



HAL
open science

**Histoire génomique d'une espèce pérenne
méditerranéenne emblématique diversité génétique,
structure, admixture, trace de sélection et adaptation
local des populations naturelles d'olivier de l'ouest du
Bassin Méditerranéen**

Lison Zunino

► **To cite this version:**

Lison Zunino. Histoire génomique d'une espèce pérenne méditerranéenne emblématique diversité génétique, structure, admixture, trace de sélection et adaptation local des populations naturelles d'olivier de l'ouest du Bassin Méditerranéen. Amélioration des plantes. Montpellier SupAgro, 2023. Français. NNT: 2023NSAM0056 . tel-04958171

HAL Id: tel-04958171

<https://theses.hal.science/tel-04958171v1>

Submitted on 20 Feb 2025

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 POUR OBTENIR LE GRADE DE DOCTEUR
DE L'INSTITUT AGRO MONTPELLIER
ET DE L'UNIVERSITE DE MONTPELLIER**

En Biologie et Ecologie Evolutives

École doctorale GAIA – Biodiversité, Agriculture, Alimentation, Environnement, Terre, Eau

**Portée par
Unités de recherche
AGAP Institut & UMR DIADE**

A genomic tale of an emblematic Mediterranean perennial species: genetic diversity, structure, admixture, footprints of selection and local adaptation of western Mediterranean Basin naturally occurring populations of olive trees

**Présentée par Lison ZUNINO
Le 19 décembre 2023**

**Sous la direction de Evelyne COSTES
Bouchaib KHADARI, Philippe CUBRY et Gautier SARAH**

Devant le jury composé de

Laurence DESPRES, Professeur des universités, Université de Grenoble Alpes

Thibault LEROY, Chargé de recherche, INRAE Toulouse

Laurène GAY, Chargée de recherche, INRAE Montpellier

Olivier FRANÇOIS, Professeur des universités, Université de Grenoble Alpes

Benjamin DAUPHIN, Collaborateur Scientifique, Swiss Federal Research Institute WSL

Rapporteuse

Rapporteur

Examinatrice

Président

Invité



**UNIVERSITÉ
DE MONTPELLIER**

 **L'INSTITUT
agro Montpellier**

Abstract

Global changes occurred for millions of years, with consecutive warming and cooling phases. For instance, multiple events of glaciation took place during the Miocene, Pliocene and Pleistocene, leading to significant species shifts or extinctions. However, current global change is distinct from past events as it is occurring very rapidly due to increased anthropogenic activities. Our assumption is that knowledge and comprehension of potential adaptation of species to climate may help to mitigate the effects of current changes. The adaptation and evolution are especially challenging for perennial plants and are a very important issue in evolutionary biology and ecology today. By looking at their genome, we can understand their potential local adaptation to the current environment. Olive tree, (*Olea europaea* L.), is an iconic perennial species from the Mediterranean Basin, which can live thousands of years. Cultivated and wild forms coexist within the same Mediterranean distribution range. Moreover, gene flow between wild and crop olives has been evidenced, leading to the emergence of admixed populations in the natural environment as well as populations escaped from cultivated olive. It is currently impossible to distinguish between genuine wild, admixed and cultivated olive trees in the natural environment because of the absence of distinct morphological features, so these forms can only be distinguished using genetic markers.

My PhD is addressing the question of local adaptation to the environment of western wild olive populations of the Mediterranean basin, through three sub-questions: (1) What is the genetic diversity and the genetic structure of naturally occurring olive trees? (2) Which parts of the genome are under selective pressure? (3) Which loci are linked to climate? To investigate the genetic structure and diversity of wild olive trees, we sampled 27 populations with 12 to 15 trees each of assumed wild olives along a north-south gradient in France, Spain and Morocco. We used newly developed target sequencing probes based on annotated genes from the publicly available reference genome 'Farga' to access part of their genome. In addition, we target sequenced accessions of cultivated olive trees from the worldwide collection of Marrakech and from a Turkish wild olive trees population. The genetic analysis revealed the occurrence of genuine wild olive trees in all countries of the sampling but also extensive presence of admixed populations in native environments. The presence of more admixed populations in the natural environment of genuine wild populations, addresses questions about their evolutionary process in that environment, which could

have been facilitated by gene flow from the cultivated olive. These results led us to look at the potential adaptation, more especially on the difference of selection of parts of genomes between admixed populations and genuine wild populations which gave clues about their parallel evolutionary histories. Moreover, we looked at their potential local adaptation to their current natural environment by examining associations between climatic and topographic data with their genomic variation by the use of genome-environment association studies. Our aim was first to determine whether the populations exhibited local adaptation to their environment and second to identify potential climate-selected genes and their putative functions. By comparing admixed and wild groups, we were able to identify potential differences in SNP selection by environment. This baseline study may pave the way for future studies aimed at determining their adaptive potential to global change.

Key-words: genome-environment association, population genomics, *Olea europaea* subsp. *europaea* var. *sylvestris*, evolutionary history, crop-to-wild gene flow

Résumé

Les changements climatiques ont lieu depuis des millions d'années, avec des phases successives de réchauffement et de refroidissement. Par exemple, plusieurs événements de glaciation se sont produits au cours du Miocène, du Pliocène et du Pléistocène, entraînant d'importants déplacements et extinctions d'espèces. Cependant, le changement global actuel est différent de ceux du passé, il se produit très rapidement en raison de l'augmentation des activités anthropiques. Dans ce cadre, la connaissance et la compréhension des mécanismes évolutifs d'adaptation des espèces à leur environnement pourrait contribuer à atténuer les effets des changements actuels. Ce mécanisme est particulièrement complexe chez les plantes pérennes et constitue actuellement un enjeu majeur d'étude en biologie évolutive et en écologie. L'apport des nouvelles techniques de séquençage et d'analyse des données en génomique permet d'étudier l'adaptation locale de ces plantes en milieu naturel sur la base de données moléculaires. L'olivier (*Olea europaea* L.) est une espèce pérenne emblématique du bassin méditerranéen, pouvant vivre des milliers d'années. La forme cultivée et la forme sauvage coexistent dans la même aire de distribution. Des flux génétiques existent entre ces deux groupes, donnant lieu à l'émergence en milieu naturel, en plus des populations sauvages, à des populations admixées. Il est actuellement impossible de distinguer les vrais oliviers sauvages des oliviers admixés et cultivés dans l'environnement naturel en raison de l'absence de traits morphologiques discriminants facilement mesurables. Cette distinction n'est possible que par l'utilisation de marqueurs génétiques. Cette thèse aborde la question de l'adaptation locale des populations d'oliviers sauvages de l'ouest du bassin méditerranéen, à travers trois sous-questions : (1) Quelle est la diversité génétique et la structure génétique des oliviers en milieu naturel ? (2) Quelles parties du génome sont soumises à la sélection ? (3) Quels sont les loci potentiellement liés aux variations environnementales ? Pour examiner la structure génétique et la diversité des oliviers sauvages, nous avons échantillonné 27 populations comprenant chacune 12 à 15 arbres, d'oliviers sauvages présumés le long d'un gradient nord-sud en France, en Espagne et au Maroc. Nous avons utilisé des sondes de séquençage ciblées sur les gènes annotés du génome de référence "Farga" pour analyser leurs caractéristiques génomiques. En plus des populations naturelles, nous avons séquencé des accessions d'oliviers cultivés de la collection mondiale de Marrakech et une population

d'oliviers sauvages originaire du sud de la Turquie. L'analyse génétique a révélé l'existence de vrais oliviers sauvages tous aussi bien dans le sud de la France, qu'en Corse, en Espagne et au Maroc ; mais aussi la présence de nombreux oliviers admixés en milieu naturel. Cette présence majoritaire d'olivier admixés au sein du même environnement que les sauvages, pose question sur le processus d'adaptation locale qui aurait pu être facilité par le flux de gènes provenant du compartiment cultivé. Ces résultats nous ont conduit à examiner les différences de traces de sélection entre le groupe admixé, le groupe sauvage et le groupe cultivé, nous renseignant ainsi sur leur histoire évolutive. Nous avons ensuite examiné l'adaptation locale potentielle du groupe admixé et du groupe sauvage à travers l'étude d'association génome-environnement. Notre objectif était d'abord de déterminer si les populations sont localement adaptées à leur environnement et ensuite d'identifier les gènes impliqués dans ce processus, ainsi que leurs fonctions. En comparant le groupe des admixés et le groupe des sauvages, nous avons pu identifier des différences dans les signatures d'adaptation locale. Ces résultats ouvrent la voie à des études futures visant à déterminer le potentiel adaptatif de l'olivier face au changement global.

Mots-clés : étude d'association génome-environnement, génomique des populations, *Olea europaea* ssp. *europaea* var. *sylvestris*, histoire évolutive, flux de gènes cultivé-sauvage

Remerciements

Avant de rentrer dans le vif du sujet, je souhaite remercier toutes les personnes qui m'ont accompagnée de près ou de loin ces trois dernières années et même plus pour certaines. Ces trois dernières années ont été très particulières dans ma vie, que ce soit au niveau du travail ou au niveau personnel et m'ont permis de grandir en tant que personne.

Je souhaite tout d'abord remercier mes quatre encadrants, sans qui rien de tout cela n'aurait été possible, merci à vous Evelyne Costes, Bouchaib Khadari, Gautier Sarah et Philippe Cubry. Merci pour votre soutien, les discussions scientifiques, votre enthousiasme et la confiance que vous m'avez accordée pour ce projet. Malgré un (long) début de thèse en confinement, vous m'avez accompagnée durant cette thèse et permis de mener au mieux ce travail. Merci Evelyne pour les échanges scientifiques qu'on a pu avoir, me permettant de mieux clarifier mes idées et mon discours. Merci à Gautier pour ton aide en bioinfo et pour ton soutien toutes les fois où j'ai débarqué dans ton bureau à l'improviste (pas facile d'être l'encadrant voisin de bureau, tu seras plus tranquille quand je serai partie). Merci à Bouchaib pour ton accompagnement sur ces trois ans, à travers les missions de terrain, mais aussi les discussions scientifiques et l'accompagnement personnel. J'ai pu gagner beaucoup en autonomie grâce à toi. Enfin, merci à Philippe, pour ces échanges scientifiques, un peu philosophiques et cet humour partagé, qui m'ont permis d'arriver jusqu'ici. Merci pour ton soutien quotidien, et d'avoir cru en moi, quand je n'y arrivais plus. Hâte d'assister à ton HDR. Merci beaucoup à vous quatre.

Je tiens aussi à remercier mon comité de suivi de thèse pour leur accompagnement tout le long de ce projet, merci à Amandine Cornille, Ivan Scotti, Joëlle Ronfort, Santiago Gonzalez-Martinez et Yves Vigouroux.

Merci aussi à mes encadrants en stage, Olivier Lepais et Rachid Hamidi, qui m'ont formée jusqu'au début de cette thèse. Ils ont été accompagnés par tous mes professeurs de l'Université de Poitiers, merci à vous pour votre formation et vos encouragements durant mon master.

Merci à toutes les personnes qui ont de près ou de loin aidé au projet, à travers la réflexion scientifique, le terrain, le labo, la bioinfo, l'administratif. Merci à Sylvia, James, Juliette, Omar, Pierre, Ronan, Hélène, Anaïs, Bénédicte, Ahmed, Gaëtan, Alexandre, Guylaine, Aude, Corine, Géraldine et Magalie, en espérant n'oublier personne.

Merci tout d'abord à tous les membres de l'équipe AFEF pour m'avoir accueillie il y a trois ans déjà, merci à Fernando, Karine, Joan, Jean-Jacques, Isabelle, Céline, Amy, Kwanho et tous les autres. Et merci beaucoup à Guillaume et Clémence, qui m'ont beaucoup épaulée durant cette dernière partie de thèse, promis, on se fait une soirée jeux bientôt.

Merci à Dynadiv, malgré les 200 mètres qui nous séparent, il a fallu deux ans pour que je passe la porte de l'équipe et je ne le regrette pas du tout. Merci à tous les gens qui m'ont accueillie chaleureusement et prêtée leur bureau pour que je puisse

travailler avec vous. Merci à Yann, Muriel, François, Julie, Christine, James, Cédric, Cécile, Maud, Nora, Cécile ainsi qu'aux doctorants. Merci aussi à Laura, Pablo, Marine, Stella, pour cette collaboration AGAP-DIADE sur l'organisation de la Journée des Doctorants qui m'a permis de finir chez vous et de vous rencontrer !

Je tiens aussi à remercier aussi les personnes du bâtiment 3/1, croisées à la pause-café, au babyfoot ou au pas des portes (conversations souvent les plus intéressantes), merci beaucoup à Estelle, Najate, Franck, Mumu, Loïc, Vincent, Maëva, Céline, Sergi, Teresa, Caro, Samer, Cédric, Florelle, Alexia et à tous ceux que j'oublie, désolée.

J'ai aussi pu rencontrer beaucoup de chouettes personnes à travers des activités transversales. Merci aux membres du CU pour m'avoir fait découvrir un autre aspect de la recherche, assez différent, mais très intéressant. Merci aussi à mes encadrants à l'Université pour l'enseignement, Mélanie et Pierre-Henri, ainsi qu'à mes co-doctorants-profs Samson, Kilian, Louise, Begüm, Florian, Arthur, bon courage pour la fin de vos thèses ! Et merci aux étudiants, j'ai adoré enseigner auprès de vous !

Enfin, je souhaite remercier les doctorants d'AGAP, en commençant avec mes ex-collègues du bureau de l'asso et représentants du CU, Dylan et Flora. C'était génial de bosser avec vous pour organiser les Agapiades/PhDays/Journées des doctorants, mais aussi de m'avoir supportée dans les hauts et les bas, et dans mes galères de déménagement (Merci aussi Léo pour le frigo !). Je crois que je vous dois toujours une soirée crêpes, oups.

Merci ensuite aux autres doctorants, ceux qui étaient là quand je suis arrivée et qui m'ont montré le chemin, j'ai l'impression que c'était hier, merci Aurélie, Kelly, Julio, Cédric, Ian et Léo ! Et merci aux nouveaux doctorants avec qui j'ai partagé ces dernières années/mois/semaines, aux nombreux MidiDoc et soirées jeux de société, Mathilde, Sofia, Clément, Baptiste, Mengyao, Joe, Mohamad et Laura. Merci aussi à Thomas, avec qui j'ai partagé deux unités, et des missions à l'étranger, merci pour ta gentillesse et ton support durant ces deux dernières années.

Je souhaite aussi remercier tous les stagiaires qui sont passés dans AGAP et qui ont dynamisé le bâtiment avec leurs bêtises, après presque un an de télétravail (okay, les doctorants ont aussi pas mal participé). Merci aux soirées plages, le festival des fanfares, et autres, ainsi que votre enthousiasme débordant. Merci à Fabian, Léa, Arthur, Olivier, Florian, Arthur, Pauline et tous les autres. Merci aussi à Sihem, pour nos fous rires et nos discussions féministes, on se voit bientôt sur les routes !

Voici venu les remerciements un peu plus personnels, de ces personnes qui sont rentrées dans ma vie, et m'ont amené à la personne que je suis aujourd'hui.

