les deux marqueurs pour les deux caractères : « gros fruit » et « débourrement tardif », sans devoir attendre que les individus produisent des fruits au bout de sept ans. Il fera germer les noix issues du croisement et attendra quelques semaines jusqu'à l'apparition des premières feuilles. Une extraction d'ADN de ces feuilles sera effectuée et après recherche desdits marqueurs, le tri pourra être fait. On voit bien que cette technologie permet de gagner de précieuses années dans les cycles de sélection.

Néanmoins, cette technique de sélection assistée par marqueurs n'est pas sans difficulté. Il faut d'abord être en possession de ces marqueurs, longs à développer en laboratoire. Chez les espèces très travaillées comme la tomate ou le blé, nous avons la connaissance du déterminisme génétique des principaux caractères d'intérêt agronomique. En revanche, pour les espèces plus orphelines en termes de recherches scientifiques, comme le noyer, ces connaissances peuvent manquer. Ainsi, un premier travail à réaliser est d'obtenir ces marqueurs. C'est là où la collection de ressources génétiques du noyer de l'INRA de Nouvelle-Aquitaine-Bordeaux devient un atout majeur car il s'agit d'un réservoir de caractères et de gènes précieux pour mener des études de génétique d'association. Dans ce cadre, le projet INNOV'noyer a pu voir le jour, à l'initiative du CTIFL de Lanxade (Centre Technique Interprofessionnel des Fruits et Légumes), en partenariat avec l'INRA de Nouvelle-Aquitaine-Bordeaux et l'Université de Californie de Davis, et financé

UNE STATION DE RECHERCHE DÉDIÉE AU NOYER

La culture du noyer dans le du Sud-Ouest est ancienne et a une valeur patrimoniale importante. À la fin des années 1980, pour réagir notamment face à une forte concurrence californienne, les producteurs et responsables de la région décident de créer une station régionale professionnelle. Elle est située à Creysse, dans le Quercy, au carrefour de trois départements à forte tradition nucicole : le Lot, la Corrèze et la Dordogne. Le verger comporte plus de 100 variétés implantées sur 30 hectares.

Actuellement le programme est axé sur trois thématiques principales : en priorité un travail de sélection qui devra permettre de diversifier la gamme variétale existante en intégrant des variétés performantes et tolérantes à la bactériose. Au passage, il s'agit également d'améliorer la production des variétés à fort potentiel par l'usage de porte-greffe vigoureux dont on attend aussi une limitation des intrants, eau et engrais. Ensuite, il convient de travailler sur la conduite des vergers modernes, notamment la taille et la densité de plantation, afin de diminuer leurs besoins en main d'œuvre et d'accélérer leur entrée en production. Enfin la mise au point de systèmes de production respectueux de l'environnement et de la santé des utilisateurs est une préoccupation forte pour la filière. Une protection efficace contre la bactériose, principale maladie du noyer, constitue pour la recherche un enjeu impératif pour le maintien de la rentabilité des vergers.

Pour en savoir plus : www.noixsudouest.fr

Chez le noyer, les fleurs mâles et femelles sont bien distinctes (ph. Eloïse Tranchand, ingénieure à la station de Creysse).

UN GRAND CONNAISSEUR DES NOYERS

Né en Algérie en 1938, Éric Germain a hésité entre médecine ou agronomie, mais étant issu du milieu agricole par son père qui possédait une production de vigne et d'agrumes dans la région de Bône, il intègre en 1958 l'école nationale d'agriculture de Grignon, dans le but de reprendre l'exploitation. En 1963, après avoir terminé son cursus suivi de son service militaire en Algérie, il revient en France où sa famille possédait également une propriété dans le Sud-Ouest. Il entre cette même année à l'INRA de manière fortuite à la suite d'une discussion avec le directeur de l'époque, lors d'une visite où il accompagne son père à l'ancienne station d'arboriculture de la Grande-Ferrade. Entre 1967 à 1970, il est adjoint au directeur de cette station et est rapidement chargé du programme d'amélioration génétique du noyer et du noisetier, avec une équipe solide de trois techniciens sous sa houlette.

En 1976, il part en Californie durant trois mois, à l'Université de Davis, pour découvrir ce qui se faisait de mieux en termes de création variétale sur le noyer. Il initie ses premières hybridations en 1977 dans le but d'obtenir une variété à fructification latérale et à floraison tardive, avec une fructification précoce, un haut potentiel de production, une faible sensibilité à la bactériose et une bonne qualité gustative.

Vingt ans après, ses premières variétés sont inscrites au catalogue officiel des variétés fruitières : 'Fernor', 'Fernette', 'Ferjean'... En parallèle, il entreprend un grand travail de prospection pour rassembler du matériel végétal venu du monde entier. En fin de carrière, il est nommé directeur adjoint de l'Unité de recherche sur les espèces fruitières et la vigne. Éric Germain est décédé en décembre 2011.

Un autre bouquet d'oiseaux d'Audubon sur des branches de *Juglans cinerea* (doc. DR).

par la Région Nouvelle-Aquitaine et le dispositif Cifre de l'ANRT (Agence Nationale pour la Recherche et la Technologie). Ce projet finance une thèse visant à connaître ces déterminismes génétiques. Tous les individus de la collection seront phénotypés sur les principaux caractères d'intérêt agronomique (port et vigueur de l'arbre, caractéristiques du fruit et du cerneau, sensibilité aux pathogènes...) et génotypés à l'aide d'une puce à ADN, développée en Californie à l'Université de Davis et contenant plusieurs milliers de marqueurs potentiels, répartis sur tout le génome. Des analyses statistiques pourront déterminer si, par exemple, une région du génome est en corrélation avec un caractère donné, tel que la taille du fruit ou la date de débourrement. C'est ce qu'on appelle une analyse de génétique d'association non ciblée ou *Genome-Wide Association Study*. Les marqueurs obtenus, en addition avec la bonne connaissance du phénotype des individus de la collection, seront les outils de demain pour initier un nouveau programme d'innovation variétale plus efficient. Il s'agit donc d'une approche multidisciplinaire faisant appel aux compétences des différents partenaires : le CTIFL pour son expertise en pathologie et biologie moléculaire, l'INRA pour sa collection et ses connaissances en phénotypage du noyer, en génétique des populations et en génomique, et l'Université de Davis qui est à l'origine du séquençage du génome du noyer et qui a mis au point cette puce de marqueurs. Grâce à cette alliance, le noyer dévoilera quelques-uns de ses secrets les mieux cachés...

Le gaulage des noyers n'est plus de mise dans les vergers modernes. Place aux machines... (ph. Syndicat professionnel de la noix et du cerneau de noix du Périgord).

Annexe 2 : Article « The INRA's Walnut Germplasm Collection, a Reservoir of Diversity Essential for the Future of French Walnut Industry » dans la revue « Nutfruit Magazine ».

The INRA's Walnut Germplasm Collection, a Reservoir of Diversity Essential for the Future of French Walnut Industry

ANTHONY BERNARD

PhD student at the University of Bordeaux working on the study of walnut genetic diversity to provide knowledge and tools for the future French improvement program. He co-authored the article Bernard, A., Lheureux, F., & Dirlewanger, E. (2018). Walnut: past and future of genetic improvement. *Tree Genetics & Genomes*, 14(1), 1.

Tree of beliefs and legends, the walnut deserves this short overview of the situation of the walnut industry in France, and the past and future contribution of the INRA's walnut germplasm collection to the new French improvement program.

Since ancient times, walnut has been the subject of many legends. Greek mythology tells us that Dionysus fell in love with King Dion's youngest daughter, Carya, whose two sisters were turned into rocks because of their jealousy. Later, Carya dies of sadness and Dionysus immortalizes her in a fertile walnut tree. She later gave her name to Artemis, who is the godess of walnut and fruit bearing trees, under the name of Artemis Karyatis. A temple dedicated to her shows walnut columns, carved in the shapes of the goddess, erected to support the capital. These are the famous caryatids, which are also found in stone at the entrance to the Erechtheion of the Acropolis of Athens and in many classical buildings such as the Louvre.

The cultivated walnut, *Juglans regia* L., is believed to be native to Central Asia, from Kashmir, in the foothills of the Himalayas, to present-day Kyrgyzstan. It first spread to Uzbekistan, northern Iran, the Caucasus region and eastern Turkey, and then to northern India, Nepal and western China. Its introduction in Europe dates from the last ice age, about 50,000 years ago. Archaeological evidence found in Périgord in France suggests that Cro-Magnon man could have consumed ancestral forms of the fruit of *Juglans regia* 17,000 years ago. In Ancient Greece, the walnut became more abundant thanks to Alexander the Great and its dispersion then increased in Spain, Portugal and North Africa. In the 16th century, the conquest of the New World allowed its arrival in South America, notably Chile, before its introduction in California in the 19th century.

World walnut production has grown steadily over the last decades and exceeded 866,000 metric tons in 2017/18 (kernel basis), according to INC data. France production was estimated at 17,500 metric tons in 2017/18, placing it sixth

Picture of a purple walnut leaf.
© Anthony Bernard

Picture of a classical walnut leaf.
© Anthony Bernard

Picture of a laciniate walnut leaf.
© Anthony Bernard

in the world and second in Europe, behind Ukraine. Although the United States, Chile and Eastern Europe dominate the world exports, the French walnut nevertheless resists global competition thanks to its recognized organoleptic qualities.

There are two appellations of controlled origin corresponding to the two main production regions, the AOC 'Noix du Périgord' and 'Noix de Grenoble'. France exports 80% of its production, mainly to Europe and mainly destined for the table nut market. The old variety 'Franquette' provides about 70% of the production and the areas allocated to walnut increased by 19% between 2000 and 2010, reaching more than 20,000 hectares. This development makes it the second largest fruit crop after apples, with more than 1,800,000 trees.

Éric Germain, a researcher from Institut National de la Recherche Agronomique (INRA)- Bordeaux, led the first walnut improvement program in France. From 1977 to 1995, he made crosses, mainly between old French varieties such as 'Franquette', 'Grandjean' and 'Marbot', appreciated for their late budbreak date and quality, and Californian varieties such as 'Pedro' and 'Chandler', known for their ability to produce fruit all along the shoot. There are indeed two main types of fruiting in walnut: in terminal position, at the end of the current season's shoot, as is the case in old French varieties, and in lateral position, all along the one-year shoot. This is the case of the American variety 'Payne', present in the pedigree of most American varieties. Two cultivars with lateral fruiting were registered in 1995: 'Fernette' and 'Fernor', resulting from a cross between 'Franquette' and 'Lara®', the latter having 'Payne' in its pedigree. A third cultivar with lateral fruiting and rather small kernels, 'Ferjean', resulting from the crossing between 'Grosvert' and 'Lara®', was registered in 1999. Finally, the last four cultivars were released in 2010: 'Feradam', 'Ferbel', 'Ferouette' and 'Fertignac'. Between 1996 and 2007, Éric Germain used a germplasm collection resulting from important prospecting work that he did in many countries between 1988 and 2000. The most interesting varieties have been crossed with plant material from different regions of the world, such as the Mediterranean, the Middle East, Central Asia and Japan. This diversification aimed to reduce the inbreeding problems previously observed in crossbreeding between French and American materials. Following INRA's strategic decisions, the improvement program was stopped in 2007 although the experimental stations had identified some promising hybrids.