Tout d'abord je souhaite remercier les quatre personnes que j'ai rencontrées durant cette thèse, doctorants et stagiaire, Emma, Valaurie, Eva et Michel. Merci à vous quatre de m'avoir supportée dans les hauts et les bas, dans le travail comme dans la vie perso. Merci Emma, pour l'humour noir partagé, pour la mise en place du plan olivier, je promets, on le fait, et le soutien moral (ok, on a peut-être plus pleuré ensemble que réellement réussi à s'aider mutuellement). Merci à Valau, coup de cœur amitié, avec qui j'ai partagé absolument tout, même le covid. Merci pour ta joie, ton chat et ton soutien inébranlable. Merci, merci, merci. Ensuite, vient les deux derniers

acolytes de cette thèse : le premier est arrivé en même temps que moi et j'en suis très heureuse. Merci Michel pour ces trois ans passés ensemble, nos débats passionnés même les vendredis soir à 23h (désolée Eva), les baignades au lac, les randos mais aussi les soirées à boire du cidre et parler lectures, ainsi que pour la découverte de l'escalade. Cette passion, partagée avec Eva, qui nous a rassemblés tous les trois. Enfin, merci Eva, cette rencontre impromptue de dernière année, qui me donne envie de rester deux ans de plus à Montpellier pour t'accompagner et te soutenir dans ta thèse, continuer de grimper, coudre et discuter avec toi, et souviens toi « Just go up ».

Je tiens aussi à remercier dans le désordre, mes petites tortues, Laura, Amélie et Clairou, hâte de reprendre mon sac à dos pour voyager avec vous. Merci à Alice et Francesca *Grazie mille. Ora avrò tempo per imparare l'italiano*. Merci aussi à Flo, pour ton soutien durant ces trois ans ! Promis, je débarque l'année prochaine. Je souhaite aussi remercier Julie, Solène et Gaëlle, mes acolytes depuis des années, avec qui j'ai passé tant de temps et qui m'ont soutenue durant ces trois ans, on se revoit bientôt à Poitiers. Merci aussi à deux profs en particulier, Mme Rambault et M. Corona, qui m'ont fait me passionner pour la biologie, m'accrocher aux études, et sans qui, cette thèse serait probablement en sociologie.

Je souhaite remercier une personne en particulier, qui a partagé ces trois années de thèse avec moi, littéralement, merci Domitille. Rencontrée un mois en présentiel avant le covid en stage à Bordeaux, on ne s'est plus quitté depuis, malgré les confinements, puis la distance avec Montpellier, on a fait nos thèses ensemble (bon je ne comprends toujours pas bien ce qui est métabolites spécialisés) et ont fini même ensemble. On ne pourra pas s'encourager à nos soutenances, mais on l'a fait depuis le début. Merci beaucoup !

Enfin, je souhaite finir ces remerciements avec ma famille. Merci à mes grands-parents, cousins, cousines, oncles, tantes et plus particulièrement à mon frère Eliot, ma petite sœur Mélisse, ma grande sœur Lucie, Sylvain et la petite Yaëlle, rayon de soleil depuis ton arrivée. Merci à vous tous pour votre aide durant ces trois ans, mais aussi depuis qu'on partage notre vie ensemble, et même si tout n'est pas toujours rose, merci d'être toujours là, et pour le plus longtemps possible. Pour finir, merci à mes parents, d'avoir toujours été présent, bienveillant et de m'avoir poussé dans mes projets, quels qu'ils soient. Merci à mon père de m'avoir enseigné la curiosité et le goût de l'environnement. Et merci à ma mère de m'avoir donnée le goût du social et l'humanité. Je vous aime.

Merci à toutes les personnes qui ont réalisé des dessins pour ce manuscrit :

Illustration Introduction et Discussion – Emma Clément

Illustration Chapitre 1 et 3 – Fabien Zunino

Illustration Chapitre 2 – Eva Coindre

Summary

Résumé substantiel.....	1
1 Introduction	3
2 Matériels et méthodes.....	6
3 Chapitre 1	7
4 Chapitre 2	9
5 Chapitre 3	10
6 Discussion et conclusion générale	11
Introduction	15
1 Global change, past and current	17
1.1 Global	17
1.2 The case of the Mediterranean Basin	23
2 How do environmental factors and their fluctuations impact species populations?	24
2.1 Species response mechanisms to cope with changing environment	24
2.2 Local adaptation: genomic variations between populations	27
2.3 The impact of domestication on natural populations: the case of crop-to-wild gene flow.	29
3 How to study local adaptation?	30
3.1 Direct approach: reciprocal transplant experiment and/or common gardens	31
3.2 Indirect approach: Genomic studies of local adaptation.....	32
4 A model of study: wild olive tree.....	33
4.1 Biology of <i>Olea europaea</i> L.	33
4.2 Evolutionary History of <i>Olea europaea</i> L.	34
4.3 Domestication history of <i>Olea europaea</i> L.	38
5 Research project ClimOliveMed and objectives of the PhD	40

Chapter 1	43
1 Introduction	47
2 Methods	50
2.1 Sampling of wild olive trees on a north-south gradient in the western Mediterranean Basin.....	50
2.2 Reference set of cultivated and eastern wild accessions.....	53
2.3 Bait design	53
2.4 Library preparation and sequencing	54
2.5 Genomic analysis	55
3 Results.....	57
3.1 Target sequencing efficiency	57
3.2 Genetic diversity and genetic structure	57
3.3 Inference of population admixture and gene flow	62
4 Discussion.....	65
4.1 Genetic variation patterns in natural and cultivated olive trees highlight the persistence of genuine wild olive populations in the western Mediterranean Basin	66
4.2 Wild versus admixed olive trees: an evolutionary history impacted by domestication.....	67
4.3 What factors could influence the prevalence of admixed populations in natural environments?	68
4.4 Consequences of extensive hybridization of wild olive via domesticated olive introgression and conservation recommendations.....	70
5 Conclusion	70
Acknowledgements.....	71
References.....	72
Chapter 2	79
1 Introduction	82

2 Methods	85
2.1 Sampling and sequencing.....	85
2.2 SNP filtering.....	86
2.3 Genetic diversity and genetic structure	87
2.4 Detecting traces of selection	87
2.5 Characterisation of outlier loci.....	89
2.6 Admixture patterns detection and potential geneflows.....	89
3 Results	90
3.1 Target sequencing characterization	90
3.2 Genomic diversity, LD decay and genomic structure	91
3.3 Detection of selection footprints.....	94
3.4 Biological significance of outlier SNPs.....	96
3.5 Admixture patterns detection and potential geneflows.....	98
4 Discussion.....	100
4.1 Three genetic patterns (wild, domesticated and admixed) and three evolutionary trajectories in the western Mediterranean Basin.....	102
4.2 Selection footprints as a genetic signature of olive: supportive information of evolutive history?	104
5 Conclusion	106
References.....	108
Chapter 3	113
1 Introduction	116
2 Methods	119
2.1 Sampling and sequencing.....	119
2.2 SNP calling and SNP filtering conducted on the basis of genetic groups: Wild, admixed and cultivated olive trees.....	122
2.3 Selection of environmental variables	123
2.4 Genomic environment association analysis	124

2.5 Gene functional annotation, prediction of mutation and gene-ontologies enrichment analysis	125
2.6 Characterisation of the candidate SNPs for association to the environment according to genotypic groups	126
3 Results	127
3.1 Identification of candidate SNPs in wild and admixed populations	127
3.2 Description of associated SNPs and genes in wild and admixed populations	131
3.3 SNP characterisation and localisation along the geographic gradient according to wild and admixed groups.....	136
4 Discussion.....	138
4.1 Local adaptation of western wild olive trees to climate?	140
4.2 Local adaptation of admixed olives trees?	142
4.3 Differential local adaptation between wild and admixed olive trees: a potential help to cope with climate change?.....	144
5 Conclusion	146
References.....	146
Discussion.....	153
1 Confirmation of previous results on olive genetic structure and new insights	155
2 Evolutionary history of the wild, cultivated and admixed olives	156
3 Selection footprints and genome-environment association	158
4 Local adaptation and adaptive potential in front of climate change.....	160
5 Importance of conservation of wild olive trees and potential impact of this research on agricultural practices	161
General conclusion.....	165
References	169
Supplementary Chapter 1	180
2 Methods	182

DNA extraction:.....	182
Library preparation:.....	182
Cluster assignment to create subgroup inside cultivated:.....	183
Supplementary Chapter 2	204
2 Methods	205
2.1 Sampling and sequencing.....	205
2.2 SNP filtering.....	205
3 Results	206
3.1 Target sequencing characterization	206
3.2 Genomic diversity, LD decay and genomic structure	207
3.3 Biological significance of outlier SNPS	209
References.....	209
Supplementary Chapter 3	210
1 Methods	211
2.1 Sampling and sequencing.....	211
2.3 Selection of environmental variables	213
3 Results	217
3.1 Identification of candidate SNPs in wild and admixed populations	217
3.2 Description of associated SNPs and genes in wild and admixed populations	227
3.3 SNP characterisation and localisation along the geographic gradient according to wild and admixed groups.....	230

Résumé substantiel



1 | Introduction

Les changements climatiques globaux existent depuis des millions d'années à travers des phases consécutives de réchauffement et de refroidissement de la planète. Ces changements entraînent de profondes modifications des environnements, avec des changements de températures, de régimes de précipitations ou encore de concentration en CO₂. Ces modifications ont impacté directement les espèces avec des changements d'aire de distribution, ayant pour conséquence des migrations ou des extinctions massives (Benton & Twitchett 2003). Depuis la période industrielle, dans les années 1800, les activités humaines ont largement impacté les déterminants physiques du climat. Les émissions de gaz à effet de serre induites par les activités anthropogènes provoquent une augmentation significative de la moyenne globale de température. Ce changement global, très rapide à l'échelle des temps géologiques, provoque une modification importante de l'environnement à laquelle les espèces sont confrontées. Il est par exemple prédit une disparition de près de 50% des espèces marines tropicales dans le scénario où la température moyenne mondiale augmente de 4°C (Calvin *et al.* 2023).

Un changement dans les conditions environnementales d'un habitat entraîne de nombreux impacts sur les espèces qui y vivent. Il existe quatre réponses possibles de la part des espèces quand leur environnement n'est plus habitable (Aitken *et al.* 2008; Kelly 2019). 1) La migration, les espèces fuient pour aller vers de nouveaux habitats aux conditions plus favorables. Cette stratégie de survie est très dépendante de la biologie de l'espèce, plus difficile à mettre en place pour les organismes à capacité de migration restreinte comme les plantes et arbres. 2) La plasticité, c'est-à-dire l'expression de différents phénotypes en fonction de l'environnement à partir d'un même génotype. Ce mécanisme permet de survivre dans des environnements aux conditions différentes. 3) L'adaptation, l'environnement exerce une pression de sélection importante et sélectionne les individus avec des caractères plus adaptés, qui auront ainsi une meilleure valeur sélective. Les individus résultants seront localement adaptés à leur environnement. 4) Si aucun des mécanismes cités précédemment ne peut être mis en place, les organismes peuvent disparaître, entraînant l'extinction de populations ou d'espèces. Chez les espèces pérennes, si la migration est un mécanisme qui prend beaucoup de temps, la plasticité et l'adaptation sont souvent

RÉSUMÉ SUBSTANTIEL

mobilisées (Alberto *et al.* 2013). L'adaptation locale d'une population est une réponse efficace et durable, qui permet la sélection de différents variants génomiques adaptés à différentes conditions environnementales. Ce mécanisme évolutif est très étudié aujourd'hui afin de comprendre son fonctionnement et sa mise en place au sein des populations. La plupart des approches utilisées passent par la création d'un jardin commun, ou par la transplantation réciproque (Kawecki & Ebert 2004; Savolainen *et al.* 2013; Alberto *et al.* 2013; Blanquart *et al.* 2013; Lascoux *et al.* 2016). Un jardin commun est un dispositif expérimental dans lequel des individus de différentes populations sont implantés dans un même environnement afin de comparer leur fitness. Dans le cas de la transplantation réciproque, les individus d'une population sont déplacés dans une autre population et inversement, et leur différence de fitness est là encore mesurée. Si les individus locaux ont une meilleure fitness dans leur environnement d'origine, ainsi qu'une meilleure fitness que les populations transplantées, alors ils sont considérés comme localement adaptés (Home vs away and local vs foreign criterias) (Kawecki & Ebert 2004). Ces approches pour étudier l'adaptation locale sont très efficaces mais peuvent être longues à mettre en place en raison des facteurs biologiques des espèces, comme c'est le cas pour les arbres, avec leur temps de génération longue. Aujourd'hui, l'apport des nouvelles technologies de séquençage permet d'étudier l'adaptation directement en milieu naturel, et plus facilement sur ces organismes (Sork *et al.* 2013; Steane *et al.* 2014). Dans ces nouvelles approches, on retrouve par exemple la détection de traces de sélection, basée sur différentes métriques tel que le déséquilibre de liaison ou le D de Tajima pour détecter les parties du génome soumises à sélection positive (Nielsen 2005). Une autre approche fortement utilisée est l'association génome-environnement, qui mesure la corrélation entre les variations alléliques et celle de métriques environnementales comme le climat (Steane *et al.* 2014).

Pour étudier le mécanisme d'adaptation des plantes, et plus particulièrement des espèces pérennes, cette thèse se concentre sur un arbre emblématique du bassin méditerranéen (BM), l'olivier, *Olea europaea* ssp. *europaea* L., pouvant vivre plusieurs centaines d'années (Ozenda 1975; Rivas Martínez *et al.* 1987). Cet arbre possède un système d'auto incompatibilité sporophytique (Saumitou-Laprade *et al.* 2017), est pollinisé par le vent (anémophile) et ses graines sont dispersées par les animaux (zoochore), ce qui entraîne un flux génétique important entre les populations (Spennemann & Richard Allen 2000). L'olivier sauvage et l'olivier cultivé sont

distribués dans la même zone géographique (Green 2002). La distinction entre les oliviers sauvages et les oliviers cultivés est complexe, en raison de l'absence de caractéristiques phénotypiques facilement mesurables sur le terrain (Terral *et al.* 2004, 2021). Dans ce contexte, les marqueurs génétiques sont nécessaires pour une identification précise (Besnard *et al.* 2013b; Khadari & El Bakkali 2018; Gros-Balthazard *et al.* 2019). Des flux de gènes ont été observés entre les oliviers cultivés et sauvages dans l'est du BM, mais très peu dans l'ouest (Gros-Balthazard *et al.* 2019; Julca *et al.* 2020). L'étude de l'histoire évolutive de l'olivier sauvage, notamment par des approches paléo-botaniques, montre qu'ils sont présents dans cette région depuis des millions d'années (Planchais 1984; Carrión *et al.* 2010). Des événements tels que la crise messinienne (5,33 Ma) ont conduit à l'apparition de deux pools génétiques distincts au sein du BM, l'un à l'est et l'autre à l'ouest (Terral *et al.* 2009; Besnard *et al.* 2013b; Khadari & El Bakkali 2018; Besnard *et al.* 2018). Plusieurs événements de glaciation durant l'Holocène ont par la suite fragmenté les populations, seules certaines ayant pu survivre dans des refuges climatiques. La fin de la glaciation a permis la recolonisation et la reconnexion entre les populations rémanentes (Carrión *et al.* 2010).

Pour les oliviers cultivés, le processus de domestication a commencé il y a environ 6000 ans, principalement dans la région du Levant, près de la frontière entre la Turquie et la Syrie (Zohary & Spiegel-roy 1975; Kaniewski *et al.* 2012; Besnard *et al.* 2013b). L'origine de l'olivier cultivé est très débattue. Certains suggèrent un événement de domestication unique dans l'est du BM, suivi de son expansion dans le reste du bassin, tandis que d'autres évoquent la possibilité d'événements de domestication multiples (Besnard *et al.* 2013a; Diez *et al.* 2015). Des recherches génomiques récentes suggèrent un scénario intermédiaire, avec une domestication primaire au Levant, suivie d'échanges génétiques avec des populations sauvages locales dans tout le BM, permettant l'implantation plus rapide des oliviers cultivés, et leur différenciation génétique, pouvant être considéré comme des événements secondaires de domestication (Julca *et al.* 2020). L'introduction de la culture de l'olivier dans la partie ouest du BM s'est produite à travers au moins deux événements majeurs, une première fois dans le sud de la péninsule ibérique, il y a environ 2 500 ans, puis une seconde fois, dans le sud de la France, lors de l'arrivée des Romains (environ 100 av. J-C) (Leveau *et al.* 1991; Langgut *et al.* 2019). La distribution actuelle des oliviers se compose de trois groupes génétiques. Les variétés du groupe Q1 dans

RÉSUMÉ SUBSTANTIEL

la partie ouest du BM sont très étroitement liées, tandis que celles des groupes Q2 et Q3, couvrant les régions centrale et est, montrent une plus grande séparation génétique, suggérant une histoire de domestication de l'olivier et de sa dispersion dans le BM complexe (Diez *et al.* 2015; Besnard *et al.* 2018).

Cette thèse fait partie du projet de recherche ClimOliveMed, qui cherche à comprendre les défis de l'adaptation au changement global, de la conservation de la biodiversité et de la production durable dans le contexte de l'oléiculture méditerranéenne. Ce doctorat se concentre sur l'étude de l'histoire évolutive des populations naturelles d'oliviers supposées sauvages dans le BM occidental. Son objectif principal est d'évaluer les preuves d'adaptation et de sélection locales au sein de ces populations et en examinant trois sous-questions spécifiques : (1) Quelles sont la diversité et la structure génétique présentes dans les populations naturelles d'oliviers ? (2) Quelles parties du génome sont soumises à pression sélective ? (3) Existe-t-il des loci génétiques susceptibles d'être associés à des variations environnementales au sein de ces populations ? Ce manuscrit est divisé en trois chapitres : 1. Preuve génomique de la présence d'oliviers sauvages et d'olivier admixés évoluant dans les mêmes environnements naturels dans l'ouest du bassin méditerranéen. 2. Traces de sélection chez *Olea europaea* subsp. *europaea* : détection dans les oliviers sauvages, admixés et cultivés de l'ouest bassin méditerranéen - Etude de leur histoire évolutive 3. Les flux de gènes des espèces cultivées vers leur apparenté sauvage entraînent-ils un différentiel d'adaptation des espèces à leur environnement naturel ? Etude de l'adaptation locale des oliviers sauvages à leur environnement.

2 | Matériels et méthodes

Vingt-sept populations ont été échantillonnées le long d'un gradient latitudinal de 13°, comprenant le Sud de la France, la Corse, l'Espagne et le Maroc. Les sites ont été choisis sur la base d'une étude précédente (Besnard *et al.* 2013a), ainsi que sur la base des données d'occurrence du Conservatoire Botanique National pour la France. Les sites proposés ont ensuite été triés, d'après des bases de données d'occupation des sols (Copernicus Land Monitoring Service 2020) pour éviter les oliveraies et les zones urbanisées et sur la base de données climatiques publiques (Fick & Hijmans 2017), pour obtenir une sélection de climats variés entre les sites et

éviter des redondances locales. Pour chaque site, 13 à 15 arbres ont été échantillonnés, en récupérant et séchant les feuilles pour l'extraction d'ADN. Au total, près de 400 arbres ont été étudiés, auxquels ont été ajoutés 135 échantillons cultivés présents dans la collection mondiale des ressources génétiques de l'olivier de Marrakech, ainsi que 10 génotypes cultivés provenant des collections françaises, et une population sauvage préalablement caractérisée du Sud de la Turquie ($n = 15$; Gros-Batkthazard *et al.*, 2019). Les données climatiques ont été récupérées pour chaque individu, en se basant sur la base de données climatiques CHELSA (Karger *et al.* 2017; Brun *et al.* 2022).

Pour accéder à l'information génomique des individus, nous avons réalisé un séquençage par capture ciblée du génome. À partir, du génome de référence 'Farga', (Julca *et al.* 2020), une variété de *O. europaea* ssp. *europaea*, 210 367 sondes ont été dessinées afin de cibler les 55 595 gènes annotés. Cela représente une taille de capture du génome de 16,8 Mb, sur une taille totale de 1,38 Gb. Les données de séquences brutes ont ensuite été nettoyées, alignées sur le génome de référence puis un appel de variant et un appel de variant conjoint ont été réalisés. Les données ont ensuite été filtrées avant d'être utilisées pour les analyses de génomique, détaillées dans les chapitres suivants.

3 | Chapitre 1

Des flux de gènes entre les espèces cultivées et leurs apparentés sauvages ont été identifiés dans plusieurs études, que ce soit chez les animaux (Johnsson *et al.* 2016; Anderson *et al.* 2021; Mary *et al.* 2022) ou les plantes (Hauser & Shim 2007; Le Corre *et al.* 2020). Ce phénomène résulte en l'apparition de variants génétiques cultivés au sein de populations naturelles. L'apparition de ces populations admixées soulève la question de l'impact des flux de gènes sur l'évolution des populations naturelles. Par exemple, l'introduction d'une nouvelle diversité génétique dans les génomes sauvages pourrait accélérer leur évolution en augmentant la fréquence d'allèles bénéfiques, mais cela peut également avoir des effets indésirables en entraînant une perte de fitness chez les hybrides par rapport à leurs parents (Ellstrand *et al.* 1999; Keller *et al.* 2000; Gering *et al.* 2019). L'identification et la caractérisation génétique des oliviers sauvages est d'une grande importance pour évaluer la diversité génétique de cette espèce, retracer son histoire évolutive et étudier son adaptation

RÉSUMÉ SUBSTANTIEL

locale. C'est particulièrement crucial pour des espèces sauvages très apparentées, comme l'olivier, où les oliviers sauvages et les oliviers cultivés cohabitent dans la même région (Zohary & Spiegel-roy 1975; Rubio de Casas *et al.* 2006) et ne peuvent être clairement distingués sur la base de caractéristiques morphologiques (Terral *et al.* 2004, 2021), mais uniquement à l'aide de marqueurs génétiques (Lumaret & Ouazzani 2001; Besnard *et al.* 2013a; Diez *et al.* 2015; Khadari & El Bakkali 2018; Gros-Balthazard *et al.* 2019; Julca *et al.* 2020). La présence de flux de gènes entre l'olivier sauvage et l'olivier cultivé a été mise en évidence dans l'est du BM, mais pas dans l'ouest (Besnard *et al.* 2013b; Khadari & El Bakkali 2018; Gros-Balthazard *et al.* 2019). Étant donné qu'il existe probablement des échanges génétiques entre les cultures et les populations sauvages, la compréhension de leurs relations génétiques peut contribuer à éclairer l'impact des cultures sur les populations naturelles, leur histoire démographique et à explorer de nouvelles sources de diversité génétique.

Dans le cadre de ce Chapitre 1, notre objectif était d'identifier les véritables populations d'oliviers sauvages en étudiant la structure génétique et les schémas de diversité des oliviers évoluant dans l'environnement naturel dans l'ouest du BM. Nous avons prélevé des échantillons d'oliviers supposés sauvages, conformément à une étude antérieure (Besnard *et al.* 2013a), sur une vaste zone géographique de l'ouest du bassin. Nous avons fait l'hypothèse que toutes ces populations étaient constituées d'oliviers sauvages. De plus, nous avons inclus des oliviers sauvages du sud de la Turquie et 145 oliviers cultivés représentatifs de la diversité génétique de l'ouest du BM. Nous avons analysé ce large échantillonnage de populations en utilisant des SNP issus d'un séquençage ciblé. Il s'agit de la première utilisation du séquençage ciblé pour étudier la variation génomique de l'olivier, permettant de balayer le génome de nombreux individus tout en accédant à plus de 140 000 SNP répartis sur l'ensemble du génome.

Nous avons identifié la présence de véritables populations d'oliviers sauvages dans la région méditerranéenne occidentale, en France continentale, en Corse, en Espagne et au Maroc. De nombreuses populations naturelles résultant d'un flux de gènes significatif entre les cultivés et les populations sauvages ont également été observées. La présence de populations d'oliviers admixées coexistant avec les populations sauvages soulève des questions sur les raisons de leur prévalence dans l'environnement naturel et sur les stratégies de conservation pour les deux groupes. Nos résultats apportent un éclairage sur l'état actuel des populations naturelles et

suscitent de nouvelles questions. La première concerne la nature du mélange, c'est-à-dire quelles allèles/régions génomiques domestiquées sont les plus susceptibles d'être retrouvées dans les génomes sauvages. La deuxième question est liée à l'adaptation locale : les oliviers sauvages sont-ils mieux adaptés localement que les oliviers admixés ?

4 | Chapitre 2

Les traces de sélection sont des variants génomiques, sélectionnés positivement, au travers des contraintes du milieu dans lequel vit l'organisme étudié. Ces contraintes, aussi appelées pressions de sélections, peuvent englober divers facteurs environnementaux, qu'ils soient abiotiques tels que la température et les précipitations, ou biotiques tels que les pathogènes (Mitchell-Olds *et al.* 2007; Jordan *et al.* 2017; Capblancq *et al.* 2023). De plus, les traces de sélection peuvent résulter de la sélection artificielle. Chez les plantes pérennes, les traces de sélection proviennent souvent de processus d'adaptation locale, fournissant des informations sur l'histoire évolutive des espèces dans leur aire naturelle.

Chez l'olivier, la forte présence d'individus admixés, résultant de l'hybridation entre des populations sauvages et cultivées, dans l'environnement naturel de l'ouest du MB, montré dans le Chapitre 1, soulève des questions sur le potentiel adaptatif des populations naturelles d'oliviers et les pressions de sélection qui ont permis la coexistence des trois groupes génétiques dans les mêmes zones géographiques. En utilisant le séquençage ciblé présenté dans le chapitre 1, et en ajoutant les informations de séquençage de scaffolds, nous avons réalisé une analyse génomique de la sélection au sein de différents groupes génétiques d'oliviers *O. europaea* ssp. *europaea* dans l'ouest du BM. Nous avons ainsi pu identifier des traces de sélection distinctes chez les oliviers cultivés, sauvages et admixés dans l'ouest du BM, permettant d'établir des hypothèses sur leur histoire évolutive respective dans cette région. La sélection observée chez les sauvages et les cultivés résulte d'une signature génétique de leur histoire évolutive ancienne dans le BM, ainsi que plus récente chez les cultivés. En revanche, les génotypes admixés, issus d'évènements d'hybridation entre des génotypes sauvages et cultivés dans l'ouest du BM, semblent avoir une histoire évolutive plus récente, et reflètent des informations évolutives issues du cultivé et du sauvage. Nous ne pouvons pas affirmer s'ils descendent directement des

RÉSUMÉ SUBSTANTIEL

génotypes sauvages avec un flux de gènes important, provenant des cultivés, ou si c'est le scénario inverse, voire une combinaison des deux.

Notre étude soulève des questions concernant l'histoire démographique des groupes génétiques sauvages, admixés et cultivés. Pour répondre à ces questions, nous proposons comme perspective l'utilisation de modèles de simulation par coalescence. De tels modèles pourraient aider à élucider quel scénario est le plus probable et à fournir davantage d'informations sur la dynamique de population des différents groupes d'oliviers de l'ouest du BM.

5 | Chapitre 3

Aujourd'hui, les organismes sont menacés par d'importantes modifications de leur environnement, dues aux changements globaux provoqués par l'activité humaine depuis l'avènement de l'ère industrielle (Calvin *et al.* 2023). Cela a entraîné un changement significatif du climat à travers le monde, notamment une augmentation de la température moyenne mondiale et des changements dans les régimes de précipitations. Dans le BM, les conséquences de ces changements globaux sont déjà constatés, et se caractérisent par une augmentation de la température moyenne 20% plus rapide que la moyenne mondiale (Lionello & Scarascia 2018). Cette région du monde est un *hotspot* de biodiversité, abritant une importante flore endémique (Medail & Quezel 1997). Le changement climatique constitue une menace importante pour la survie et la distribution des espèces dans le BM. Pour comprendre comment les populations naturelles pérennes peuvent réagir face aux changements globaux, il est important d'étudier les mécanismes de sélection impliqués dans les changements climatiques passés. Cela permettrait d'aider à prédire les réponses des espèces face au changement global actuel et de mettre en place des mesures de conservation (Steane *et al.* 2014; Hoffmann *et al.* 2015).

Dans ce troisième chapitre, nous avons étudié l'adaptation locale des populations d'oliviers sauvages et admixés de l'ouest du BM. Nous avons exploré les associations génomiques potentielles avec le climat grâce à une analyse d'association génome-environnement (GEA), utilisant des variables climatiques (CHELSA; Karger *et al.* 2017), des données d'altitude (Danielson & Gesch 2011), ainsi que le jeu de données de séquençage des populations naturelles issu des chapitres précédents. Sur la base des résultats d'association de chaque population, sauvage et admixée,

nous avons examiné les annotations du génome de référence (Julca *et al.* 2020) pour déterminer les fonctions potentielles, corrélées à ces variables environnementales. Nous avons identifié plus de 1000 SNP associés à des facteurs environnementaux pour chaque groupe génétique. Ces découvertes nous ont permis de caractériser l'adaptation locale des oliviers sauvages et des oliviers admixés, d'élucider les fonctions potentielles des gènes candidats pour cette adaptation et de comprendre l'importance des facteurs climatiques. Nous avons mené en plus, une analyse comparative entre les deux groupes génétiques pour évaluer la présence importante des oliviers admixés dans l'environnement naturel.

Les résultats ont permis de montrer que l'olivier sauvage de l'ouest, *Olea europaea* L., est localement adapté à son environnement, principalement en relation avec des variables de températures. Chez les admixés, l'analyse n'a pas permis de mettre en évidence de réelle adaptation locale, pourtant testée sur des individus échantillonnés dans les mêmes habitats. Cependant, le génome des oliviers admixés conserve en moyenne 80% des variants génomiques détectés comme associés au climat chez les sauvages, alors que génétiquement beaucoup plus proche des oliviers cultivés. Il contient, de plus, d'autres variants à priori sélectionnés, cette fois, issus du cultivé, apportant des gènes de résistance aux pathogènes par exemple. Ainsi, les populations admixées ont une diversité génétique plus importante, avec une augmentation de leur variation génétique latente, résultat de l'hybridation entre le cultivé et le sauvage. Cela pourrait potentiellement permettre une réponse adaptative plus rapide de ces populations à des environnements nouveaux ou changeants. Ce troisième chapitre ouvre la voie à la prédiction de futures adaptations ou au calcul de la différence génomique d'adaptation actuelle des populations naturelles d'oliviers dans l'ouest du BM, à celle requise pour survivre dans le futur.