The INRA's walnut collection is however still maintained by INRA - Bordeaux, more precisely at the *Prunus-Juglans* Genetic Resources Center. Aware of the importance of safeguarding biodiversity, Éric Germain knew that creating a reservoir of genetic diversity is an essential basis for initiating an improvement program. Thus, this rich collection represents a large part of the species of the genus *Juglans*, twelve different species and more than 400 accessions. For the only cultivated walnut, *Juglans regia*, the collection has nearly two hundred accessions coming from all over the world.

Visually, this collection presents a wide diversity of observable characteristics, or 'phenotypes'. Within the species *Juglans regia*, we notice trees with purple foliage and others with laciniated foliage. Some produce nuts with red kernels. The largest nuts can reach 10 centimetres in diameter (e.g., nuts of the 'Bijou' variety), and others do not exceed the size of a hazelnut. The earliest trees emerge in the first week of March while the later trees wait until the last week of April. This great phenotypic variability is important for the breeder, who has a wide range of possible genitors to create a new variety as needed. The INRA's collection is undeniably one of the most important in the world.

Today, the French walnut industry is facing increased global competition. It also has to cope with the already visible effects of climate change, with an increase in average temperatures and the appearance of late spring frosts, which can cause production losses. In 2017, production in the Périgord peaked at 12,000 tonnes instead of the usual 20,000 tonnes because of late spring frosts. In addition, future and existing pathogens must be taken into account in a possible breeding program. Examples include damage caused by bacterial infections caused by *Xanthomonas arboricola* pv. *juglandis*, and damage caused by two forms of anthracnose, the fungi *Gnomonia leptostyla* and *Colletotrichum acutatum*. It is therefore in the interest of the sector to have varieties with a short cycle, capable of avoiding spring frosts while offering the fruit on the market stalls relatively early in the season to resist American competition. If these criteria are essential, a French variety should also have high productivity, great vigour of the tree and an organoleptic quality always present. Nowadays, focusing on the knowledge of genetic diversity seems to be a promising strategy to help the breeder to achieve these varieties, even if orchard management improvements have to be made in parallel. INRA and the Ctifl (Technical Center for Fruits and Vegetables) are now exploring the genetic diversity of the INRA's germplasm collection to conduct association genetics studies and to choose new genitors suitable to face new challenges of walnut industry. The new breeding program will be led by the Ctifl.

Annexe 3 : Synthèse des travaux réalisés en pathologie.

1. Introduction

Dans ce contexte de changement climatique, la pression de pathogènes nouveaux ou déjà existants tend à devenir plus importante. En arboriculture fruitière, la qualité du fruit et la phénologie sont évidemment des critères prioritaires de recherche mais un verger est implanté pour de nombreuses années. Il est donc indispensable d'identifier des sources de tolérance à ces pathogènes dans le but de développer des variétés plus résilientes. Chez le noyer, deux pathogènes (un champignon et une bactérie) sont particulièrement sources de pertes significatives chez les producteurs·trices.

2. Synthèse bibliographique concernant les deux principaux pathogènes du noyer

Le complexe d'espèces *Colletotrichum acutatum*, ascomycète responsable d'une forme d'anthracnose

Dans le cadre de la validation de sa Licence Professionnelle Génomes et Biotechnologies pour l'Amélioration des Plantes, à l'Université de Toulouse III Paul Sabatier, Joris Molina a pu travailler de mars à août 2017 sur le sujet suivant : « étude de la sensibilité et/ou de la tolérance de différentes accessions de noyer à *Colletotrichum acutatum* ». L'encadrement de Joris s'est inscrit dans le déroulement de la présente thèse. Joris a d'abord réalisé une synthèse bibliographique sur ce complexe d'espèces de champignon et son interaction avec le noyer, dont voici un résumé.

Quelques mots sur Colletotrichum sp., un genre répandu

Le genre de champignon *Colletotrichum* sp. appartient à l'embranchement des Ascomycètes et abrite un grand nombre d'espèces phytopathogènes responsables d'une maladie : l'anthracnose. Celle-ci a pour conséquence des pertes économiques à travers le monde sur une grande diversité de végétaux (Dean et al. 2012). Parmi les espèces phytopathogènes, *C. acutatum* J.H. Simmonds (forme téléomorphe : *Glomerella acutata*) est un complexe d'espèces très répandu (Damm et al. 2012). Il est constitué de plusieurs espèces distinctes génétiquement mais qui sont difficilement différenciables morphologiquement (mycélium dense, blanc à gris, avec acervules orange à rose à la surface). Auparavant, *C. acutatum* fut décrit comme une espèce à part entière mais en 2012, l'analyse phylogénétique multi-locus de plus de 300 isolats a permis l'identification de 29 espèces au sein du complexe *C. acutatum sensu lato* (Damm et al. 2012). Dans cette étude, les espèces les plus représentées sont : *C. acutatum sensu stricto*, *C. cuscutae*, *C. fioriniae*, *C. godetiae*, *C. lupini* et *C.*

nymphaeae. L'identification moléculaire grâce aux régions Internal Transcribed Spacer (ITS) de l'ADN ribosomal amplifiées par PCR (Liu et al. 2012) permet maintenant de différencier *C. acutatum* et *C. gloeosporioides*, longtemps confondus en raison de morphologies très proches (Peres et al. 2005).

C. acutatum possède une large gamme d'hôtes comme, entre autres, la pomme (Jones 1996), la fraise (Xie et al. 2010), l'amande (Förster and Adaskaveg 1999) et la noix (Giraud 2012). De ce fait, les tissus infectés et les symptômes induits varient d'un hôte à l'autre, ce qui le rend peu spécifique et d'autant plus difficile à étudier. Une étude a justement montré, après inoculations croisées, que certains isolats de *C. acutatum* sont majoritairement non spécifiques à un hôte (Freeman et al. 1998). Néanmoins, les symptômes les plus communément retrouvés sont des nécroses mais on peut aussi trouver des piqûres sur divers appareils végétatifs tels que les fleurs, les racines, les feuilles et les fruits. Il est connu aujourd'hui que *C. acutatum* n'engendre pas toujours des nécroses. En effet, il peut alterner les phases biotrophes et nécrotrophes. Sur l'amande par exemple, il infecte les jeunes feuilles puis instaure une longue période de biotrophie avant de nécroser celles-ci lors de son grandissement cellulaire et de son expansion sur le végétal (Diéguez-Uribeondo et al. 2005). S'il est peu spécifique à un hôte, ce pathogène est hautement spécialisé quant aux tissus qu'il attaque sur chacun de ces hôtes. En effet, le tissu juvénile peut être attaqué, comme c'est le cas pour les feuilles d'amandier. Dans d'autres cas, il touche la fleur comme chez l'oranger ou le fruit comme chez la pomme où il n'induit pas de symptôme sur feuille. Enfin, ce champignon peut aussi produire des symptômes post récolte, notamment chez la myrtille avec l'apparition de pourriture lors du stockage.

Le cycle général de reproduction de C. acutatum

a) *processus d'infection* – cette espèce peut coloniser la surface des tissus végétaux sous forme d'épiphytes ou d'endophytes dans certaines circonstances, tant chez les hôtes que chez les non-hôtes, sans produire de symptômes (Freeman et al. 2000; Horowitz et al. 2002; Jones et al. 1996). Dans tous les cas, des acervules avec des conidies abondantes sont présents sur les tissus de leur hôte. Une fois que les conidies sont fixées sur le tissu hôte, celles-ci germent et créent dans la plupart des cas des appressoria (Curry et al. 2002). Ces appressoria engendrent un filet infectieux (ou haustorium) qui peut être extra-cellulaire et le champignon va ensuite croitre et coloniser les tissus. Une nécrose va apparaitre et le champignon va produire des acervules grâce à l'énergie qu'il aura puisée au sein du végétal (Wharton and Diéguez-Uribeondo 2004). A l'heure actuelle, il n'y a pas d'information sur le fait que le

champignon soit capable de percer la cuticule du végétal grâce à la pression exercée par l'appressorium ou si celui-ci est opportuniste et rentre par des stomates ou autres points d'ouverture (par exemple, une blessure antagoniste). Malgré tout, après pénétration dans l'hôte, *C. acutatum* peut développer deux stratégies d'infection : une stratégie intracellulaire hémi-biotrophe et une stratégie subcuticulaire nécrotrophe.

b) *dispersion* – elle se fait grâce au vent et la pluie. Les acervules contenant les conidies peuvent être éclatés par la pluie et les conidies peuvent ainsi se disséminer par projection. En conditions sèches, un fort vent peut aussi conduire les conidies à se déplacer et donc infecter des plantes à proximité.

c) *survie* – le champignon présente plusieurs « formes » de survie. Les fruits momifiés présentent la principale source d'infection pour les prochaines générations de fruits ; les conidies sont présentes dans ces fruits momifiés et se dispersent ensuite grâce au vent et à la pluie comme expliqué précédemment (Adaskaveg and Förster 2000). Les conidies peuvent aussi survivre pendant de longues périodes dans les débris ou le sol dans des conditions sèches. Cependant, le champignon est incapable de rivaliser ou de se multiplier sur des débris de feuilles ou dans le sol dans des conditions humides où d'autres organismes sont actifs (Eastburn and Gubler 1990).