6 | Discussion et conclusion générale

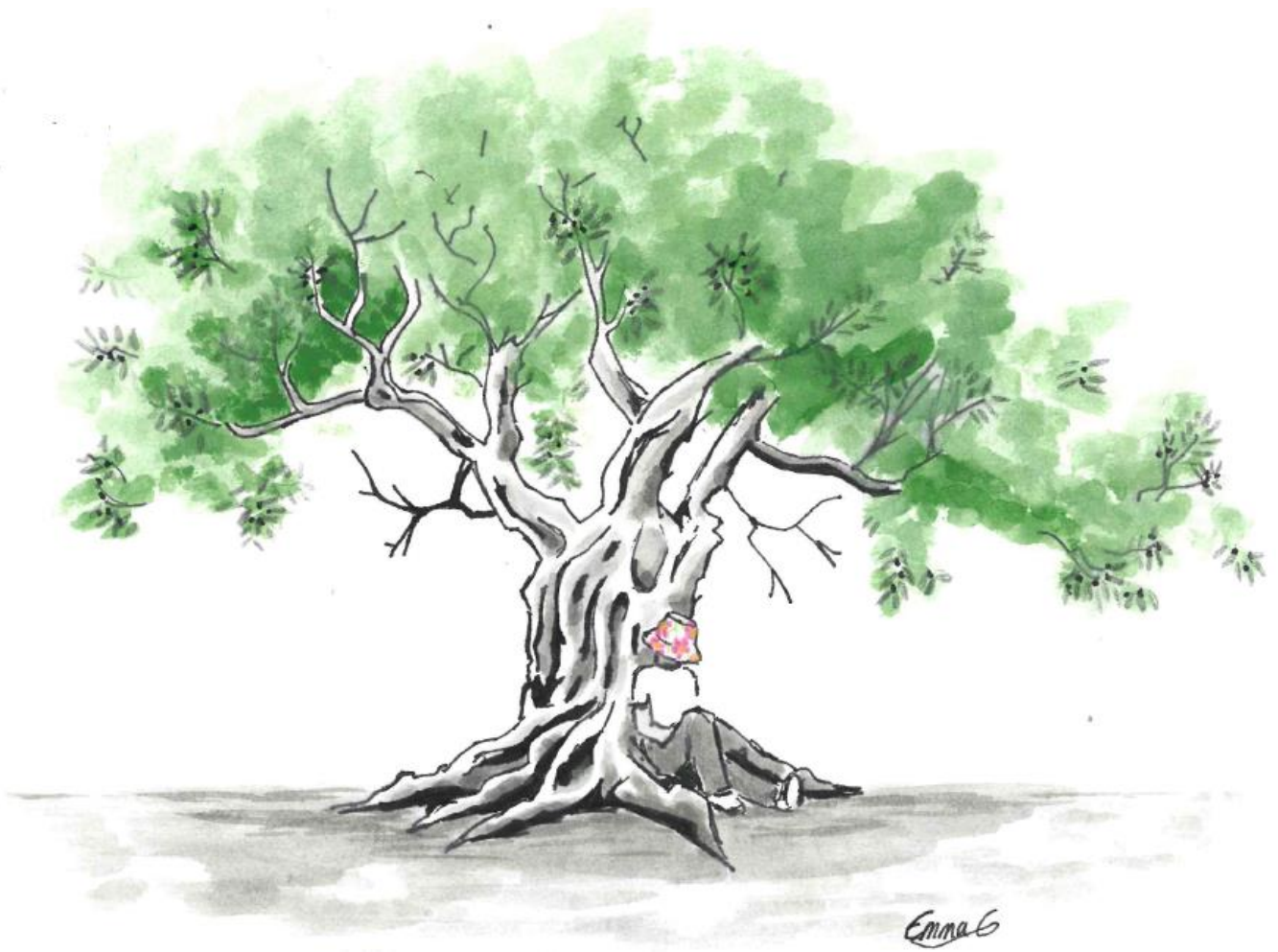
Les travaux conduits au travers de cette thèse permettent de mettre en lumière les mécanismes d'adaptation d'une espèce pérenne à son environnement et au sein d'environnements très différents. L'étude de l'olivier *Olea europaea* L. au sein de l'ouest du BM souligne une histoire évolutive complexe, scindée en différent compartiment génétique, comportant chacun une trajectoire évolutive différente. L'utilisation du séquençage ciblé sur plusieurs centaines d'oliviers le long d'un gradient

RÉSUMÉ SUBSTANTIEL

climatique nord/sud dans l'ouest du BM nous a donné des éléments de réponse sur cette histoire évolutive. Au travers du Chapitre 1, nous avons pu accéder à la structure génétique des populations naturelles d'oliviers sauvages au sein de l'ouest du BM. Nous avons ainsi pu confirmer des résultats d'études précédentes, mais aussi mettre en avant la présence de l'olivier sauvage en France continentale pour la première fois. Nous avons également trouvé de nombreuses populations admixées, résultant d'événements d'hybridation entre les oliviers sauvages et les oliviers cultivés, dans la même zone de répartition que les populations sauvages, posant des questions sur leur prépondérance dans l'environnement naturel. Cela soulève des questions sur leur adaptation, en parallèle de l'investigation de l'adaptation locale chez les sauvages. Avec le Chapitre 2, nous avons pu mettre en évidence l'existence de trois histoires évolutives différentes au sein des oliviers du BM, en lien avec les trois groupes observés (sauvage, cultivé et admixé). Malgré une différence importante, ces histoires ont évolué de façon interconnectée et posent la question de l'importance de ces connexions dans le temps. Des analyses supplémentaires de modélisation d'histoire démographique ainsi que des analyses d'introggression permettraient de répondre à cela. Le Chapitre 3 a permis de mettre en évidence l'adaptation locale de l'olivier sauvage à l'environnement, et plus particulièrement en lien avec les variables de températures. En revanche, nos résultats n'ont pas fourni de preuves d'une adaptation locale pour les oliviers admixés trouvés dans les mêmes habitats. Cependant, les individus admixés conservent 80 % des variants détectés comme adaptés localement chez les sauvages. Ils apportent également une nouvelle diversité génétique issue des cultivés dans les populations naturelles. L'augmentation de la variation génétique latente due à l'hybridation pourrait potentiellement faciliter une réponse plus rapide à l'adaptation de ces populations naturelles face aux changements globaux.

Les résultats de cette thèse vont permettre d'établir la base pour de nouvelles études sur une espèce pérenne, emblématique d'un milieu très fortement impacté par le changement climatique, que ce soit au travers d'étude d'introggression, de l'étude de son l'histoire démographique ou de prédiction d'adaptation face au changement climatique. Les résultats obtenus ici permettront d'adapter les préconisations actuelles de conservation des populations sauvages et admixées. De plus, ils permettront d'accompagner l'évolution des pratiques oléicoles, via l'utilisation de variétés potentiellement mieux adaptées au climat local, limitant ainsi les besoins en irrigation, dans un contexte de raréfaction de l'eau, ressource nécessaire à la vie.

Introduction



1 | Global change, past and current

1.1 | Global

Global change and anthropization are causing rapid changes of ecosystems. These phenomena occurred for millions of years, with consecutive warming or cooling phases. For instance, several events of glaciation occurred during the Miocene, Pliocene and Pleistocene (23 Ma to -11,700 years ago) causing important shifts or extinctions of species (Benton & Twitchett 2003). The climatic changes led to deep modifications of environments through temperature changes, cooler or warmer, precipitation regimes, CO₂ concentrations, or airstream with strong local variations, changing habitat conditions of living organisms. The pace of these changes can vary significantly, ranging from very rapid shifts to processes that unfold over extended periods, for several thousand years (Figure 1; Scott & Lindsey (2020)). Through pollen studies, analysis of fossil records, and monitoring changes in CO₂ concentration, ample evidence has revealed substantial alterations occurring across space and time on Earth, exerting significant impacts on organisms. The impact on plants and trees of past climatic evolution is well documented, where they persisted in several climate changes (temperature, precipitation, CO₂ concentration), with changes occurring rapidly in a few decades or very slowly in thousands of years (Davis & Shaw 2001).

INTRODUCTION

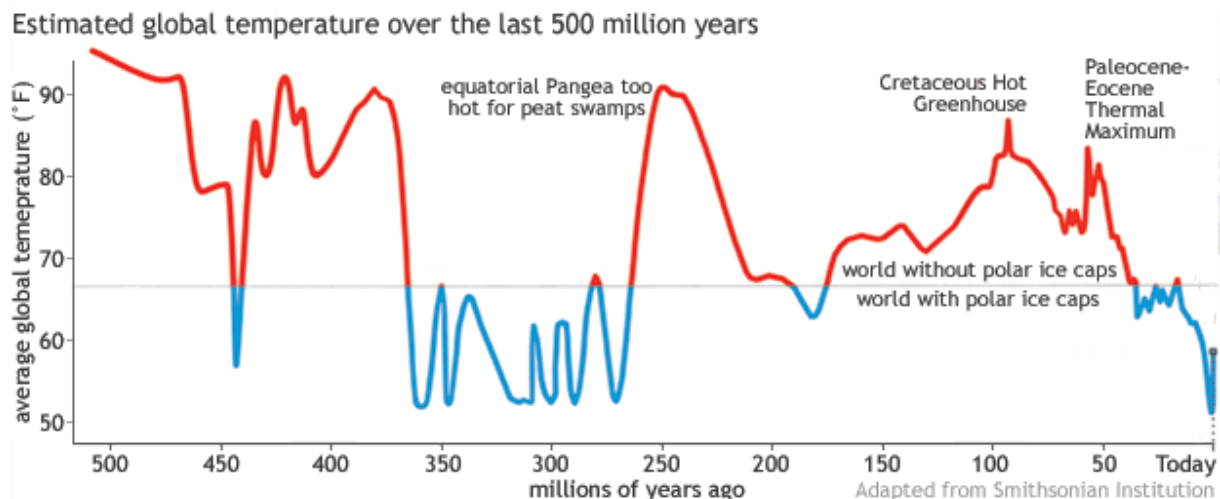


Figure 1. Figure from Scott & Lindsey (2020). Preliminary results from a Smithsonian Institution project led by Scott Wing and Paul Huber, showing Earth's average surface temperature over the past 500 million years. For most of the time, global temperatures appear to have been too warm (red portions of line) for persistent polar ice caps. The most recent 50 million years are an exception. Image adapted from Smithsonian National Museum of Natural History.

Since the 1800s, with the industrial period, human activities largely impact the current climate. The greenhouse gases (CO_2 , methane, etc.) emissions induced by anthropogenic activities from 1850 to 2022 led to a rise of average global temperature of Earth, around 1.15 to 1.65°C depending on the areas (Calvin *et al.* 2023). This global change is characterised by its high speed, resulting in strong modification of selective pressure on the current environment and organisms, with extreme environmental changes (Calvin *et al.* 2023). In 2023, four scenarios of global change were proposed depending on the evolution of temperature: a rise of 1.5°C, 2°C, 3°C or 4°C. These four scenarios are presented in Figure 2, from the Intergovernmental panel on climate change (IPCC) in March 2023 (Calvin *et al.* 2023). In all scenarios, the increase in the global average temperature is impacting the maintenance of organisms in their habitat to a greater or lesser extent, and may even threaten their survival as a species. For instance, it is estimated that under a scenario with a temperature increase of over 4°C, approximately 50% of marine tropical species will become extinct (Figure 3; Calvin *et al.* (2023)).

With every increment of global warming, regional changes in mean climate and extremes become more widespread and pronounced

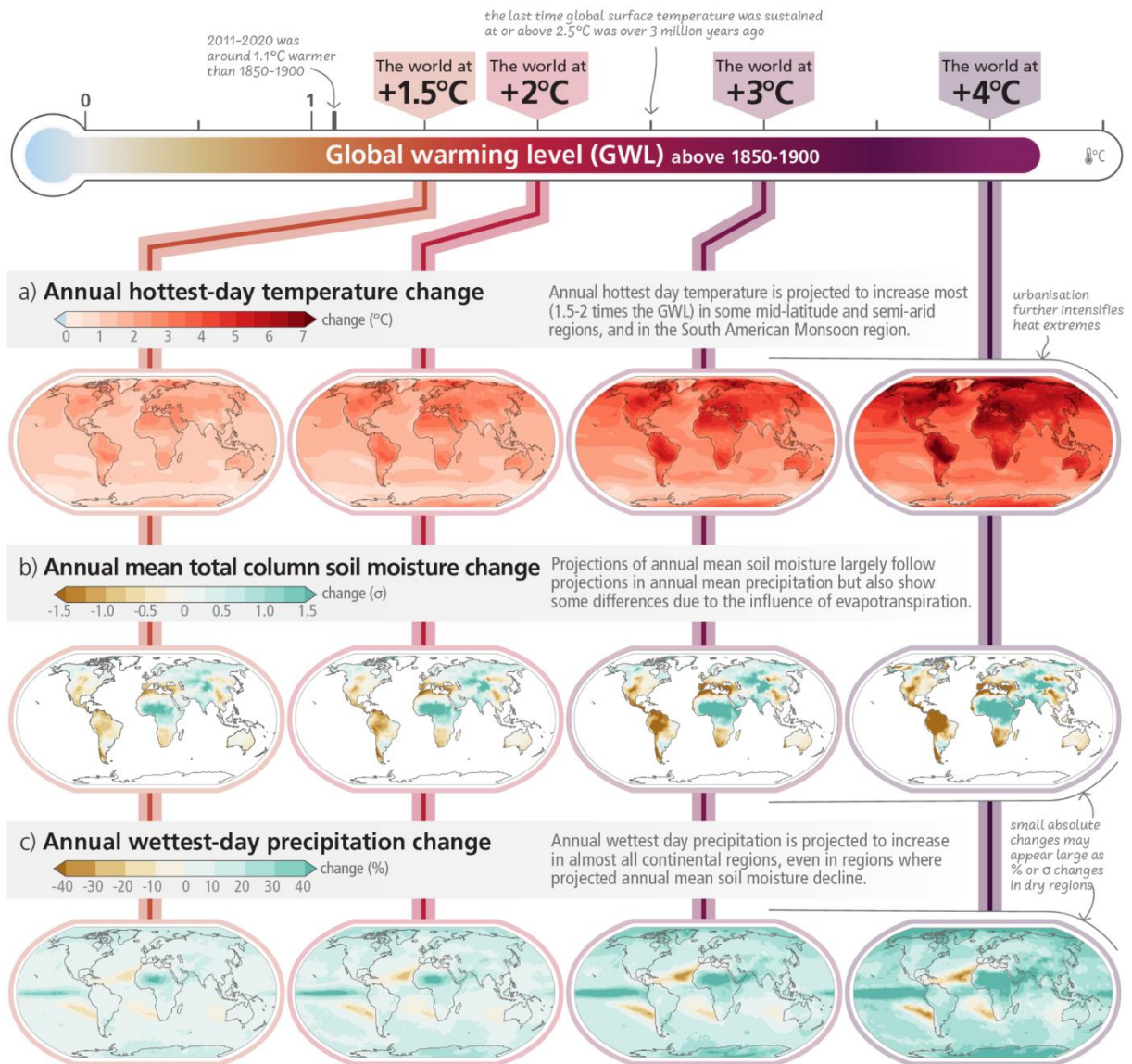


Figure 2. Figure and caption from IPCC report 2023 (Calvin *et al.* 2023). Projected changes of annual maximum daily maximum temperature, annual mean total column soil moisture and annual maximum 1-day precipitation at global warming levels of 1.5°C, 2°C, 3°C, and 4°C relative to 1850–1900. Projected (a) annual maximum daily temperature change (°C), (b) annual mean total column soil moisture (standard deviation), (c) annual maximum 1-day precipitation change (%). The panels show CMIP6 multi-model median changes. In panels (b) and (c), large positive relative changes in dry regions may correspond to small absolute changes. In panel (b), the unit is the standard deviation of interannual variability in soil moisture during 1850–1900. Standard deviation is a widely used metric in characterizing drought severity. A projected reduction in mean soil moisture by one standard deviation corresponds

INTRODUCTION

to soil moisture conditions typical of droughts that occurred about once every six years during 1850–1900. The WGI Interactive Atlas (<https://interactive-atlas.ipcc.ch/>) can be used to explore additional changes in the climate system across the range of global warming levels presented in this figure (Calvin *et al.* 2023).