L'interaction C. acutatum – noyer

Aujourd'hui, il existe encore peu d'information quant à l'interaction entre *C. acutatum* et le noyer. Néanmoins, le genre *Colletotrichum* sp. est apparu sur noyer dans les années 1998-2000 comme le prouvent des études qui ont révélé sa présence, en faible quantité, sur des noix présentant des symptômes de Brown Apical Necrosis (BAN) et provenant de France et d'Italie (Belisario et al. 2002). Il a aussi été détecté plus tard dans des bourgeons de noyer en dormance ainsi que dans des nécroses sur noix en compagnie d'autres pathogènes comme *Xanthomonas arboricola* pv. *juglandis* provoquant la bactériose. De ce fait, sa présence a été rapportée à une attaque de type saprophyte et non à une attaque primaire (Giraud 2012). Par ailleurs, en 2008, des taches marron avec des sporulations orange ont été observées sur noix en France et des souches de *Colletotrichum* sp. ont été découvertes. La virulence sur noix a pu ensuite être prouvée par vérification du postulat de Koch (Giraud 2012). Mais c'est véritablement en 2011 que le complexe *C. acutatum* a provoqué des pertes importantes dans le sud-est de la France. Dans la continuité, des analyses PCR ont été réalisées entre 2013 à 2015 au CTIFL de Lanxade et ont révélé que ce complexe d'espèces *C. acutatum* était le complexe le plus présent sur noyer en France avec 99% des souches analysées ; 1% restant attribué à *C. gloeosporioides*. De plus, des analyses effectuées en 2015 par le Laboratoire

Universitaire de Biodiversité et Ecologie Microbienne (LUBEM) ont permis d'identifier en France en particulier deux espèces du complexe *C. acutatum* : *C. fioriniae* et *C. godetiae* (Donnée interne de la SENuRA). La présence de ce champignon en France remonte donc aux années 1990 avec un développement et un accroissement de virulence de nos jours jusqu'à engendrer des pertes de récolte. Cette augmentation de virulence peut être due au changement de lutte chimique effectuée au cours de ces dernières années ou bien au changement climatique qui a peut-être favorisé sur la noix le développement de ce pathogène plutôt qu'un autre (Ramain 2012).

C. acutatum possède une forme sexuée mais il semblerait qu'elle ne joue qu'un rôle mineur et elle n'a jamais été observée sur noyer. De ce fait, sur noyer, la reproduction de ce champignon est essentiellement asexuée et se manifeste par la formation des acervules orange à la surface des tissus nécrosés. Un cycle épidémiologique de *C. acutatum* sur noyer a néanmoins été proposé. Au printemps, un premier pic de projection de conidies semble être observé d'avril à mai au moment du débourrement et de la floraison. Des infections quiescentes s'établissent sur les fruits immatures. Cela forme l'inoculation primaire dont les principales sources sont des fruits contaminés au sol et dans les arbres ainsi que les bourgeons (Verhaeghe and Giraud 2015). Le champignon engendre alors des nécroses sur fruits et produit des fructifications, sources de contaminations secondaires. En effet, un second pic de projection de conidies semble enregistré en été. Au final, la noix chute prématurément. Il est important de noter toutefois que ces symptômes peuvent être confondus avec ceux d'autres maladies telles que la bactériose ou ceux de l'anthracnose liée à *Gnomonia leptostyla* mais celle-ci entraîne des taches avec la partie centrale préférentiellement blanche à gris clair (Figure 3).

Enfin, les recherches concernant différentes substances actives pouvant être efficaces contre *C. acutatum* sont toujours en cours dans les stations régionales. Des bactériophages sont également à l'essai, en partenariat avec la filière chilienne, mais aujourd'hui, aucun produit phytosanitaire n'est préconisé. Aucune source bibliographique ne fait mention de pertes de récolte dues à ce pathogène autre part qu'en France. Le cas de la France est donc particulier, voire émergent au regard des autres pays producteurs.

Xanthomonas arboricola pv. *juglandis*, bactérie responsable de la bactériose

Dans le cadre de la validation du niveau L3 et de son apprentissage en entreprise, au CTIFL, à l'école d'ingénieurs Agrocampus Ouest option Horticulture, Morgane Vincent a pu

travailler de mars à août 2018 sur le sujet suivant : « *Colletotrichum acutatum* et *Xanthomonas arboricola* pv. *juglandis* : mise au point d'une méthode d'inoculation d'organes détachés de noyer en conditions contrôlées ». L'encadrement de Morgane s'est inscrit dans le déroulement de la présente thèse. Morgane a d'abord réalisé une synthèse bibliographique sur cette bactérie et son interaction avec le noyer, dont voici un résumé.

La bactériose et ses symptômes

La bactériose est la principale maladie affectant la production de noix dans le monde, ainsi qu'en France (Lindow et al. 2014). *X. arboricola* pv. *juglandis* est une bactérie gram négatif. Le pathovar *juglandis* en fait un pathogène spécifique de ce genre. Il est responsable de la bactériose chez le noyer cultivé ainsi que chez ses espèces apparentées. Cette bactérie peut infecter tous les tissus de l'hôte dont les fruits et la rend responsable de pertes considérables pouvant réduire à 80% la production si elle n'est pas contrôlée (Germain et al. 1999). Les infections sur fruits mènent à une chute prématurée des fruits affectant le rendement. Si l'infection est plus tardive, la nécrose n'atteint pas le cerneau mais colore la coque et réduit la qualité des noix vendues en coque.

Chaque organe de l'arbre peut exprimer des symptômes. Les feuilles et les fruits présentent des piqûres de nécroses noires huileuses entourées d'un halo jaune caractéristique. La nécrose peut se situer sur la partie apicale du fruit, symptôme typique de BAN causé par *X. a.* pv. *juglandis* (Moragrega and Özaktan 2010). Quant aux fleurs mâles et femelles, elles deviennent noires et tombent de l'arbre. Un chancre vertical avec suintements d'exsudats, le Vertical Oozing Canker (VOC), peut également se former au niveau du tronc ou des pousses (Hajri et al. 2010).

Le cycle d'infection du noyer

Ce cycle débute au moment du débourrement. En effet, la bactérie se multiplie quand l'humidité de l'air et l'humectation du végétal sont élevées et que les températures sont comprises entre 16°C et 29°C. Sous l'action de la pluie, les bactéries se propagent. Elles pénètrent dans l'hôte par des ouvertures naturelles comme les stomates ou les tissus blessés. Elles atteignent également les fleurs puis les fruits et finissent par provoquer une infection. Si les conditions favorables sont réunies, un inoculum secondaire peut alors causer des dommages supplémentaires à partir des feuilles et des rameaux malades (Lamichhane 2014).

Si le foyer de contamination est important, une quantité élevée d'inoculum se retrouvera dans les bourgeons et la contamination l'année suivante sera d'autant plus

importante. Le pollen des chatons est une cause possible de contamination mais l'inoculum présent dans les écailles des bourgeons dormants des fleurs femelles et des feuilles est la principale source des infections. La propagation de la maladie est essentiellement liée au nombre d'évènements pluvieux suivant le débourrement chaque année (Lindow et al. 2014). Aucun moyen curatif n'existe contre cette maladie. La gestion actuelle consiste en l'utilisation de produits cupriques à effet bactériostatique. L'application commence dès le débourrement puis s'effectue à intervalles réguliers jusqu'à la récolte. Cependant, l'efficacité de cette méthode est compromise par l'émergence de souches résistantes au cuivre (Moragrega and Özaktan 2010). La méthode la plus efficace pour gérer la quantité de l'inoculum reste la prophylaxie.

Photographies des principales maladies du noyer.
A gauche, les symptômes caractéristiques sur noix de la bactériose provoquée par *X. a.* pv. *juglandis*. A droite, ceux de l'anthracnose provoquée par *C. acutatum*. Photographies de la Station Expérimentale de la Noix de Creysse.

3. Méthodes de caractérisation de la sensibilité des accessions à ces deux principaux pathogènes

En raison du nombre important de caractères cités précédemment et étudiés durant la thèse, la caractérisation de la sensibilité aux deux principaux pathogènes du noyer a été confiée chaque année à des stagiaires. L'encadrement de ces stagiaires s'est inscrit dans le déroulement de la présente thèse. En effet, aucun protocole de routine n'a encore été développé à ce jour pour ces caractérisations. Le but de ces différents stages a été donc de mettre au point des méthodes de culture de ces pathogènes et de leur inoculation sur différents organes détachés du noyer ; organes détachés puisque des tests en conditions de verger ne sont pas réalisables. Les deux principaux pathogènes du noyer sont d'une part, le complexe

d'espèces *Colletotrichum acutatum*, ascomycètes responsables d'une forme d'anthracnose et d'autre part, la bactérie *Xanthomonas arboricola* pv. *juglandis*, responsable de la bactériose. Aucun moyen de lutte efficace n'existe à ce jour pour ces deux maladies. Ainsi, plusieurs travaux ont été réalisés sur ces pathogènes durant les saisons de printemps/été en 2017, 2018 et 2019. Ce qui suit est une synthèse du « Matériels et Méthodes » de ces différents stages.

Etude préliminaire de la sensibilité de l'ensemble des accessions de la collection à *Colletotrichum acutatum*

Dans le cadre de la validation de sa Licence Professionnelle Génomes et Biotechnologies pour l'Amélioration des Plantes, à l'Université de Toulouse III Paul Sabatier, Joris Molina a pu travailler de mars à août 2017 sur le sujet suivant « Etude de la sensibilité et/ou de la tolérance de différentes accessions de noyer à *Colletotrichum acutatum* ». Joris a mis en œuvre un test pour apprécier la sensibilité de 253 accessions de la collection à ce pathogène selon les protocoles qui suivent.

Repiquage de l'isolat de C. acutatum

L'isolat du complexe de *C. acutatum* du CTIFL, préalablement vérifié par PCR (Liu et al. 2012), provient d'une noix nécrosée de l'UEA de Toulenne. L'isolat est repiqué au scalpel, en conditions stériles, sur du milieu Potato Dextrose Agar (PDA) : 4 g/L d'extrait de pomme de terre, 20 g/L de dextrose et 15 g/L d'agar.

Prélèvement et désinfection des folioles de noyer

Un total de 5 folioles (4 inoculées et 1 témoin) est prélevé pour chacune des 253 accessions de la collection, en prenant soin de prélever aléatoirement tout autour de l'arbre afin d'avoir un échantillonnage plus représentatif de l'arbre. Le prélèvement est effectué environ 45 jours après la date de débourrement (stade Cf + 45 j), stade auquel les folioles ne sont pas trop agées, mais ont atteint leur dimension définitive. Les folioles sont rincées dans un premier temps à l'eau du robinet afin d'éliminer toutes les impuretés grossières et subissent ensuite 4 bains de 50 secondes chacun en conditions stériles (1 bain d'éthanol à 70% et 3 bains d'eau distillée autoclavée). Les folioles, toujours en conditions stériles, sont déposées dans des boites de Pétri de 15 cm de diamètre sur 4 couches de papier buvard préalablement autoclavé et ce papier est ensuite imbibé d'eau distillée autoclavée afin de maintenir une humidité dans les boites. La face adaxiale est contre le papier pour laisser libres les stomates.

Préparation d'une solution de spores et inoculation des folioles

Il est nécessaire de préparer un étalement de spores 3 jours avant l'inoculation. Des gouttes de spores prélevées dans la boite de repiquage sont mélangées à 1 mL d'eau distillée autoclavée. 200 µL de cette solution sont étalés sur du milieu PDA et mis en étuve à 32°C durant 3 jours. A la fin de cette durée, les spores ont germé et sont prêtes à être utilisées en solution. Une concentration de spores de 1.10^5/mL est souhaitée pour observer des symptômes et est obtenue grâce aux recommandations classiques de comptage à la cellule de Malassez.

La foliole témoin est piquée à l'aide d'une aiguille et 10 µL d'eau stérile sont déposés sur ces points de piqûre. Deux piqûres sont réalisées sur le limbe et deux autres sur la nervure centrale, à chaque extrémité de la foliole. Pour les folioles test, 10 µl de solution de spores sont déposés sur les points de piqûres. Les boites de Pétri sont placées en chambre de culture dans des conditions adéquates pour le maintien de la foliole et le développement des symptômes : 23°C jour et 18°C nuit, avec un photopériodisme de 16h jour et 8h nuit et avec une hygrométrie de 80%.

Prise d'images et analyses

La prise d'images des folioles s'effectue 5 jours après l'inoculation avec un appareil photo numérique en positionnant les folioles sur un papier bleu contrastant. Le programme « APS Assess 2.0 », développé par Lakhdar Lamari en 2008, permet ensuite d'analyser les images et de quantifier les symptômes, donc les nécroses. A l'aide d'une surface de 30 x 30 mm servant de témoin sur le papier bleu, l'aire de la nécrose délimitée par le programme est donnée en mm².

Etude précisée de la sensibilité d'un groupe d'accessions à *Colletotrichum acutatum* et à *Xanthomonas arboricola* pv. *juglandis*

Dans le cadre de la validation du niveau L3 et de son apprentissage en entreprise, au CTIFL, à l'école d'ingénieurs Agrocampus Ouest option Horticulture, Morgane Vincent a pu travailler de mars à août 2018 sur le sujet suivant : « *Colletotrichum acutatum* et *Xanthomonas arboricola* pv. *juglandis* : mise au point d'une méthode d'inoculation d'organes détachés de noyer en conditions contrôlées ». Sur la base des résultats de Joris, elle a pu tester différentes méthodes d'inoculation et différents organes détachés. Morgane a donc testé la sensibilité d'un groupe d'accessions aux deux pathogènes, selon les protocoles qui suivent. Le fait de tester différentes méthodes d'inoculation permet d'observer leurs avantages et inconvénients. Aussi, si les résultats concordent entre les organes, la fiabilité du protocole

sera augmentée. Par ailleurs, il est très probable d'observer des différences de comportement du groupe d'accessions en fonction du pathosystème étudié.

Choix du groupe d'accessions de la collection

Sur la base des résultats de Joris et de la faible littérature existante, 6 accessions ont été sélectionnées pour cette étude précisée, afin de tester différentes méthodes d'inoculation des deux pathogènes. Ce nombre plus restreint d'accessions a aussi permis d'augmenter le nombre de répétitions de folioles et de tester des inoculations sur fruits détachés. Au final, 6 accessions ont été retenues : 3 accessions de *J. regia* et 3 autres appartenant aux espèces apparentées :

- 'Serr' (*J. regia*) : considérée faiblement sensible à *X. a.* pv. *juglandis* (Tsiantos et al. 2008) – donnée manquante pour *C. acutatum*.
- 'Franquette' (*J. regia*) : considéree moyennement sensible à *X. a.* pv. *juglandis* (Öztakan et al. 2008) – donnée manquante pour *C. acutatum*.
- 'UK 6-2' (*J. regia*) : considéree comme fortement sensible aux deux pathogènes par observation de symptômes en vergers.
- 'SB 24-2' (*J. sieboldiana)* : donnée manquante pour *X. a.* pv. *juglandis* – considérée comme faiblement sensible à *C. acutatum* (résultats de Joris Molina).
- 'CN 1-9' (*J. cinerea)* : donnée manquante pour *X. a.* pv. *juglandis* – considérée comme faiblement sensible à *C. acutatum* (résultats de Joris Molina).
- Une accession de *J. nigra* : espèce considérée comme faiblement sensible à *X. a.* pv. *juglandis* (Belisario et al. 1999) – donnée manquante pour *C. acutatum*.

Repiquages de l'isolat de C. acutatum et de l'isolat de X. a. pv. juglandis

Le protocole de repiquage de l'isolat de *C. acutatum* est strictement identique à celui décrit précédemment ; il s'agit du même isolat. L'isolat CFBP 7179 de *X. a.* pv. *juglandis* quant à lui provient de l'Agence Nationale de Sécurité Sanitaire de l'alimentation, de l'environnement et du travail (ANSES), et plus particulièrement d'un symptôme de chancre vertical suintant d'un noyer français. Cette maladie, forme la plus agressive de la bactériose, assure une infection rapide de cet isolat. Il est repiqué sur du milieu Levure Peptone Glucose Agar (LPGA) : 10 g/L d'extrait de levure, 5 g/L de peptone, 10 g/L de glucose et 18 g/L d'agar. Ces repiquages sont maintenus à température ambiante, permettant le développement des bactéries.

Prélèvement et désinfection des folioles et des fruits immatures de noyer

Cette fois, 12 folioles et 12 fruits immatures sont prélevés pour obtenir 2 témoins (t) et 10 organes inoculés et ce, pour différentes méthodes d'inoculation testées (MI) et présentées dans le tableau suivant.

Pathogène	Folioles		Fruits		Total
X. a. pv. *juglandis*	(MI1) pulvérisation (MI2) piqûre + dépôt	10 + 2t 10 + 2t	(MI4) injection	10 + 2t	24 folioles 12 fruits
C. acutatum	(MI1) pulvérisation (MI2) piqûre + dépôt (MI3) piqûre + gélose	10 + 2t 10 + 2t 10 + 2t	(MI4) injection (MI3) piqure + gélose	10 + 2t 10 + 2t	36 folioles 24 fruits
					60 folioles **36 fruits**

Récapitulatif des différentes méthodes d'inoculation.

Les folioles sont toujours prélevées au stade Cf + 45 j et les fruits immatures au stade Gf + 45 j, soit 45 jours après le dessèchement des stigmates à la nouaison, comme conseillé dans la littérature (Bandi et al. 2015). Le protocole de désinfection est strictement identique à celui décrit précédemment, pour les deux organes. Les folioles sont toujours déposées dans des boites de Pétri sur 4 couches de papier buvard autoclavé et imbibé d'eau distillée autoclavée. Les fruits quant à eux sont déposés dans des boites à œufs alvéolées autoclavées et imbibées d'eau distillée autoclavée afin de maintenir une humidité dans les boites.

Préparation des solutions d'inoculation des deux pathogènes

La suspension de spores de *C. acutatum* est préparée selon le protocole décrit précédemment. La suspension de *X. a.* pv. *juglandis* est souhaitée à une concentration de 1.10^8 UFC/mL pour obtenir des symptômes, selon la littérature (Bandi et al. 2015; Moragrega et al. 2011). Il faut d'abord prélever une colonie du milieu LPGA, puis la mettre en suspension dans 5 mL de tampon phosphate salin (7,65 g de NaCl, 0,724 g de Na_2PO_4 et 0,210 g de $NaPO_4$ pour 1 L). Ce tampon isotonique permet de rincer les bactéries et enlever les traces de milieu PDA. La turbidité est mesurée au spectrophotomètre et un travail interne au CTIFL montre qu'une DO de 0,2 à 600 nm correspond à une concentration de 1.10^8 UFC/mL.

Mise en œuvre des différentes méthodes d'inoculation des folioles

- MI1, pulvérisation des folioles pour les deux pathogènes : il s'agit de l'aspersion de la solution d'inoculation sans blessure au préalable grâce à un pulvérisateur classique pour plantes d'intérieur (témoin : eau distillée)

- MI2, piqûre et dépôt de la solution d'inoculation pour les deux pathogènes : il s'agit du même protocole que celui de Joris. Néanmoins, la nervure centrale n'est plus testée, au regard des résultats qu'il a obtenu. Cette fois, 6 piqûres de 10 µL sont réalisées sur le limbe avec 3 de chaque côté, pour augmenter le nombre de répétitions (témoin : eau distillée)
- MI3, piqûre et dépôt de morceaux de gélose colonisés par le mycélium de *C. acutatum* : cette méhode a été utilisée pour l'étude du chancre du châtaignier provoquée par *Cryphonectria parasitica* (Newhouse et al. 2014). Ici, 4 morceaux de gélose sont déposés sur le limbre avec 2 de chaque côté (témoin : gélose PDA saine).

Mise en œuvre des différentes méthodes d'inoculation des fruits immatures

- MI3, piqûre et dépôt d'un morceau de gélose colonisé par le mycélium de *C. acutatum* : du Parafilm® M est déposé sur le morceau de gélose pour le maintenir sur le fruit (témoin : gélose PDA saine).
- MI4, injection de la solution d'inoculation pour les deux pathogènes : à l'aide d'une seringue, il est facile de traverser l'épicarpe du fruit, qui formera le brou. 3 injections de 10 µL sont réalisées tout autour du fruit (témoin : eau distillée).

Au final, les boites de Pétri contenant les folioles inoculées et les boites à œufs alvéolées contenant les fruits immatures inoculés sont placés en chambre de culture, dans les mêmes conditions décrites précédemment. 5 jours après inoculation, les prises d'images et leur traitement sont réalisés également de la même façon que dans le protocole mis en place par Joris, en ce qui concerne les folioles. Pour les fruits, les mesures de nécrose ont été effectuées à l'aide d'une règle graduée, après ouverture des fruits.

4. Résultats de caractérisation de la sensibilité des accessions à ces deux principaux pathogènes

Evaluation préliminaire de la sensibilité de l'ensemble des accessions à *C. acutatum*

Pour rappel, la collection de ressources génétiques *Juglans* de l'INRAE est une collection disponible uniquement *ex situ*. Le choix est donc très limité pour mener à bien des travaux sur les intéractions plantes-pathogènes. Néanmoins, la mise en place d'un essai préliminaire visant à évaluer la sensibilité de l'ensemble des accessions à *C. acutatum* via

l’utilisation des organes détachés à mis en exergue certains niveaux de tolérances, chez les espèces apparentées au noyer cultivé.

Les résultats issus des travaux de Joris Molina ont été valorisés dans le « Full Article » suivant, lié au poster présenté au XXX International Horticultural Congress (IHC 2018) à Istanbul en Turquie, du 12 au 16 août 2018, sous le nom de « The French walnut improvement program: preliminary investigations ».