Future climate change is projected to increase the severity of impacts across natural and human systems and will increase regional differences

Examples of impacts without additional adaptation

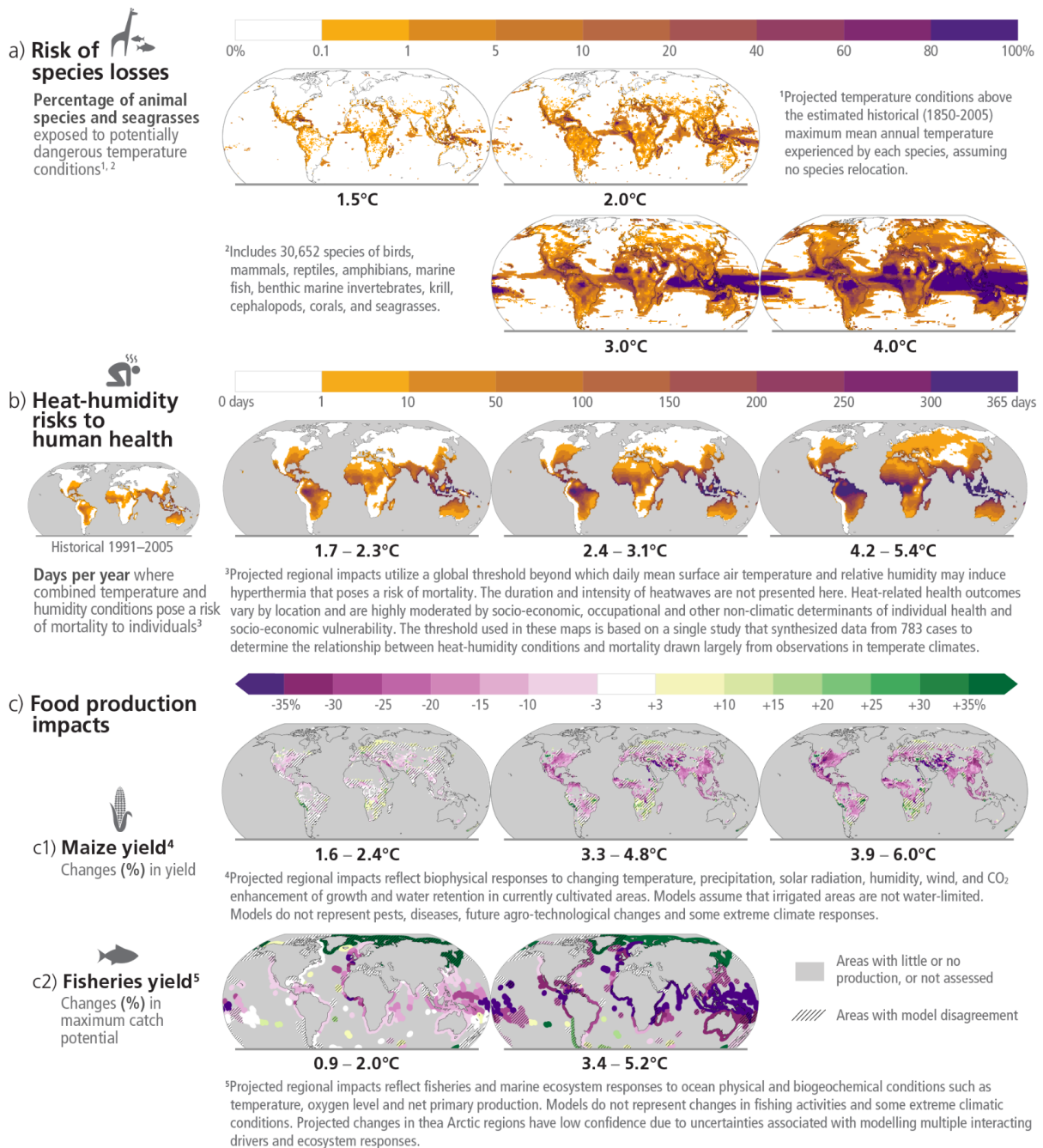


Figure 3. Full figure and caption from IPCC report (Calvin *et al.* 2023). Projected risks and impacts of climate change on natural and human systems at different global warming levels (GWLs) relative to 1850-1900 levels. Projected risks and impacts shown on the maps are based on outputs from different subsets of Earth system and impact models that were used to project each impact indicator without additional adaptation. WGII provides further assessment of the impacts on human and natural systems using these projections and additional lines of evidence. (a) Risks of species losses as indicated by the percentage of assessed species

INTRODUCTION

exposed to potentially dangerous temperature conditions, as defined by conditions beyond the estimated historical (1850-2005) maximum mean annual temperature experienced by each species, at GWLs of 1.5oC, 2oC,3oC and 4oC. Underpinning projections of temperature are from 21 Earth system models and do not consider extreme events impacting ecosystems such as the Arctic. (b) Risks to human health as indicated by the days per year of population exposure to hyperthermic conditions that pose a risk of mortality from surface air temperature and humidity conditions for historical period (1991-2005) and at GWLs of 1.7°C–2.3°C (mean = 1.9°C; 13 climate models), 2.4°C–3.1°C (2.7°C; 16 climate models) and 4.2°C–5.4°C (4.7°C; 15 climate models). Interquartile ranges of GWLs by 2081–2100 under RCP2.6, RCP4.5 and RCP8.5. The presented index is consistent with common features found in many indices included within WGI and WGII assessments (c) Impacts on food production: (c1) Changes in maize yield by 2080–2099 relative to 1986–2005 at projected GWLs of 1.6°C–2.4oC (2.0°C), 3.3°C–4.8oC (4.1°C) and 3.9°C–6.0oC (4.9°C). Median yield changes from an ensemble of 12 crop models, each driven by bias-adjusted outputs from 5 Earth system models, from the Agricultural Model Intercomparison and Improvement Project (AgMIP) and the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP). Maps depict 2080–2099 compared to 1986–2005 for current growing regions (>10 ha), with the corresponding range of future global warming levels shown under SSP1-2.6, SSP3-7.0 and SSP5-8.5, respectively. Hatching indicates areas where { < }70% of the climate-crop model combinations agree on the sign of impact. (c2) Change in maximum fisheries catch potential by 2081–2099 relative to 1986–2005 at projected GWLs of 0.9°C–2.0°C (1.5°C) and 3.4°C–5.2°C (4.3°C). GWLs by 2081–2100 under RCP2.6 and RCP8.5. Hatching indicates where the two climate-fisheries models disagree in the direction of change. Large relative changes in low yielding regions may correspond to small absolute changes. Biodiversity and fisheries in Antarctica were not analysed due to data limitations. Food security is also affected by crop and fishery failures not presented here.

1.2 | The case of the Mediterranean Basin

Nowadays, the Mediterranean Basin (MB) is one of the places most impacted by the ongoing global change, with an increase of average temperature 20% faster than the world average. It is considered as a global change hotspot (Lionello & Scarascia 2018). The current and recent climate of the MB is very distinctive with a temperate climate, hot and dry summers, and mild and wet winters. Within the MB, there are a wide range of climatic variations, depending on the regions, in relation to the heterogeneous topography and environments. This region is considered as a hotspot of biodiversity, containing an average of 10% of the world's plants, including numerous endemic species (Medail & Quezel 1997). The last decades were recorded as warmer and dryer during winters (Lionello *et al.* 2006, 2012). In the future, an increase in daily and annual temperature range and a reduction in precipitation in the basin are predicted, which will have an impact on the environment. Even a change of average 2°C of global warming scenario will likely produce a critical environmental situation related to warming of land areas in summer and widespread reduction of precipitation, differentially across the MB (Lionello & Scarascia 2018). This change in the environment could lead to a critical change in distribution of species, with potential extinction as well.

Changes in environments, whether rapid or gradual, compel species to respond in order to survive. Nowadays, global change at a regional and local scale is challenging for organisms, leading to numerous selective pressures. Unlike what happened in the past, the current global change is occurring very rapidly because of increased anthropogenic activities. Today, studies face challenges in comprehending the diversity of responses of biodiversity to climate change, to anticipate whether species can adapt quickly enough, and understand the life-history traits roles in adapting to environmental changes (Nogués-Bravo *et al.* 2018). Our assumption is that knowledge and comprehension of past climatic events and their impact on species evolution may help mitigate the effect of current changes. Several previous studies have addressed these major events and have contributed to increased scientific knowledge (Steane *et al.* 2014; Jordan *et al.* 2017). In order to understand how plant species, respond to environmental challenges due to global changes, it is important to ascertain the relative influence of the different processes that have shaped the within-species structure of traits during past climate changes.

2 | How do environmental factors and their fluctuations impact species populations?

2.1 | Species response mechanisms to cope with changing environment

The fate of a species, animal or plant, under evolving selective pressure can be achieved in four different ways: extirpation, migration, phenotypic plasticity and adaptation (Aitken *et al.* 2008; Kelly 2019). In the case of extirpation, the population fails to compete, leading to a forced disappearance (Aitken *et al.* 2008). The second strategy is the migration of individuals to track ecological niches in the surroundings. This evolutionary strategy is more easily used by species that are able to move than by fixed species, such as plants. But these latter can migrate by slowly moving with their offspring. This strategy can provide a rapid and easy response to global change. The third strategy to cope with environmental changes is phenotypic plasticity (Kelly 2019), i.e. the ability of a given genotype to express different phenotypes according to different environmental conditions. If the genotype global fitness increases, this mechanism can be considered as adaptive. The last possibility is the adaptation, arising when populations experience contrasted environmental conditions that are spatially distributed. In this case, adaptive traits are selected thus increasing the fitness of individuals in their local environment, resulting in local adaptation (Kawecki & Ebert 2004; Hoban *et al.* 2016) (Figure 4). All of these mechanisms were found during past climate change (Figure 5), for example the migration then extinction of thermophilus plants during Miocene global cooling (Erdei *et al.* 2007), or the extinction of 52% of cool-temperate European trees because of glaciation (Svenning 2003) or again significant plant migrations on several hundred to thousand kilometres because of warming temperature of 5°C (Davis & Shaw 2001; Nogués-Bravo *et al.* 2018).

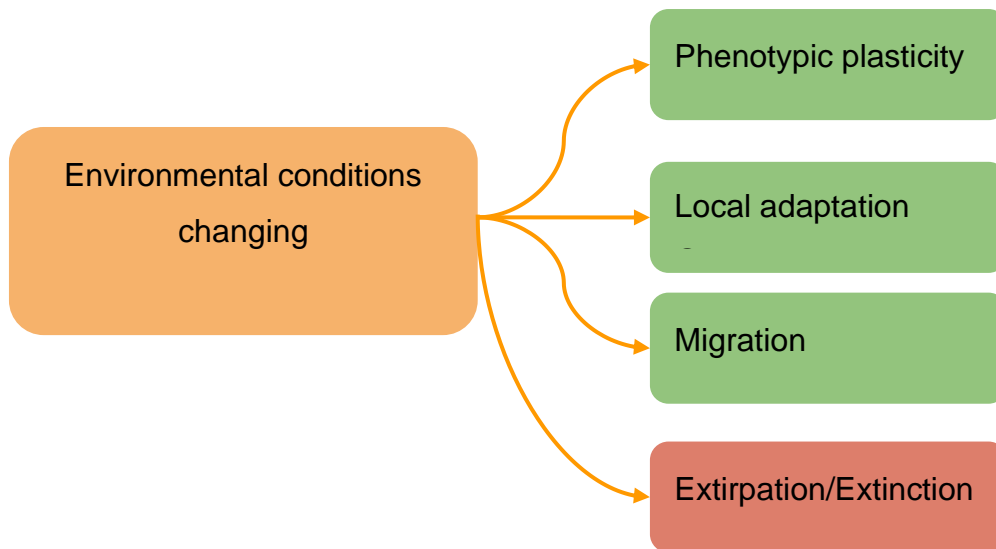


Figure 4. Species response mechanisms to cope with changing environment conditions.

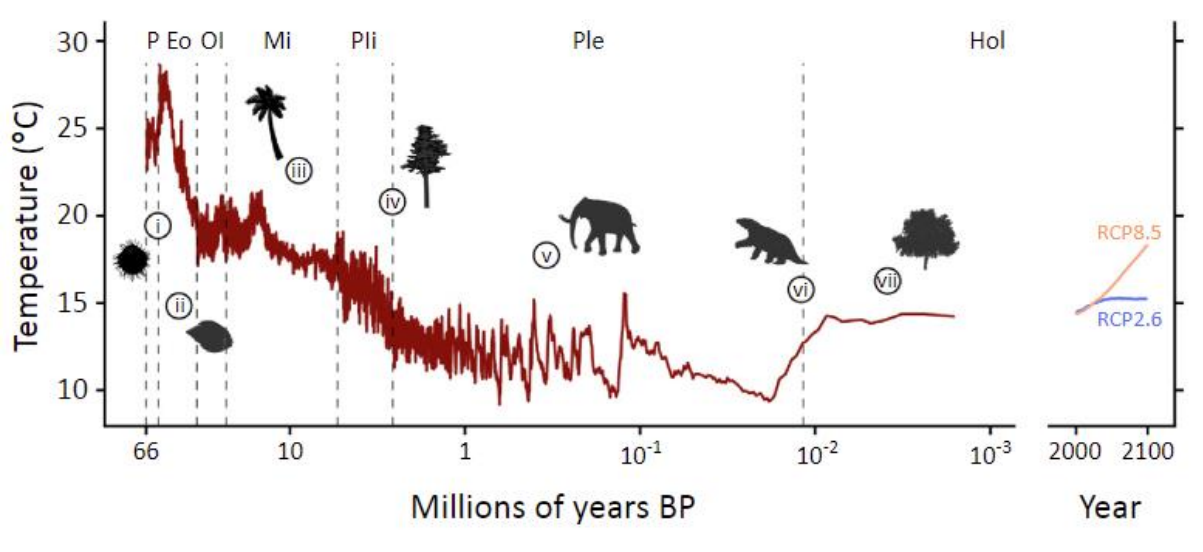


Figure 5. Example of biodiversity response to past climate change (direct or indirect response) along the Cenozoic (past 66 million years). Figure from Nogués-Bravo *et al.* (2018). The case of plants and trees is represented with (iii) (iv) and (vii). The first one (iii) represents the shifts southward followed by extinction of thermophilous plants during the late Miocene global cooling (Erdei *et al.* 2007). The second one (iv) represents the extinction of 52% of cool-temperate European trees because of glaciation cycles at the end of Pliocene (Svenning 2003). The last one (vii) represents the migration of North America plants, in a range of 450 to 2200 km in less than 1000 years because of warmings of 5°C (Davis & Shaw 2001).