The French walnut improvement program: preliminary investigations

A. Bernard[1,2], E. Dirlewanger[2] and F. Lheureux[1]

[1]Ctifl, centre de Lanxade, Prigonrieux, France; [2]UMR 1332 BFP, INRA, Univ. Bordeaux, Villenave d'Ornon, France

Abstract

With 36,000 tons of in-shell walnuts produced in 2016, France is in 9th world position among major producers. Walnut orchards increased by almost 20% between 2000 and 2010, and walnut is the most important crop other than apple. Nevertheless, the varietal choice does not appear to be adapted to the new future constraints and the new French walnut improvement program, led by the Ctifl, has to take into account the global warming context and the reduced use of plant protection products. In this way, two preliminary studies have been conducted regarding the effect of climate change on walnut phenology under French climate conditions, and concerning an emerging disease in French walnut orchards, a form of anthracnose, due to *Colletotrichum acutatum*. On the one hand, as regards phenology aspect, chronological data have been collected for the two most produced cultivars in France thanks to the late Eric Germain, former head of breeding program at INRA of Bordeaux from 1977 to 2007 and thanks to the work of the two French walnut experimental stations of Creysse and SENuRA. A significant advance in bud break date and male and female flowering dates was found. On the other hand, as concerns the emerging disease, a qualitative detached leaf assay (absence or presence of necrosis) have been carried out in order to quickly find out which accessions of the French germplasm (including 220 accessions of *J. regia*, the cultivated walnut tree, and 39 accessions of related species) are tolerant. Results show that, after 7 days of spores inoculation on detached leaflets, all *J. regia* accessions have specific symptoms of anthracnose due to *C. acutatum* (black to brown spots with orange acervules) while two related species, *J. cinerea* and *J. sieboldiana*, do not present any symptom, revealing a potential tolerance to *C. acutatum*.

Keywords:

Juglans regia L., French walnut improvement program, chronological data, phenology, *Colletotrichum acutatum*, detached leaflet assay

INTRODUCTION

If the main goals of walnut breeding in France are a lateral bearing habit, an increased yield, a larger fruit size, a light color, an organoleptic quality, a thicker shell, an ease of cracking, and a tolerance to walnut blight (Bernard et al., 2018), two important traits must be considered: a well adaptation to French climatic conditions, with a late budbreak due to late spring frosts, and the tolerance to *Colletotrichum acutatum*, an emerging pathogen.

Colletotrichum spp. is an ascomycete genus of fungi ranked among a 'Top 10' fungal plant pathogen list based on scientific/economic importance (Dean et al., 2012). Among the species, *C. acutatum* J.H. Simmonds (sexual stage: *Glomerella acutata*) is a widespread complex of 29 species (Damm et al., 2012) responsible for a form of anthracnose and has a wide range of host crops worldwide such as apple (Jones et al., 1996), strawberry (Xie et al., 2010) and almond (Förster & Adaskaveg, 1999). In French walnut orchards, *Colletotrichum* spp. is known to have emerged in 1998-2000 from the observation of walnuts with symptoms of Brown Apical Necrosis (Belisario et al., 2002) and it has been characterized as necrotrophic in walnut in France since 2008. Symptoms classically include the appearance of black to brown spots with orange acervules, which often leads to a completely necrotic walnut that falls prematurely. There is currently no effective mean of control to prevent this pathogen and one of the levers would be the search for tolerant genotypes.

In parallel, climate change has impacted plant phenology in last decades, particularly in favor of an advance in flowering dates (Chmielewski et al., 2001; Menzel et al., 2006; Jochner et al., 2015). If late spring frosts are known to be damaging in French walnut, it was also shown that early frost damages could negatively influence fruit yield in the following year (Charrier et al., 2017). It is therefore important to assess the effect of environment on the phenology of the main cultivars currently produced in France.

Because *C. acutatum* attack and global warming can both lead to loss of French walnut production, the objectives of this study are, on the one hand, to determine which available accessions have or do not have necrosis caused by an isolate of *C. acutatum* using an *in vitro* detached leaflet assay (DLA) with observation of symptoms at 7 days after inoculation (DAI) and on the other hand, to study the budbreak and the male/female flowering dates of 'Franquette' and 'Lara' cultivars of the last decades using chronological data.

MATERIALS AND METHODS

Evaluation of the tolerance of French germplasm to *Colletotrichum acutatum* using detached leaf assay

Fungal isolation: Several walnuts that showed typical symptoms of anthracnose disease caused by *C. acutatum* were collected in spring 2016 from trees in arboriculture experimental unit of INRA in Toulenne, near Bordeaux (France). After cleaning with water and 70% ethanol, several small square-shaped pieces of necrotic walnut husks were cut with a scalpel and placed on Potato Dextrose Agar (PDA) medium with streptomycin sulfate to inhibit growth of bacteria. Successive subculturing was performed to obtain several purified colonies.

Plant materials collection and surface sterility obtaining: 220 *J. regia* accessions and 39 of related species (including *Rhysocaryon* section (*J. nigra*, *J. hindsii*, *J. microcarpa*, *J. major*, *J. mollis*) and *Cardiocaryon* section (*J. cinerea*, *J. sieboldiana*, *J. cathayensis*, *J. mansdhurica*)) were screened. All

accessions are grafted trees located in the arboriculture experimental unit in Toulenne (latitude 44°34'37.442''N – longitude 0°16'51.48''O), near Bordeaux (France). Five mature and healthy leaflets from compound leaves randomly taken from all over the tree were hand-collected for each accession, from May to July 2017 to globally respect the age of leaflets according to budbreak dates. Surface sterility was obtained by a rinse in 70% ethanol for 50 seconds and three consecutive rinses in sterile distilled water (SDW) for 50 seconds. Each of the five leaflets is placed, on laminar flow hood, in an annotated Petri dish with adaxial surface in contact with four layers of sterile blotting paper moistened with SDW, in order to maintain leaflet turgidity with free stomata and to prevent evaporation until symptoms are observed.

Leaflets inoculation and growing conditions: A solution of germinated spores from *C. acutatum* isolate with a concentration of 1.10^5 spores/mL is produced thanks to a hemocytometer. One leaflet by accession, chosen as control, is wounded with a sterile needle and inoculated with a 10 µL drop of SDW. The four other leaves are equally wounded but inoculated with a 10 µL drop of the solution of spores. Four inoculation points are performed, two on the blade and two on the central rib. All the Petri dishes are sealed with Parafilm during transport to the climatic chamber. Growing conditions, trade-off between the host and the pathogen, are as follows: temperature of 23±1°C/day and 18±1°C/night, artificial light for 16h/day and 8h/night, and 80% humidity rate.

Symptoms observation and image processing: Symptoms are recorded 7 DAI with image capture. Image processing is then realized using APS Assess 2.0 software (Lamari, 2008) to measure the size of the necrosis.

Fungal identification: Koch's postulates could not be verified on all the 259 accessions. Thus, in order to prove the presence of *C. acutatum* in the typical necrosis, a molecular identification was performed on a randomly selected accession by PCR using specific primers based on the ITS region to *C. acutatum* (CaInt2: 5′-GGGGAAGCCTCTCGCGG-3′ and ITS4: 5′-TCCTCCGCTTATTGATATGC-3′) and to *C. gloeosporioides* (CgInt: 5′-GGCCTCCCGCCTCCGGGCGG-3′ and ITS4: 5′-TCCTCCGCTTATTGATATGC-3′) (Liu et al., 2012), a related species often confused with and also found to be pathogenic in French walnut orchards. The PCR reactions were carried out using DNA samples extracted with the Qiagen DNeasy Plant Mini Kit according to the manufacturer's instructions for both plants and fungi samples. The amplification reactions were performed in 23 µL with 2 µL DNA matrix, 5 µL 5X buffer, 0.05 µL dNTP, 0.25 µL each primer and 0.2 µL Taq polymerase. PCR conditions were as follows: denaturation at 94°C for 5 min, 35 cycles of 94°C for 1 min, 59°C for 2 min and 72°C for 2 min, and final extension at 72°C for 10 min. The PCR products were checked on 1% agarose gel by electrophoresis at 130V for 25 min. All PCR experiments were repeated two times with identical results.

Study of budbreak date and male/female flowering dates using historical data

Plant materials and phenotypical traits observations: Phenotypical data have been collected for 'Franquette' and 'Lara' cultivars, for a period from 1989 to 2016, and on three different sites: arboriculture experimental unit in Toulenne (latitude 44°34'37.442''N – longitude 0°16'51.48''O), walnut experimental station of Creysse (latitude 44°53'16.529''N – longitude 1°36'23.838''E) and walnut experimental station of SENuRA (latitude 45°8'38.565''N – longitude 5°17'34.323''E). Budbreak dates and male/female flowering dates have been observed according to IPGRI descriptors for walnut (IPGRI, 1994) improved by the late Eric Germain, former head of breeding program at INRA of Bordeaux from 1977 to 2007 (Table 1).

Table 1. Walnut descriptors

Phenological trait observed	Description	Unit	Picture
Date of bud break	When over 50% of terminal buds have enlarged and bud scales have split exposing the green of the leaves inside	In the format DDMMYYYY	
Peak male bloom date	When maximum pollen shedding occurs	In the format DDMMYYYY	
Peak female bloom date	Date of maximum pistillate flower receptivity	In the format DDMMYYYY	

RESULTS AND DISCUSSION

Evaluation of the tolerance of French germplasm to *Colletotrichum acutatum* using detached leaf assay

Fungal identification: H 109-3 accession was chosen randomly to perform fungal molecular identification. PCR products show the presence of *C. acutatum* and the absence of *C. gloeosporioides* in necrosis of H 109-3 samples (Fig. 1).

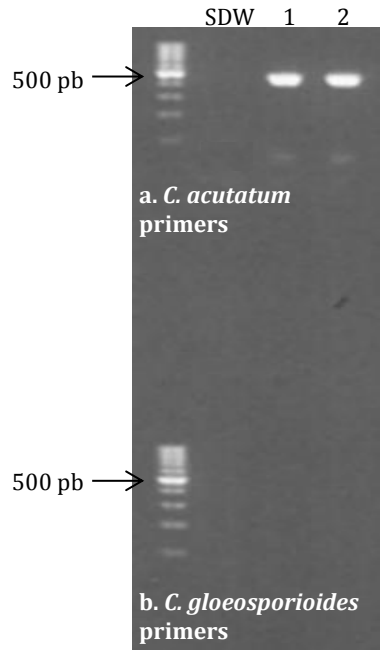


Figure 1. Molecular identification of *C. acutatum* on randomly selected H 109-3 accession necrosis. 1 + 2: necrosis samples of H 109-3; SDW: sterilised distilled water. *C. acutatum* is present in H109-3 necrosis as revealed with *C. acutatum* primers (a) and *C. gloeosporioides* is absent in H109-3 necrosis as revealed with *C. gloeosporioides* primers (b).