INTRODUCTION

Each of these response mechanisms is limited according to species biology. The effectiveness of these mechanisms in response to the environment can be more or less beneficial depending on different organisms. One of the fastest response, especially in the case of significant events, is migration to a more suitable environment. However, this option is only feasible under specific conditions, such as the absence of a fixed life form or one with a short generation time. In the case of trees, migration takes a long time, and it relies on the presence of a habitable environment within the range of the organism's dispersal capacity (Davis & Shaw 2001; Aitken *et al.* 2008). The second fastest response mechanism to major environmental change is plasticity. The expression of a new phenotype with improved fitness in the environment allows better survival. This mechanism depends on the ability to express phenotypic traits or behaviour with higher fitness in a given environment. This mechanism is found in many species, including trees, and it could contribute to the potential for their habitat expansion and reduce the risk of extinction (Hampe & Petit 2005; Petit & Hampe 2006). The last possibility is adaptation. This slower mechanism is the result of species evolving within their environments, leading to a sustainable and effective response to that environment with improved fitness. Local adaptation in multiple populations of the same species allows for the emergence of different genetic variants adapted to these environments, and the sharing of these variants among populations can occur through gene flow. For example, the case of the distribution of *Pinus sylvestris* and *Pinus pinaster* was modelised to predict distribution species under global warming in the MB. They indicate for these species that populations from southern Spain would have higher relative survival in northern territories, that would suggest a potential pre-adaptation of their populations to warmer climate, in this case, locally adapted provenances to warmer conditions would benefit from climate change at the expense of other provenances. On the other hand, populations from northern Spain would probably disappear, their phenotype would not be adapted (Benito Garzón *et al.* 2011). In plants, a meta-analysis of adaptation studies has shown that nearly 45% of them are locally adapted to their environment (Leimu & Fischer 2008).

2.2 | Local adaptation: genomic variations between populations

Local adaptation is an important evolutionary mechanism that occurs when individuals exhibit higher average fitness in their local environment compared to other environments, and they display better fitness in contrast to individuals from different locations. The local conditions, which change over time and space, determine which traits are favoured by natural selection (Savolainen *et al.* 2013). It results from the evolution of individuals in response to the change of variable selection. To determine if a population is locally adapted to its environment, fitness measurement can be employed: a population, locally adapted to its habitat, will have to fulfil two conditions: i) having a higher fitness in this habitat than in another one, i.e. the “home versus away” criterion and ii) having a higher fitness in its habitat than another population moving into it, i.e. the “local versus foreign” criterion as defined by (Kawecki & Ebert 2004) (Figure 6; Savolainen *et al.* (2013); Vanwallendael *et al.* (2019)).

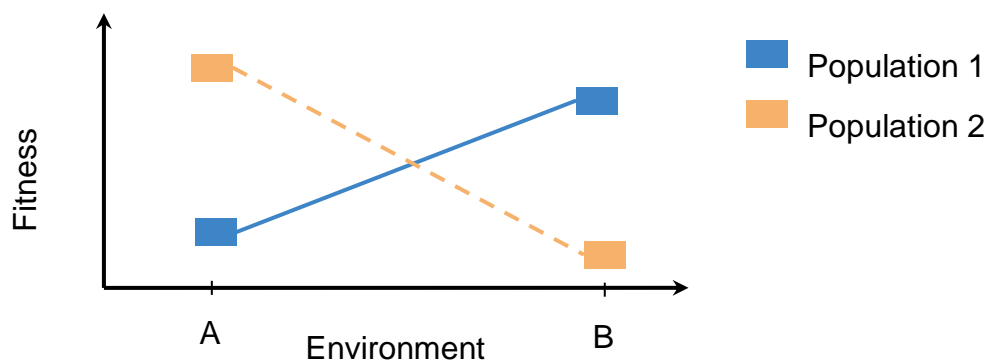


Figure 6. Representation of the evolutionary mechanism of local adaptation to two environments based on figure from Vanwallendael *et al.* (2019). The filled square corresponds to population fitness means.

For local adaptation to occur, genomic variation must take place within the genome, driven by evolutionary forces. Evolutionary forces are defined in four types: mutations, selection (natural and artificial), genetic drift, and gene flow/migration. Mutation is a highly important phenomenon in evolution. This force allows for the emergence of new variants within populations, leading to modifications within the genome through deletions, duplications, transpositions, nucleotide variations, or other types of changes. These genomic variations can be beneficial, neutral, or deleterious,

INTRODUCTION

resulting in variations between individuals and populations, and they can be maintained based on the selective forces acting on the genome. The variability arising from these mutations may or may not persist depending on other evolutionary forces.

Migration is an evolutionary force through which populations move from their initial habitat, often on a larger scale than usual, leading to a redistribution in environments (Dingle & Drake 2007). It can also take the form of gene flow through reproduction. Gene flow enables the exchange of genomic variants between populations, and may result in a homogenization among them. This can have both negative or positive effects, either hindering local adaptation through the homogenization of variants or, conversely, ensuring the introduction of new variants necessary for the adaptive potential of the population (Lenormand 2002; Kremer *et al.* 2012). For example, some tree species have long-distance gene flow, which can increase their fitness by increasing their genetic variation and thus help for setting up in new environments (Petit & Hampe 2006; Kremer *et al.* 2012).

Genetic drift is an event defined as random fluctuations in allele frequencies from one generation to another. Conversely, within large populations, the number of draws of alleles is high, and there will be little to no difference in allele frequencies between generation 1 and generation 2 (Serre 2006). This phenomenon help to maintain Hardy-Weinberg equilibrium, and can maintain standing genetic variation, meaning the presence of multiple alleles at a locus within a population, which could enable a quicker response to a changing environment (Barrett & Schluter 2008). While, within small populations, the draws will be less numerous, resulting in a loss of equilibrium. Allelic frequencies will therefore change due to random sampling variations, which can lead to the fixation of random or deleterious alleles or the loss of variants (Serre 2006).

Selection can be divided into two categories: natural and artificial. Natural selection is defined by the fact that some individuals within a population exhibit phenotypic variation and differential reproductive success, corresponding to fitness, in relation to genomic variations within a changing environment. This means that the habitat exerts positive selective pressures, leading to the disappearance of some genotypes in favour of others better adapted. Artificial selection is a selection carried out by humans on plants and animals to choose traits considered beneficial to them, a process known as domestication. This process can mitigate the effects of natural selection and promote genotypes that would not survive in the wild in favour of

agronomic traits. This mechanism can lead to a loss of genetic diversity within cultivated groups compared to their wild relatives (Bourguiba *et al.* 2012).

Exploring local adaptation and how it works provides fundamental knowledge in understanding evolutionary mechanisms. Indeed, when adaptation is local, the evolutionary process may be studied in action, allowing investigation on the ways in which genetic variation may shape the patterns of phenotypic variability (Kawecki & Ebert 2004). Investigating local adaptation, especially by genomic approach, has become an important goal for both crop and animal breeding facing climate change (Howden *et al.* 2007). Knowledge on this evolutionary process is particularly important to better understand impacts on natural perennial populations, predict their long-term behaviour in response to climate change and implement conservation measures (Hoffmann *et al.* 2015).

2.3 | The impact of domestication on natural populations: the case of crop-to-wild gene flow.

As explained earlier, gene flow can lead to the homogenization of genomic variants between populations and the introduction of beneficial or deleterious alleles. Exchanges have been observed between wild and cultivated populations in both directions, crop to wild or wild to crop, in animals and plants. (Hauser & Shim 2007; Johnsson *et al.* 2016; Le Corre *et al.* 2020; Anderson *et al.* 2021; Mary *et al.* 2022). These introgressions of new genetic diversity inside wild genomes can accelerate their adaptation to new environment by increasing the frequency of favourable alleles, but can also be negative, resulting in outbreeding depression, i.e. a loss of fitness in hybrids compared to their parents (Ellstrand *et al.* 1999; Keller *et al.* 2000; Gering *et al.* 2019). Gene flow has been observed several times in perennial species such as chestnut, poplar, or apple (Chenault *et al.* 2011; Jelić *et al.* 2015; Feurtey *et al.* 2017; Cornille *et al.* 2019). But its evolutionary consequences in the natural environment and in wild populations are still misunderstood, especially in perennial species, because of fitness measurement can be complicated with these species. Apple tree studies quantified fitness within wild groups impacted by introgression from the cultivated, and showed a higher fitness of admixed individuals compared to wild ones (Feurtey *et al.* 2017; Cornille *et al.* 2019). Nevertheless, this result is only one example of benefit from crop-to-wild gene flow but several other studies are required. The potential impact

INTRODUCTION

from cultivated via gene flow must be taken into account when considering local adaptation.

3 | How to study local adaptation?

Most methods to study local adaptation of plants are based on quantification of fitness or phenotypes in common garden or reciprocal transplant experiment (Fournier-Level *et al.* 2011; Brachi *et al.* 2013; Capblancq *et al.* 2023) or on the assessment of the genomic variation of phenotypic traits among natural populations (Savolainen *et al.* 2013; Jordan *et al.* 2017; Leroy *et al.* 2020). Based on phenotypic trait analyses, these methods are considered as direct approaches. This kind of study, using common garden or reciprocal transplant experiments to study local adaptation, was the only way, for decades, to understand the impact of natural selection on the phenotypic variation (Sork 2017). Recently, with the advent of next-generation sequencing, other methods, based exclusively on genome variation analysis, became available. They rely on genotyping individuals from natural heterogeneous environments without directly investigating phenotypic traits (Savolainen *et al.* 2013; Steane *et al.* 2014; Whitlock & Lotterhos 2015; Capblancq *et al.* 2018). These new, indirect methods, allowed to investigate the association between environmental gradients and genomic variation of population (Savolainen *et al.* 2013; Sork *et al.* 2013; Tiffin & Ross-Ibarra 2014) and to avoid lengthy common garden or reciprocal transplant, especially for long-lived plants such as trees (Sork *et al.* 2013; Jordan *et al.* 2017)

3.1 | Direct approach: reciprocal transplant experiment and/or common gardens

Direct methods involve the use of a common garden experiment or reciprocal transplantation approach, that includes at least two populations of plants in a minimum of one to several different environments. In order to detect local adaptation, individuals from different environments are introduced into new environments alongside the existing population of the same species. Fitness measurements are then taken to determine whether local individuals exhibit a higher fitness compared to foreign individuals (Blanquart *et al.* 2013; Lascoux *et al.* 2016). Local adaptation is evidenced when individuals respect two criteria: i) they have better fitness in their local environment than individuals from the other environment and ii) they have better fitness in their own environment (Kawecki & Ebert 2004; Figure 7). Experiments on reciprocal transplantation have been mainly used for short-living plant species because the adaptive phenotypic traits are easy to measure (for instance in *Arabidopsis thaliana*; Fournier-Level *et al.* 2011; Brachi *et al.* 2013). Despite long periods of experiments, phenotypic traits are commonly measured in one or multiple common gardens in forest trees (Sork *et al.* 2013). However, in situations where establishing common gardens or reciprocal transplant experiments is challenging due to species-specific biological factors, limited access to diverse plant populations, or other logistical constraints.

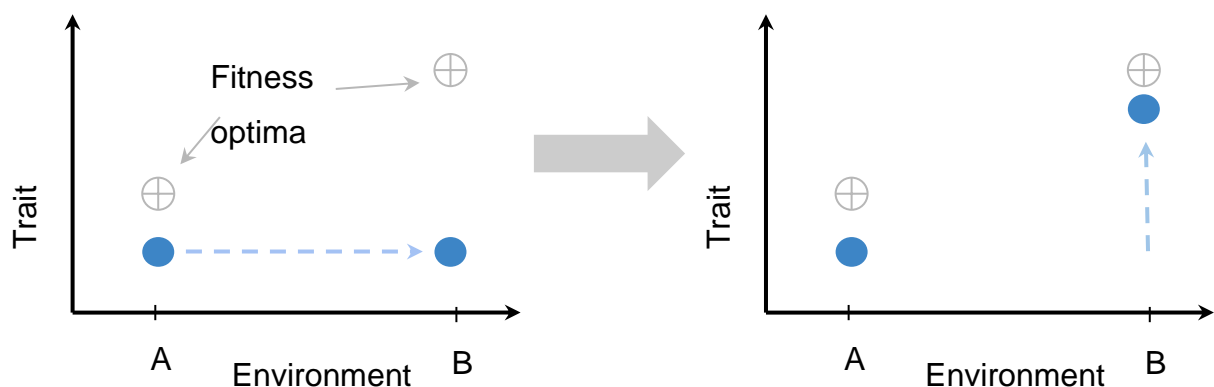


Figure 7. Representation of the evolutionary mechanism of local adaptation to two environments based on figure from (Vanwallendael *et al.* 2019). The filled circle corresponds to population trait means. The target points are the optimum trait values for particular environment (A or B)

3.2 | Indirect approach: Genomic studies of local adaptation

When establishing a common garden or reciprocal transplant experiment is challenging, the use of indirect approaches is possible. These methods have emerged with the availability of population genome scans in natural environments. Indirect approaches rely on genomic variation analyses without knowledge on the phenotypes. Indeed, local adaptation results from spatially variable selective pressure which can affect part of the genome (Tiffin & Ross-Ibarra 2014). Fitness effects may leave a strong pattern on the genome, on an evolutionary time scale. Several methods are used to identify specific genes and alleles that may be responsible for local adaptation. There are various approaches available, here we will present two: detection of selection footprints and genomic environment association analyses (GEA).

The first approach aimed to detect genomic regions under selection by identification of selective sweep (Hoban *et al.* 2016). It relies on the assumption that most of the genome is evolving neutrally with only a minor subset of loci evolving under positive selection. A selective sweep arises when a mutation is selected that reduces variability in linked sites as its frequency increases toward its fixation. The local intraspecific genomic variability would decrease, and the frequency spectrum locally shifts toward low frequency variants. Calculation of frequency spectrum, variability and/or linkage disequilibrium allows to determine selective sweep (Nielsen 2005). This approach is mostly used to confirm signals detected by F_{ST} -based methods which is commonly used to detect outliers (Whitlock & Lotterhos 2015; Hoban *et al.* 2016). Loci with significantly high F_{ST} compared to the rest of the genome are assumed to be under positive selection and are considered as outliers, based on the genetic differentiation of loci between populations (Hoban *et al.* 2016). The method issue is that the detection of part of the genome being under positive selection indicates biotic or abiotic pressures on individuals, but not the origin of this selection.

The second potential approach is the genome environment association (GEA), which takes into account the correlations between allelic variations and the environment, when a loci is significantly associated to an environment variable, it is considered as potentially selected by this variable (De Mita *et al.* 2013; Steane *et al.* 2014; Jordan *et al.* 2017; Rellstab *et al.* 2017). This approach needs access to multiple populations, as well as environmental data from each population site (Steane *et al.* 2014; Hoban *et al.* 2016).

The limit of both approaches is the difficulty to distinguish the genomic association to environment from the genomic variations caused by demographic history (Whitlock & Lotterhos 2015; Frichot & François 2015; Hoban *et al.* 2016). Incorporating a correction for confounding factors, such as a structure matrix derived from genomic data minimally influenced by linkage disequilibrium, can assist in refining this model and reducing the occurrence of false positives (Lotterhos & Whitlock 2015).

Nowadays, the increase of these techniques has facilitated the study of natural populations, especially tree populations. The extended generation times of trees make it difficult to investigate local adaptation through common garden or reciprocal transplant experiments. This long generation time also complicates the fast adaptive response to a changing environment with the global change and threatened trees (Aitken *et al.* 2008). Sork *et al.* (2013) emphasised the significance of identifying species or populations at risk using studies on local adaptation to adapt forest conservation measures.

To investigate the impact of climate change on plants, and more particularly on perennial species, this PhD will focus on a tree species of key importance in the MB, *Olea europaea* subsp. *europaea* L., to better understand local adaptation and its footprints on its genome.

4 | A model of study: wild olive tree

4.1 | Biology of *Olea europaea* L.

Olive tree, *Olea europaea* subsp. *europaea* L. is a perennial species emblematic of the MB, and more particularly of the meso-Mediterranean climate alongside the evergreen oak (Ozenda 1975; Rivas Martínez *et al.* 1987). Olive tree can live more than thousands of years. The olive tree is pollinated by wind (anemophilous) and seeds are disseminated by birds (zoochoria), it is an allogamous species with a self-incompatibility system (Spennemann & Richard Allen 2000; Saumitou-Laprade *et al.* 2017). Because pollination and seed dissemination may occur over a long geographic distance (more than 50 km; Spennemann & Richard Allen 2000), an extensive gene flow is expected, causing connection between wild populations as a metapopulation (Besnard *et al.* 2018). Its presence around the MB can be explained by its specific temperature requirements for flowering. A first stage

INTRODUCTION

requires a minimal accumulation of low temperatures to complete the chilling requirement (a minimum of 7.2°C), then a period of heat accumulation is needed allowing budbreak and bud growth (Rallo & Martin 1991; Pérez-López *et al.* 2008; Haberman *et al.* 2017). The lower threshold temperature for olive fruit growth is around 15°C to 26°C (Pérez-López *et al.* 2008).

This tree is diploid ($2n = 46$) (Green & Wickens 1989) and can hybridise with other olive subspecies such as *O. europaea* subsp. *cuspidata* (Besnard *et al.* 2007, 2008), or *O. europaea* subsp. *laperrinei* (Besnard *et al.* 2008). Its genome size is approximately 1.38GB for the Farga reference genome (Cruz *et al.* 2016; Julca *et al.* 2020). Its linkage disequilibrium is low with less than 250 bp for r^2 to be under 0.2 (D'Agostino *et al.* 2018; Zhu *et al.* 2019).

According to botanical knowledge, *Olea europaea* subsp. *europaea* is composed of two botanical varieties in the MB, *Olea europaea* subsp. *europaea* var. *europaea* and var. *sylvestris* which are the cultivated form and the wild form (or oleaster) respectively (Green 2002). Wild and cultivated species live in the same distribution area in the MB (Zohary & Spiegel-roy 1975; Rubio de Casas *et al.* 2006). It is very complicated to distinguish wild from cultivated using morphology measured alone, the shape stone allows to distinguish the wild (round-shaped stone) from cultivated. However, numerous intermediate morphotypes have been found, ranging from elliptical to more tapered and were complicated to associate to wild or cultivated (Terral *et al.* 2004, 2021). Only DNA markers can provide certainty (Lumaret & Ouazzani 2001; Besnard *et al.* 2013b; Diez *et al.* 2015; Khadari & El Bakkali 2018; Gros-Balthazard *et al.* 2019; Julca *et al.* 2020) Additionally, gene flow has been observed between cultivated and wild olive trees in the MB (Gros-Balthazard *et al.* 2019; Julca *et al.* 2020), but little impact was identified in the western MB according to the observed genetic pattern (Besnard *et al.* 2013a; Khadari & El Bakkali 2018; Gros-Balthazard *et al.* 2019).

4.2 | Evolutionary History of *Olea europaea* L.

The studies on wood-coal (Carrión *et al.* 2010), pollen analyses (Planchais 1984) and fossilised olives stone (Suc 1984; Terral *et al.* 2004) allowed to trace the demographic history of *Olea europaea* across the MB. Wild olives have been present in this region for millions of years. The Messinian crisis which occurred between 5.59

to 5.33 million years ago, resulted in a significant colonisation of this species from North Africa to Southern Europe (Terral *et al.* 2009). It resulted in two gene pools of olive trees around the MB: one gene pool in the east, the other one in the west. The western gene pool comprises populations from southern France to North Africa, including Spain, Italy, Morocco and Algeria. The eastern gene pool is located in Syria, Turkey and also includes Greece (Figure 8; Besnard *et al.* 2013b, 2018; Diez *et al.* 2015; Khadari & El Bakkali 2018). Populations in Europe and North Africa were affected by multiple glaciation events during the Holocene, leading to significant extinctions and population shifts. The surviving populations remained in refuge areas (Figure 9 from Carrión *et al.* 2010). At the end of glaciation, favourable environmental conditions made possible the recolonization leading to the extension and potential reconnection between populations from refuges (Planchais 1984; Terral *et al.* 2009). Subsequently, human activities, using olive for heating oil, but not considered as domestication activity at this period, became the primary driver in landscape and vegetation transformation, promoting the expansion of olive trees (Planchais 1984; Terral *et al.* 2009; Carrión *et al.* 2010). This demographic history has led to the current distribution (Figure 10; Carrión *et al.* 2010).

INTRODUCTION

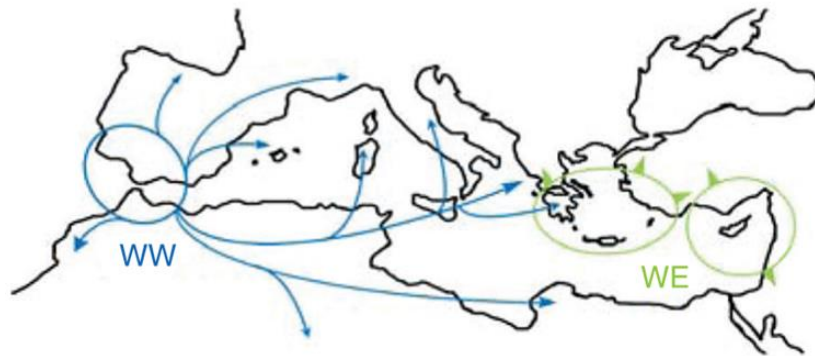


Figure 8. Distribution of *Olea europaea* subsp. *europaea* var. *sylvestris* with western wild (WW) oleaster populations and eastern wild (WE) oleaster populations that expanded extensively during the Holocene. Figure from Besnard et al. (2018). Contact is thought to occur in the central Mediterranean Basin (Besnard et al., 2013, 2018; Diez et al., 2015; Khadari & El Bakkali, 2018).

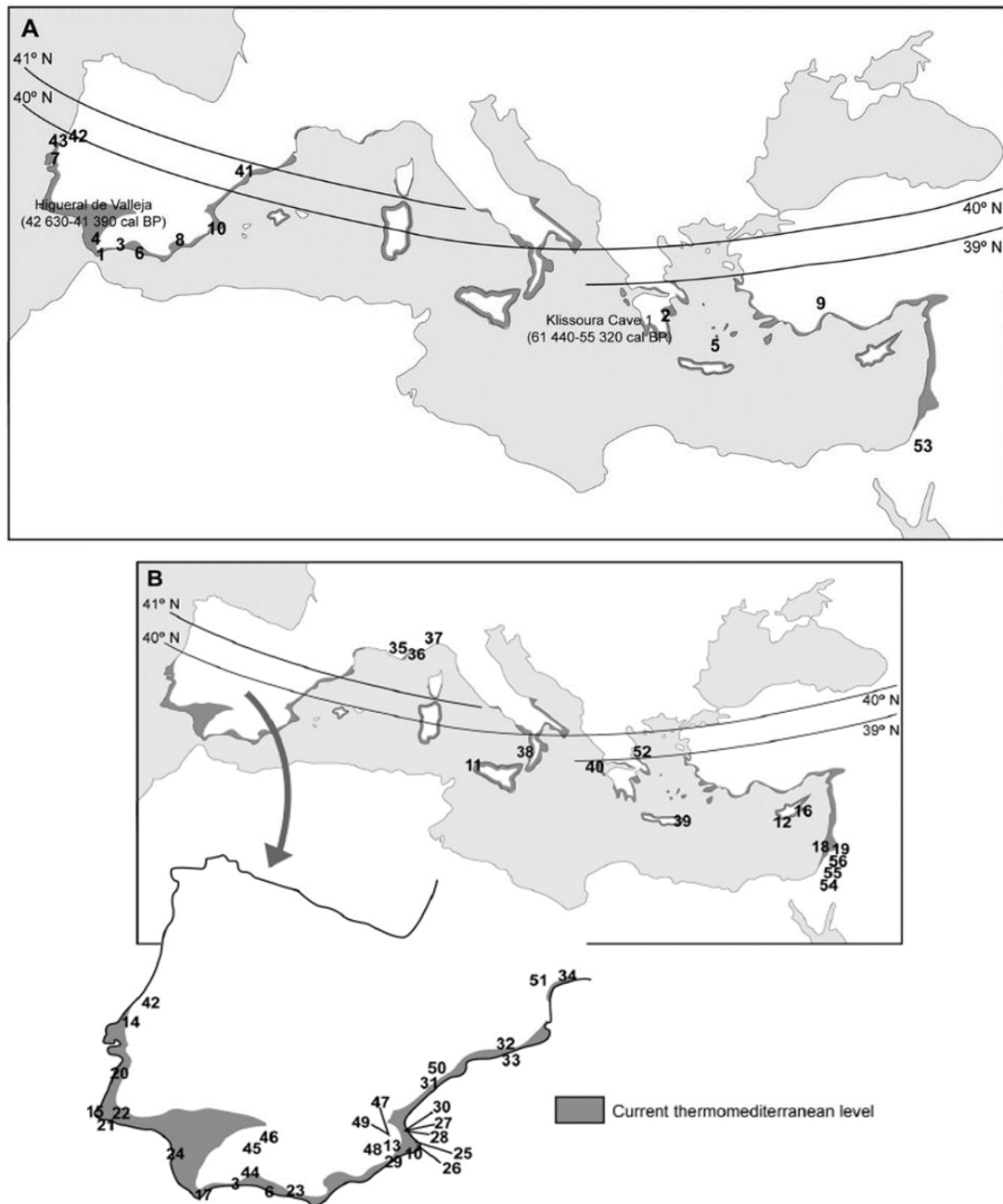


Figure 9. Refuge areas of *Olea europaea* in the Northern Mediterranean Basin during A) the Pleniglacial and B) the Early–Middle Holocene. The presence of *Olea europaea* in these refuges was determined based on the presence of wood-charcoal (Figure from Carrión et al. (2010)).

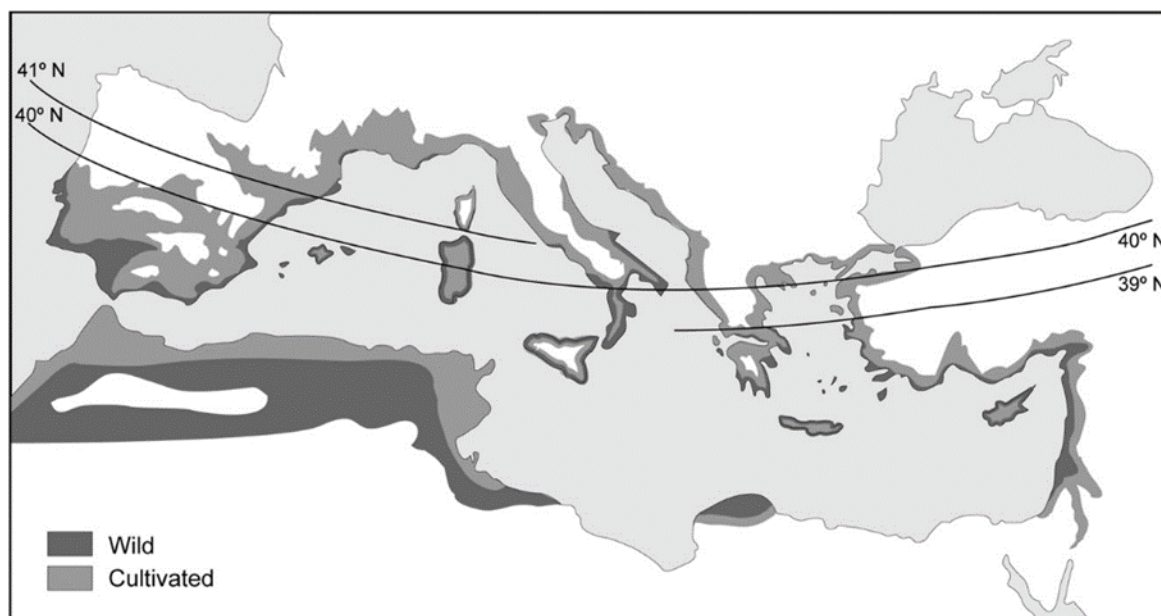


Figure 10. Present-day range of the wild and cultivated olive (*Olea europaea* L.) in the Mediterranean basin (Figure from Carrión et al. (2010)).

4.3 | Domestication history of *Olea europaea* L.

The process of olive tree domestication began approximately 6,000 years ago (Zohary & Spiegel-roy 1975; Kaniewski *et al.* 2012; Newton *et al.* 2014). It is usually considered that the primary centre of olive domestication, originating from wild progenitors, is located in the Middle East, near the border between Turkey and Syria (Zohary & Spiegel-roy 1975; Kaniewski *et al.* 2012; Besnard *et al.* 2013b). However, a question persists regarding whether cultivated olives arose from a single domestication event in the Levant, followed by subsequent diversification, or if it was the result of multiple independent domestication events. Several studies argue in favour of a scenario involving a single primary domestication event in the eastern MB, followed by the spread of the initially domesticated olive trees and their diversification in the central and western MB. These processes are considered as essential in the evolutionary history of olive trees (Besnard *et al.* 2013b, 2018). However, recent genomic research supports the possibility of a second domestication event occurring in the central MB (Diez *et al.* 2015). A recent study by Julca *et al.* (2020) argues in favour of an intermediate scenario, a domestication continuum. It consists of a primary domestication event in the Levant, followed by events of hybridization with the local wild genetic pools throughout the Mediterranean Basin. The arrival of the cultivation

of olive tree in the western MB would originated from at least two major events of introduction, a first one in the southern Iberian Peninsula, 2,500 BP years ago (Langgut *et al.* 2019) and a second one later, in southern France, during the Roman arrival in the first century BC (Leveau *et al.* 1991; Figure 11). The cultivation of the olive in later periods in the northern Iberian peninsula could have caused the olive trees to become more resistant to continental conditions (Carrión *et al.* 2010).

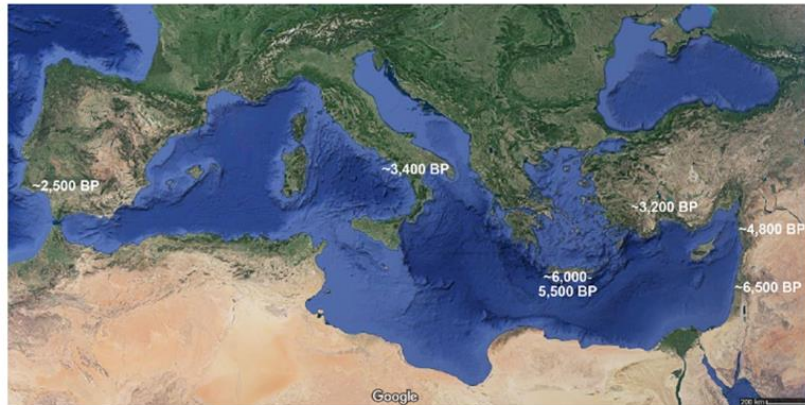


Figure 11. Suggested assumption regarding the onset of olive cultivation in the Mediterranean Basin (figure from Langgut *et al.* (2019)).