Symptoms observation and tolerant accessions: Inoculation of detached walnut leaflets with a solution of spores of *C. acutatum* resulted, for susceptible accession H 109-3, in a clearly defined necrosis which rapidly spread within few days (Fig. 2). If literature does not prove the presence of symptoms on leaflets in French walnut orchards, this work firstly permitted to show that *C. acutatum* is able to induce necrosis on detached leaflets within 7 DAI whereas the symptoms are mostly observed on fruits. The primary objective of this detached leaflet assay was to perform a quantitative measure of necrosis according to genotypes. However, statistical analysis performed on the repeatability of the test (Friedman test) showed that it was not able to be quantitative but only qualitative with presence or absence of necrosis. Among the 259 accessions tested, all the *J. regia* accessions showed a presence of necrosis at 7 DAI. Nevertheless, seven accessions among related species showed an absence of necrosis at 7 DAI. There were the four *J. cinerea* accessions (CN 13-3, CN 1-9, CN 19-2 and CN 8-1) and the three *J. sieboldiana* accessions (SB 1-30, SB 24-1 and SB 6-9) (Fig. 2). These *Juglans* species found to be tolerant to *C. acutatum* on our detached leaflet assay, *J. cinerea* and *J. sieboldiana*, seem to be also tolerant to other diseases according to literature. Indeed, in Belisario et al., 1999, *J. cinerea* and *J. sieboldiana* were reported to be the most resistant species to bacterial blight caused by *Xanthomonas campestris* pv. *juglandis*. In another work of Belisario et al., 2008, *J. cinerea* and *J. sieboldiana* were also reported to be the most resistant species to *Gnomonia leptostyla*, the other agent responsible for another anthracnose form. In Pollegioni et al., all *J. regia* accessions were found to be susceptible to *G. leptostyla*. They also found an effect of leaf age, as we noted in our work. In Tsiantos et al., 2007, some *J. regia* cultivars (‘Amigo’, ‘Ashley’, ‘Serr’, ‘Grand

Jean') were less susceptible to *C. acutatum* but in our case, all *J. regia* accessions could have developed necrosis at 7 DAI. This evidence of global susceptibility of *J. regia* to pathogens was also found in Arnaudov et al., 2015. Indeed, even if sensibility levels do exist, all genotypes showed symptoms of *X. arboricola* pv. *juglandis* and *G. leptostyla*. If there is little information available on the tolerance level of walnut cultivars to *C. acutatum*, this first germplasm screening suggests that the sources of strict tolerance will not be found in cultivated walnut accessions but in related species of *Juglans* genus, making it difficult in terms of genetic improvement by conventional techniques such as breeding. In addition, investigation of the tolerance level of the accessions with a detached leaflet assay was chosen, but if there are advantages such as symptoms observation on many individuals at the same time (Mo et al., 2007) and space saving, there are also many drawbacks such as a not always accurate representation of natural conditions and sometimes, an absence of correlation with orchard trials.

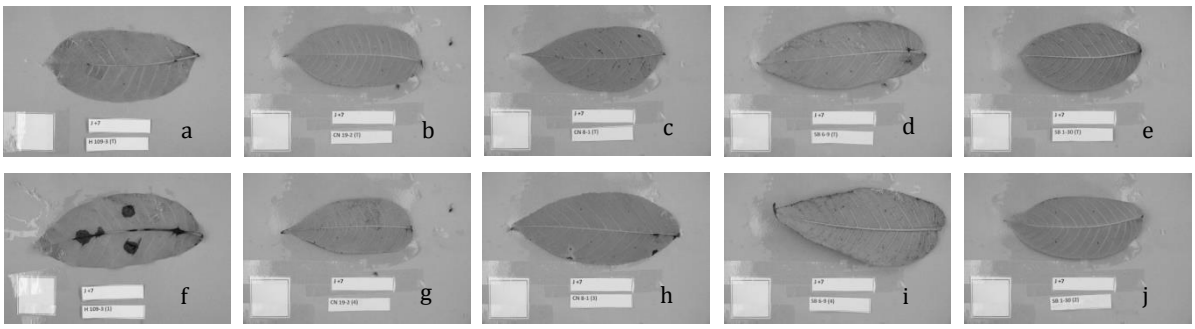
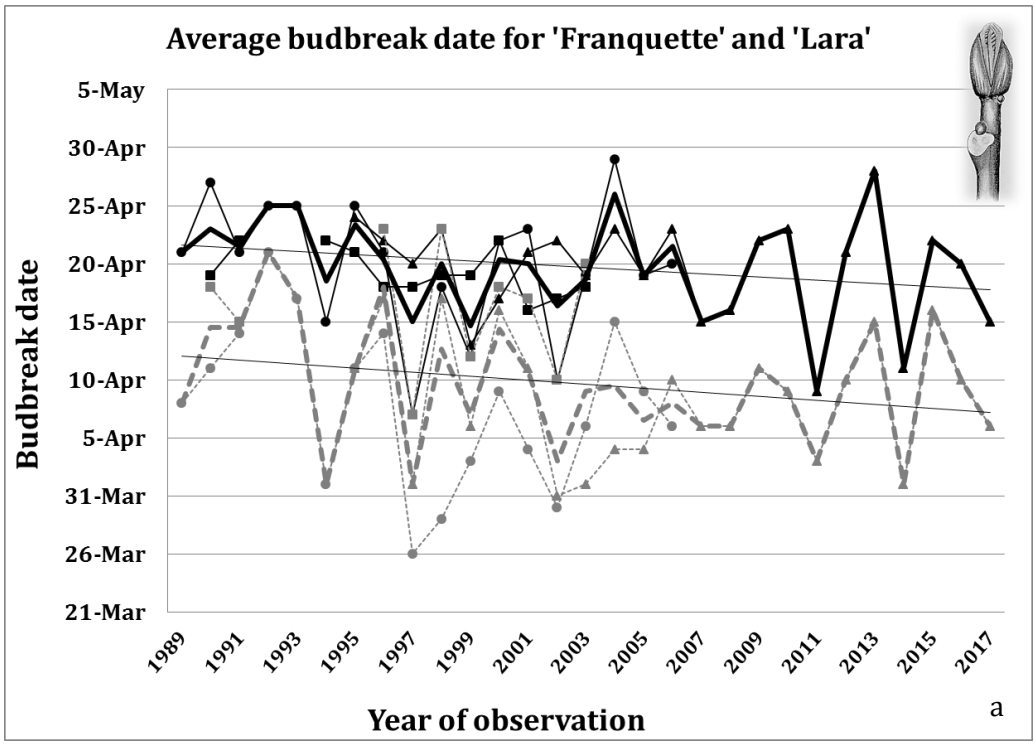


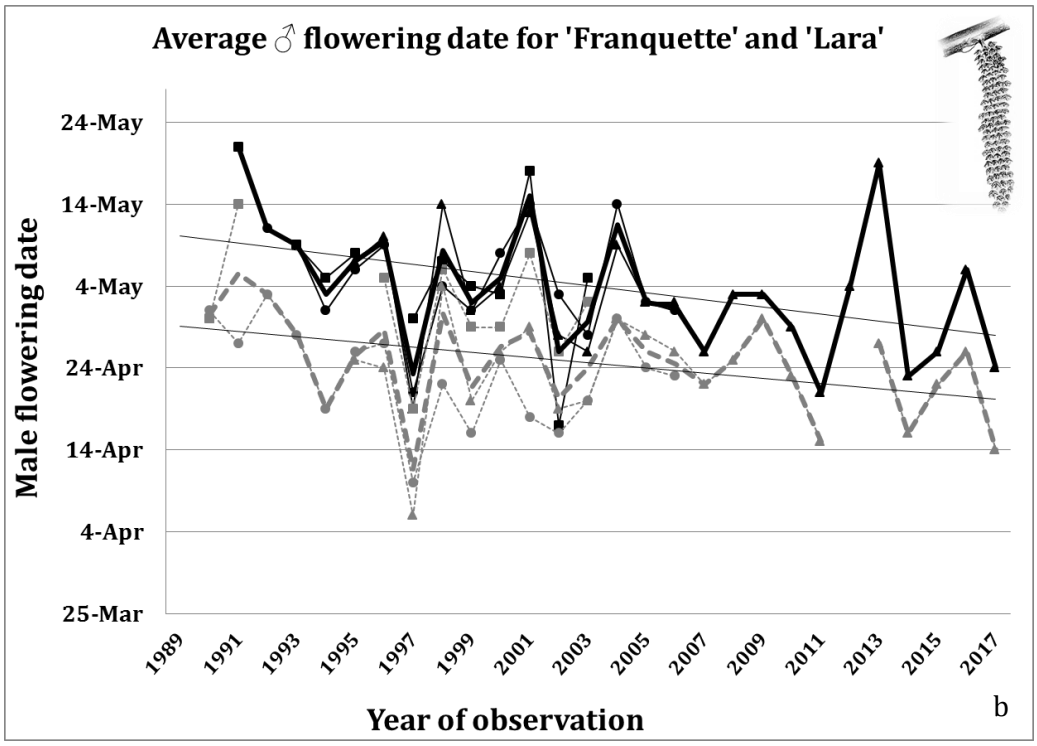
Figure 2. Observation of symptoms caused by *C. acutatum* at 7 DAI. For control leaflets, at the top, there is no evolution in the size of necrosis due to needle wound and SDW drop at 7 DAI (a, b, c, d and e). For infected leaflets, at the bottom, H 109-3 accession (f) shows necrosis at 7 DAI with characteristic symptoms: black to brown spots with orange acervules. On the contrary, for two accessions of *J. cinerea*: CN 19-2 (g) and CN 8-1 (h) and two accessions of *J. sieboldiana*: SB 6-9 (i) and SB 1-30 (j), there is no evidence of necrosis development at 7 DAI.

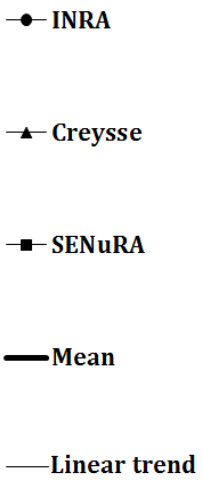
Study of budbreak date and male/female flowering dates using chronological data

Over the years, there is a tendency to an advance of phenology in 'Franquette' and 'Lara' cultivars under French climatic conditions for budbreak date and for male/female flowering dates (Fig. 3). Data are quite comparable for the three experimental stations and they follow the same fluctuations between the two cultivars. As concerns the budbreak graph, there is two years when the advance of the date was very significant: 1994 and 1997. In parallel, in France, the months of March 1994 and 1997 were recorded as being the warmer of the last thirty years, with that of 2017. In March 1994 and 1997, the mean temperature was 10.7°C against 8.7°C usually (Météo France data). On the contrary, winter 2013 vas recorded as colder than the average and there is a delay in phenology in 2013. Overall, the same results are found regarding the male flowering and the female flowering dates. It seems that the

phenology of walnut tree is sensitive and directly related to the weather conditions preceding the events, even if phenology is highly heritable, with great differences between genotypes.







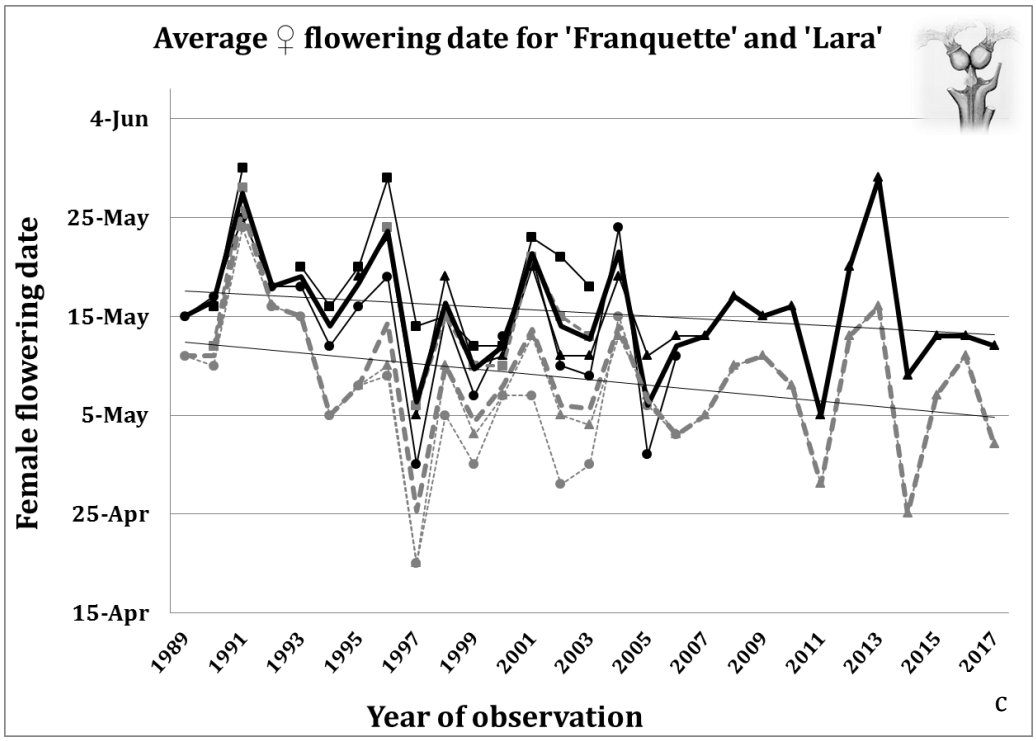


Figure 3. Phenological traits observed for 'Franquette' (solid and black line) and 'Lara' (dotted and grey line) cultivars from 1989 to 2017 on 3 experimental sites (INRA, Creysse and SENuRA). (a) Budbreak date, (b) Male flowering date, (c) Female flowering date.

CONCLUSION

One of the major challenges of the French walnut improvement program is to determine the most moderate levels of susceptibility to *C. acutatum* within the germplasm available. Thus, a number of detached fruits assays will be set up during summer 2018 on contrasting genotypes to obtain a repeatable quantitative protocol. This protocol could then be tested on the *J. regia* collection, to determine the less-susceptible accessions. In parallel, the germplasm collection will be observed regarding phenology. With pathological and phenological characterization of the collection, and with genotyping thanks to the help of the 600K SNP Affymetrix® array already initiated, we will have elements to find out genomics regions involved with those two traits by performing a Genome-Wide Association Study. This knowledge could be useful for the choice of future parents for the new breeding program.

ACKNOWLEDGEMENTS

We would like to express special thanks to the late Eric Germain, former head of breeding program at INRA of Bordeaux from 1977 to 2007. His remarkable work, then continued by Francis Delort, has given us the opportunity to use numerous data, collected during more than 30 years. We thank the teams of the two experimental stations of Creysse and SENuRA for providing additional phenological data. We really want to thank Eloïse Tranchand (Experimental Station of Creysse) for agreeing to use her pictures (presented in Table 1) on phenological stages. We thank also the team of the

Arboriculture Experimental Unit of INRA in Toulenne and the *Prunus/Juglans* Genetic Resources Center for the maintenance of the collection. Then, we thank the 'Région Nouvelle-Aquitaine' for funding the project and the 'Cifre' convention by 'ANRT' (Agence Nationale de la Recherche et de la Technologie) for funding the doctoral fellowship. Finally, we thank Joris Molina (Ctifl trainee for spring/summer 2017) for his involvement in the pathological tests.

Literature Cited

Arnaudov, V., Gandev, S., Dimova, M. (2015). Susceptibility of Some Walnut Cultivars to *Gnomonia leptostyla* and *Xanthomonas arboricola* pv. *juglandis* in Bulgaria. Agroznanje. *15*, 41–54.

Belisario, A., Maccaroni, M., Corazza, L., Balmas, V., Valier, A. (2002). Occurrence and Etiology of Brown Apical Necrosis on Persian (English) Walnut Fruit. Plant Dis. *86*, 599–602.

Belisario, A., Scotton, M., Santori, A., Onofri, S. (2008). Variability in the Italian population of *Gnomonia leptostyla*, homothallism and resistance of *Juglans* species to anthracnose. For Pathol. *38*, 129–145.

Belisario, A., Zoina, A., Pezza, L., Luongo, L. (1999). Susceptibility of species of *Juglans* to pathovars of *Xanthomonas campestris*. For Pathol. *29*, 75–80.

Bernard, A., Lheureux, F, Dirlewanger, E. (2018). Walnut: past and future of genetic improvement. Tree Genetics & Genomes. *14*, 1.

Charrier, G., Chuine, I., Bonhomme, M., Améglio, T. (2017). Assessing frost damages using dynamic models in walnut trees: exposure rather than vulnerability controls frost risks. Plant Cell Environ.

Chmielewski, F.M., Rötzer, T. (2001). Response of tree phenology to climate change across Europe. Agric For Meteorol. *108*, 101–112.

Damm, U., Cannon, P.F., Woudenberg, J.H.C., Crous, P.W. (2012). The *Colletotrichum acutatum* species complex. Stud Mycol. *73*, 37–113.

Dean, R., Van Kan, J.A.L., Pretorius, Z.A., Hammond-Kosack, K.E., Di Pietro, A., Spanu, P.D., et al. (2012). The Top 10 fungal pathogens in molecular plant pathology. Mol Plant Pathol. *13*, 414–430.

Förster, H., Adaskaveg, J.E. (1999). Identification of Subpopulations of *Colletotrichum acutatum* and Epidemiology of Almond Anthracnose in California. Phytopathology. *89*, 1056–1065.

Jochner, S., Menzel, A. (2015). Does flower phenology mirror the slowdown of global warming? Ecol Evol. *5*, 2284–2295.

Jones, A.L., Ehret, G.R., Meyer, M.P., Shane, W.W. (1996). Occurrence of Bitter Rot on Apple in Michigan. Plant Dis. *80*, 1294–1297.

Lamari, L. Assess 2.0: image analysis software for plant disease quantification. (2008). St. Paul, MN: American Phytopathological Society.

Liu, B., Louws, F.J., Sutton, T.B., Correll, J.C. (2012). A rapid qualitative molecular method for the identification of *Colletotrichum acutatum* and *C. gloeosporioides*. Eur J Plant Pathol. *132*, 593–607.

Menzel, A., Sparks, T.H., Estrella, N., Koch, E., Aasa, A., Ahas, R., et al. (2006). European phenological response to climate change matches the warming pattern. Glob Change Biol. *12*, 1969–1976.

Mo, J.Y., Navi, S.S., Li, X., Guo, T., Yang, X.B. (2007). Detached Leaf Assay: A Rapid Screening Technique to Study Foliar Diseases of Soybean. Plant Pathology Presentations and Posters.

Pollegioni, P., Van der Linden, G., Belisario, A., Gras, M., Anselmi, N., Olimpieri, I., et al. (2012). Mechanisms governing the responses to anthracnose pathogen in *Juglans* spp. J Biotechnol. *159*, 251–264.

Tsiantos, J., Vagelas, I.K., Rumbos, C.I., Chatzaki, A., Rouskas, D., Gravanis, F.T. (2007). Evaluation of resistance of cultivated walnut varieties and selections *to Xanthomonas arboricola* pv. *juglandis* in Greece. Cost 873 WG Manag Comm Meet. Athens, Greece.

Xie, L., Zhang, J., Wan, Y., Hu, D. (2010). Identification of *Colletotrichum* spp. isolated from strawberry in Zhejiang Province and Shanghai City, China. J Zhejiang Univ Sci B. *11*, 61–70.

Evaluation précisée de la sensibilité de 6 accessions à *C. acutatum* et *X. a.* pv. *juglandis*

Sur la base des résultats de Joris Molina, les principales difficultés rencontrées dans la mise en place de tels tests de sensibilité à *C. acutatum* ont été l'obtention d'une répétabilité entre les folioles d'une même accession et d'un pouvoir discriminant. Néanmoins, l'effet du génotype a été clairement observé.

Morgane Vincent a alors pu tester différentes méthodes d'inoculation (dont celle mise en place par Joris) et différents organes détachés (folioles mais également fruits) sur un panel beaucoup plus restreint de 6 accessions, permettant par la même occasion de multiplier les réplicats. Ainsi, Morgane a dans un premier temps amélioré le protocole mis en place par Joris sur foliole. Elle a pu augmenter le nombre de réplicats techniques, passant de 4 à 6 piqûres, en ce qui concerne la méthode d'inoculation MI2 pour *C. acutatum*. Le nombre de réplicats biologiques a aussi été augmenté, passant de 4 à 10 folioles testées. En plus du duo « piqûre + dépôt de suspension de spores », Morgane a aussi testé le duo « piqûre + gélose avec mycélium » (MI3) et la pulvérisation de suspension de spores pour *C. acutatum* sans piqûre préalable (MI1). Sur fruits détachés cette fois, Morgane a testé l'injection directe de suspension de spores (MI4) et le duo « piqûre + gélose » (MI3).

Le traitement statistique a permis de vérifier la fiabilité des résultats obtenus par Joris sur les folioles. Comme attendu, les accessions de *J. sieboldiana* et *J. cinerea* sont les moins sensibles. 'Franquette' se classe comme moyennement sensible. L'accession de *J. nigra* et 'UK 6-2' sont sensibles et 'Serr' est fortement sensible, comme indiqué par dans le tableau suivant. Le phénotypage a révélé des différences significatives entre les accessions, montrant ainsi un certain pouvoir discriminant.

Accession	Nombre d'observations	Moyenne nécrose (mm)	Moyenne témoins	Ecart-type	Minimum	Maximum
J. cinerea	60	2.38	0.75	1.67	0.37	8.78
J. regia 'Franquette'	60	6.38	0.26	7.28	0.29	33.17
J. nigra	60	10.37	NA	12.63	0.37	58.94
J. regia 'Serr'	60	22.71	0.40	16.94	0.73	65.31
J. sieboldiana	60	1.19	0.17	2.04	0.09	9.07
J. regia 'UK 6-2'	60	12.15	0.50	12.65	0.46	39.39

Résumé des données d'inoculation des folioles par *C. acutatum*.

Les méthodes d'inoculation MI1 et MI3 n'ont donné aucun symptôme visible sur les folioles. Si la méthode MI1 de pulvérisation sans blessure n'a pas fonctionné, cela veut potentiellement dire que les conidies ne peuvent pas pénétrer la foliole sans cette blessure. La méthode M3, avec dépôt de gélose, n'a pas fonctionné non plus car la gélose a fini par sécher, tuant l'inoculum. En ce qui concerne les fruits détachés, seule la méthode MI3 a fonctionné. Néanmoins, la mesure des nécroses à la règle graduée n'est pas des plus précises. De ce fait, la corrélation des résultats entre organes détachés n'a pas été validée. De plus, les fruits des accessions des espèces apparentées se conservent très mal et brunissent rapidement après avoir été détachés.

Concernant les travaux avec *X. a.* pv. *juglandis*, aucun symptôme ne s'est développé sur foliole ou sur fruit. Les folioles détachées perdent leur turgescence au bout de 7 jours, or d'après la littérature, il est courant d'observer des symptômes au bout de 2 à 4 semaines (Soltani and Aliabadi 2010). L'échec de contamination des fruits est plus étonnant et une des raisons pourrait être que l'isolat n'est pas assez « agressif ».

Le protocole d'inoculation idéal doit être facile à réaliser et peu couteux afin de l'appliquer à grande échelle sur l'ensemble de la collection. Par exemple, la méthode MI2 pour *C. acutatum* est simple à réaliser, mais elle prend beaucoup trop de temps. Quelques perspectives d'amélioration ont donc été proposées par Morgane. Il serait par exemple judicieux d'abandonner les tests sur organes détachés au profit de plantes en pots pour se rapprocher le plus possible des conditions d'infection naturelle, pour les deux pathogènes. Attention toutefois à la provenance des plantes en pots. Issues de greffage ou de culture *in*

vitro, le comportement des plantes peut différer. La notion d'isolat est également importante et il serait judicieux d'en tester plusieurs. Enfin, il serait intéressant de refaire une troisième fois le test sur les accessions des espèces apparentées. Un croisement interspécifique pourrait être envisagé pour d'une part, chercher dans la descendance un éventuel individu tolérant, et d'autre part, pour étudier ce caractère quantitatif via cartographie génétique et détection de QTL.

Enfin, un autre axe à approfondir est l'étude des mécanismes de défense chez les noyers, via la mise en place d'études de génétique fonctionnelle. Un très récent travail a d'ailleurs mis en exergue le rôle des transporteurs de sucres, notamment ceux de la famille des SWEET (sugar will eventually be exported transporter), dans le processus d'infection du noyer par *X. a.* pv. *juglandis* (Jiang et al. 2020). 25 gènes *SWEET* ont été distingués dans le génome de 'Chandler' (gènes *JrSWEET*) de par leur structure et leur profil d'expression dans différents tissus. Des jeunes plants issus de culture *in vitro* ont été inoculés avec un isolat de *X. a.* pv. *juglandis* et les auteurs·trices ont montré que la translocation des sucres, médiée par ces transporteurs SWEET de l'hôte vers le pathogène, n'est finalement pas l'élément déclencheur du développement de la maladie. Ces gènes sont ciblés par des effecteurs d'activation de transcription TALE (pour transcription activator-like effector) de *X. a.* pv. *juglandis* et il semblerait que ces effecteurs soient absents chez cette souche. Cela suggère par ailleurs que ce pathogène pourrait utiliser d'autres stratégies d'infection.

Étude des ressources génétiques *Juglans* en vue de la mise en place d'une sélection assistée par marqueurs

Les travaux réalisés sur le noyer ont consisté en l'exploitation des riches ressources génétiques disponibles à l'INRAE de Nouvelle-Aquitaine-Bordeaux, afin d'apporter les outils qui pourront être utilisés dans un nouveau programme de création variétale mené par le CTIFL, centre opérationnel de Lanxade. En effet, au regard du développement économique important de la noix, le choix variétal en France ne semble pas suffisant pour répondre aux futures nouvelles contraintes telles que la concurrence mondiale et le changement climatique. Le travail de prospection que l'on doit principalement à l'équipe d'Éric Germain a permis de rassembler sur l'UEA de Toulenne la majeure partie des espèces du genre *Juglans* et de nombreuses accessions de noyer cultivé, *Juglans regia* L. L'exploitation de ses archives accumulées pendant 30 ans a permis de rendre publiques d'importantes données chronologiques de phénotypage concernant cette collection. Ces données ont permis de montrer l'avancée de la phénologie des deux variétés témoins 'Lara' and 'Franquette', en lien avec le changement climatique. Grâce à un ensemble de 13 marqueurs SSR, des allèles spécifiques aux espèces *Juglans* ont été identifiés et la structure de la collection a été étudiée. Cette structure montre deux sous-groupes principaux, l'un comprenant des accessions d'Europe de l'est et d'Asie et l'autre, d'Europe de l'ouest et des Etat-Unis. Aussi, une core collection a été définie pour réaliser des études de GWAS sur les principaux caractères d'intérêt agronomique, de la fleur au fruit, grâce à l'utilisation d'une puce de 600 000 SNP mise au point par l'Université de Davis en Californie. Des associations entre des SNP et plusieurs caractères liés à la phénologie ont été mises en évidence, grâce aux données des archives et à celles nouvellement acquises. Un SNP fortement lié à la date de débourrement des feuilles et fleurs femelles a été identifié sur le chromosome 1 et co-localise avec un QTL détecté en parallèle sur une descendance F_1. Un marqueur de type KASP a été validé avec du matériel végétal de l'Université de Davis. D'autres associations ont également été identifiées pour le type de dichogamie et le type de fructification, ce caractère intervenant directement sur le rendement, et ont mené à la définition de gènes candidats. D'autres analyses GWAS ont été conduites sur les caractères liés au fruit, comme la taille de la noix, son poids, le rendement au cassage et la force nécessaire pour rompre la coque. En parallèle, des méthodes utilisant des techniques de phénotypage robustes ont été développées, comme l'utilisation de la microtomographie à rayons X pour mesurer tous les caractères morphologiques, sans casser la noix. Enfin, un travail de comparaison de l'efficacité des deux types de marqueurs utilisés dans ces travaux, SSR et SNP, a été mené. Les résultats montrent que les 13 marqueurs SSR donnent des résultats similaires à plusieurs milliers de SNP en ce qui concerne les étapes de détermination de structure et de construction de core collections, incontournables dans le management des ressources génétiques. A terme, les résultats de ces travaux permettront d'initier une sélection assistée par marqueurs pour la création de nouvelles variétés, dans le cadre d'un nouveau programme d'amélioration qui sera mené par le CTIFL. Ces nouvelles variétés seront aptes à répondre aux critères recherchés dans les années à venir, prenant en compte le changement climatique.
Mots-clés : noix, *Juglans*, GWAS, ressources génétiques, core collection, phénologie.

Study of walnut genetic resources for the implementation of marker-assisted selection

The present work carried out on walnut consisted in exploiting the rich genetic resources available at the INRAE of Nouvelle-Aquitaine-Bordeaux, in order to provide the tools that can be used in a new breeding program conducted by the CTIFL, centre opérationnel de Lanxade. In view of the walnut's significant economic development, the choice of varieties in France does not seem to be sufficient to meet future new constraints such as global competition and climate change. The prospecting work carried out mainly by Éric Germain's team has made it possible to gather most of the species of the genus *Juglans* and numerous accessions of cultivated walnut, *Juglans regia* L, at the UEA of Toulenne. The use of his archives accumulated over 30 years permitted to make public important chronological phenotyping data concerning this collection. These data showed the advance of the phenology of the two control varieties 'Lara' and 'Franquette' in relation to climate change. Using a set of 13 SSR markers, alleles specific to *Juglans* species were identified and the structure of the collection was studied. This structure shows two main subgroups, one comprising accessions from Eastern Europe and Asia and the other from Western Europe and the USA. Also, a core collection was defined to carry out GWAS studies on the main traits of agronomic interest, from flower to fruit, using a 600,000 SNP chip developed by the University of Davis in California. Associations between SNPs and several phenology-related traits were identified using archives and newly acquired data. A SNP strongly related to the budbreak date of leaves and female flowers was identified on chromosome 1 and co-localized with a QTL detected in parallel on a F_1 progeny. A KASP marker was validated with plant material from the University of Davis. Other associations were also identified for dichogamy and bearing habit, a trait directly affecting yield, and led to the definition of candidate genes. Further GWAS analyses were conducted on fruit-related traits such as nut size, weight, breaking yield and force required to break the shell. In parallel, methods using robust phenotyping techniques have been developed, such as the use of X-ray microtomography to measure all morphological traits without breaking the nut. Finally, work was carried out to compare the efficiency of the two types of markers used in this work, SSRs and SNPs. The results show that the 13 SSR markers give similar results to several thousands of SNPs with regard to structure determination and construction of core collections, which are essential in the management of genetic resources. In the long term, the results of this work will make it possible to initiate marker-assisted selection for the creation of new varieties, within the framework of a new improvement program to be carried out by the CTIFL. These new varieties will be able to meet the criteria sought in the coming years, taking into account climate change.
Keywords: walnut, *Juglans*, GWAS, genetic resources, core collection, phenology.

Unité Mixte de Recherche
UMR 1332 Biologie du Fruit et Pathologie
Bat C1 - 71 Avenue Edouard Bourlaux, CS20032, 33882 Villenave d'Ornon Cedex (France)
Tel: +33 (0)5 57 12 23 53 - Fax: +33 (0)5 57 12 25 41