Following genetic investigations, the actual distribution of cultivated olive trees is presented in Figure 12 (Diez *et al.* 2015; Besnard *et al.* 2018). They are separated into three groups according to three main genetic clusters, Q1 in the western MB, Q2 in the western and central MB and Q3 in the eastern MB (Diez *et al.* 2015; Besnard *et al.* 2018). The Q1 cultivated varieties are very closely related while in Q2 and Q3 groups, varieties are genetically separated with limited close relation as described in (Figure 13; (Diez *et al.* 2015)). Interestingly, despite the close genetic relationships between wild and cultivated olive trees in the eastern MB, it appears that wild olive populations in the western MB have experienced less or even no introgression by cultivated olives. The genetic structure of this western gene pool is very low (Khadari & El Bakkali 2018). This suggests a more complex and nuanced history of olive tree domestication and dispersion in the MB.

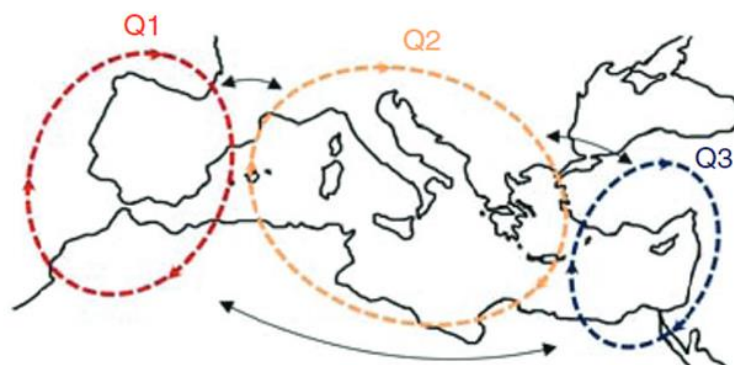


Figure 12. Distribution of cultivated olive trees according to their three main genetic clusters: Q1, Q2 and Q3, figure from Besnard *et al.* (2018) based on Diez *et al.* (2015).

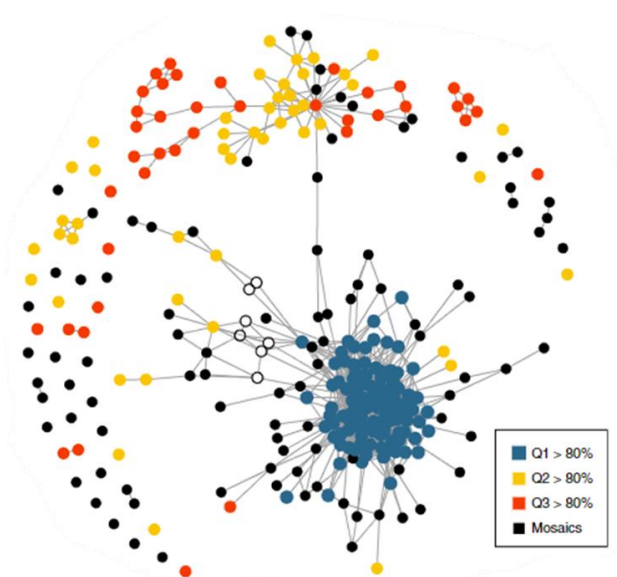


Figure 13. First-degree relative connections between different cultivars olives from different gene pool Q1 (in blue), Q2 (in yellow) and Q3 (in red), figure from Diez *et al.* (2015).

5 | Research project ClimOliveMed and objectives of the PhD

This thesis is proposed as a part of the ClimOliveMed project. This project aims to research the challenges of global change adaptation, biodiversity conservation and sustainable production within the context of Mediterranean olive cultivation. The project is in collaboration between academic institute as well as professionals within the olive sector from diverse countries around the MB. The project is structured into three distinct work packages (WPs). The first WP is primarily concerned with

comprehensive investigations into the olive tree, focusing on various olive varieties and their potential applications in mitigating the adverse impacts of climate change. The second WP is dedicated to the olive tree production system of and more specially about analysing the perception of farmers about the climate change on their production on their agricultural practices. The final WP is involved in facilitating constructive dialogue among all actors within the olive sector, including research institutions and international bodies. The overarching goal of this collaborative effort is to conceive and implement adaptive strategies for the olive cultivation in response to the challenges posed by global change.

This PhD is part of the first WP and focuses on the evolutionary history of populations of wild olive trees in the western Mediterranean Basin. Its primary objective is to evaluate the evidence of local adaptation and selection within these populations. This inquiry is explored through three specific sub-questions:

1. What is the genetic diversity and the genetic structure present in naturally occurring olive tree populations?
2. Which parts of the genome are under selective pressure?
3. Are there genetic loci that are potential associations with environmental variations within these populations?

The PhD is divided in three chapters: 1. Genomic evidence of genuine wild versus admixed olive populations evolving in the same natural environments in western Mediterranean Basin. 2. Selection footprints in *Olea europaea* subsp. *europaea*: Detection in Wild, Admixed and Cultivated olive trees from the Western Mediterranean Basin - Clues to their evolutionary history? 3. Does crop-to-wild gene flow lead to maladaptation of perennial species to their natural environment? Insights from the study of local adaptation of wild olive trees to their environment.

Chapter 1



Genomic evidence of genuine wild versus admixed olive populations evolving in the same natural environments in western Mediterranean Basin

Published – *PLOS One*

<https://doi.org/10.1371/journal.pone.0295043>

Lison Zunino^{1,2}, Philippe Cubry², Gautier Sarah¹, Pierre Mournet^{1,3}, Ahmed El Bakkali⁴, Laila Aqbouch¹, Stéphanie Sidibé-Bocs^{1,3}, Evelyne Costes¹ and Bouchaib Khadari^{1,5*}

Affiliations:

¹AGAP Institut, University of Montpellier, CIRAD, INRAE, Institut Agro, Montpellier, France

²DIADE, University of Montpellier, CIRAD, IRD, Montpellier, France

³CIRAD, UMR AGAP Institut, Montpellier, France

⁴INRA, UR Amélioration des Plantes et Conservation des Ressources Phytogénétiques, Meknes, Morocco

⁵Conservatoire Botanique National Méditerranéen (CBNMed), UMR AGAP Institut, Montpellier, France

*Corresponding author

E-mail: bouchaib.khadari@cirad.fr; b.khadari@cbnmed.fr (BK)

Abstract

Crop-to-wild gene flow is a mechanism process widely documented, both in plants and animals. This can have positive or negative impacts on the evolution of admixed populations in natural environments, yet the phenomenon is still misunderstood in long-lived woody species, contrary to short-lived crops. Wild olive *Olea europaea* L. occurs in the same eco-geographical range as domesticated olive, i.e. the Mediterranean Basin (MB). Moreover, it is an allogamous and anemophilous species whose seeds are disseminated by birds, i.e. factors that drive gene flow between crops and their wild relatives. Here we investigated the genetic structure of western MB wild olive populations in natural environments assuming a homogenous gene pool with limited impact of cultivated alleles, as previously suggested. We used a target sequencing method based on annotated genes from the Farga reference genome to analyze 27 western MB olive tree populations sampled in natural environments in France, Spain and Morocco. We also target sequenced cultivated olive tree accessions from the Worldwide Olive Germplasm Bank of Marrakech and Porquerolles and from an eastern MB wild olive tree population. We combined PCA, sNMF, pairwise F_{ST} and TreeMix and clearly identified genuine wild olive trees throughout their natural distribution range along a north-south gradient including, for the first time, in southern France. However, contrary to our assumption, we highlighted more admixed than genuine wild olive trees. Our results raise questions regarding the admixed population evolution pattern in this environment, which might be facilitated by crop-to-wild gene flow.

Keywords: crop-to-wild gene flow, population genomics, genetic diversity, admixture, woody perennial species, *Olea europaea* L.

1 | Introduction

Gene flows between domesticated species and their wild relatives have been identified in several studies in animals [1–3] and plant species [4,5]. This phenomenon is noted when cultivated genomic variants occur in unmanaged naturally occurring populations in natural environments. These admixed populations raise the question of the impact of gene flow on the evolution of natural populations. For instance, the introgression of new genetic diversity inside wild genomes can accelerate their evolution by increasing the frequency of favorable alleles, but can also be detrimental resulting in outbreeding depression, i.e. a loss of fitness in hybrids compared to their parents [6–8].

Cultivated to wild introgressions variants may be found in several plant species. Crop-to-wild gene flow occurs in maize and teosinte, its closest relatives, for instance, which has been found to lead to the acquisition of herbicide resistance in teosinte and, consequently, to high frequency of teosinte forms in maize fields [4]. In some perennial species such as apple, major introgressions with the spread of alleles from the cultivated gene pool to wild populations in Europe have been documented [9]. The resulting admixed populations showed higher fitness than wild apple trees [9,10]. Gene flow from domesticated relatives has also been reported in natural chestnut and poplar populations. Admixed poplar populations have been found in France [11] and along the Danube River in natural environment [12], while the same scenario has been observed in chestnuts in Japan [13]. All these studies have proposed conservation measures for in-situ and ex-situ preservation of genuine wild populations by limiting gene flow, by replanting genuine wild genotypes far from domesticated forms or by protecting the connections of wild metapopulations which can breed and thereby protect themselves from random genetic deterioration [10,12]. The evolutionary

CHAPTER 1

consequences of crop-to-wild gene flows in the natural environment and on wild populations are still misunderstood, especially in perennial species.

Olive tree, (*Olea europaea* L.) is an iconic perennial species from the Mediterranean Basin (MB) which can live thousands of years. Cultivated (*Olea europaea* var. *europaea*) and wild (*Olea europaea* var. *sylvestris*) forms coexist within the same Mediterranean distribution range [14,15]. Wild olive trees have an ancient evolutionary history in the MB [16] indicated that three plastid lineages with a probable common ancestor dating from the Middle to Upper Pleistocene had diversified long before the Last Glaciation Maximum (26,500 to 19,000 BP [17]). They were subsequently impacted by glaciation, while some wild populations persisted in refugia [16,18]. Olive lineages have been isolated in two distant areas, which could explain the current population genetic structure profile of wild olive trees. According to previous genetic studies, two main gene pools are identified, one in the eastern and another in the western/central MB [16,19–21]. This eastern/western genetic differentiation is also found in other plants in the MB [22,23]. Cultivated olive trees emerged with the domestication of olive trees around 6,000 years BP [14,24,25]. It is generally considered that the center of primary olive domestication, from wild progenitors, is located in the Middle East, near the border between Turkey and Syria [14,16,24].

It is currently impossible to distinguish between genuine wild, admixed and cultivated olive trees in the natural environment because of the absence of easily measurable discriminating morphological traits in the field. The use of geometric component of shape stone allows to distinguish the wild (round-shaped stone) from cultivated morphotype [26,27]. However, numerous intermediate morphotypes have been found, ranging from elliptical to more tapered, reflecting the complex history and evolution

processes related to human-associated migration [19,21,26,28]. Genuine wild and admixed olive trees can only distinguish using genetic markers [19,21,28–31]. The genetic diversity of cultivated olive trees is close to eastern MB wild olive trees [20,30,31], hence making it difficult to study gene flow between eastern MB wild and cultivated accessions. Conversely, in western MB, the genetic diversity of wild olive trees growing in natural areas is clearly different from cultivated accessions [16,21,28], thereby enabling the identification of genuine wild olive trees as previously reported by [31] using allozyme markers. The genetic pattern observed in naturally occurring populations was little impacted by crop-to-wild gene flow [21,28,30]. The well-known genetic differentiation makes it a relevant model for investigating the genetic structure of populations in their natural environments and to infer potential gene flow between cultivated and wild olive.

Here, we investigated the genetic diversity of naturally occurring olive tree populations in the western MB. We assumed that the genetic pool of wild olive tree in the western Mediterranean area has not been impacted by introgressions from domesticated forms. We addressed the following questions: (1) What is the genetic structure of spontaneous olive trees in the western MB? (2) Are there genuine wild olive populations in this range? (3) Is there crop-to-wild gene flow in this region? We analyze genome-wide SNPs in olive trees from 27 natural sites ranging from southern France, Spain and Morocco. We included DNA from wild trees previously sampled in southern Turkey for the purpose of comparing diversity in these populations with the genetic pattern in the eastern MB wild gene pool [30]. We sought to identify crop-to-wild gene flow and patterns of admixtures using data of cultivated accessions from western MB from the Worldwide Olive Germplasm Bank of Marrakech (WOGBM) and Porquerolles obtained with the same sequencing strategy [32,33].

2 | Methods

2.1 | Sampling of wild olive trees on a north-south gradient in the western Mediterranean Basin

Sampling of 27 assumed wild olive tree sites was conducted along a north-south gradient from southern France to southern Morocco in 2021 and early 2022. Sites were selected via the Conservatoire Botanique National Simethis database (<http://simethis.eu>) for southern France and northern Spain whereas for Corsica, central and southern Spain and Morocco, the delineation of wild populations was based on plastid polymorphism as reported by [21]. In addition to information from the Simethis database, we used environmental criteria to limit sampling of admixed olive populations and disregard olive orchards, agricultural and urban environments [34]. This resulted in the selection of 27 sites (Table 1; Fig 1A). At each of them, 13 to 15 individuals were sampled, representing a total of 400 sampled wild olive trees. At each site and for each tree, leaves were collected and immediately dried in silica gel for subsequent DNA extractions.

Table 1. Summary of the localisation of the natural olive populations sampled for the study.

Site	n	Country	Latitude	Longitude	Localisation	Source
F01	15	Continental France	42.9368	3.0126	Leucate	This study
F02	15	Continental France	43.5407	3.3033	Lac des olivettes - Valhan	This study
F03	15	Continental France	43.7706	3.7919	Cazevieille	This study
F04	15	Continental France	43.8774	4.7328	Avignon	This study
F06	15	Continental France	43.3826	6.3611	Plain des Maures - Gonfaron	This study
F07	15	Continental France	43.6894	7.3029	MontBoron - Nice	This study
F08	15	Corsica France	42.6538	9.0638	Ile Rousse	This study
F09	15	Corsica France	42.4037	8.6989	Manso - Calvi	This study
F10	15	Corsica France	41.7503	8.8688	Filitosa - Propriano	This study
F11	15	Corsica France	41.3723	9.202	Bonifacio	This study
S12	15	Spain	42.2355	3.2188	Roses	This study
S13	15	Spain	41.4208	1.976	Barcelone	This study
S14	15	Spain	41.0202	0.9348	L'Hospitalet del Infant	This study
S15	15	Spain	40.3415	0.3858	Peníscola	This study
S16	15	Spain	38.8029	0.1952	Xàbia	This study
S17	15	Spain	38.3737	-3.507	Santa Elena	This study
S18	15	Spain	37.2563	-6.2085	Aznalcar - Sevilla	This study
S19	15	Spain	36.7656	-3.8496	Malaga	This study
S20	15	Spain	36.0615	-5.6695	Tarifa	This study
M21	15	Morocco	35.79	-5.9248	Cap Spartel	This study
M22	15	Morocco	35.7828	-5.5153	Douar Dakchire	This study
M23	13	Morocco	34.8685	-5.3526	Douar Nefzi	This study
M24	15	Morocco	33.5341	-5.9082	Bouquachmir	This study
M25	15	Morocco	33.0998	-5.5883	Moyen-Atlas - M'rit	This study
M28	15	Morocco	31.2104	-8.0398	Marrakech-Asni	This study
M29	13	Morocco	30.6315	-9.3704	Ameskroud-Idmine	This study
M30	15	Morocco	31.111	-9.6907	Agadir - Essaouira	This study
OST	15	Turkey	36.1136	33.43209	Tisan	[30]

n : sample size

CHAPTER 1

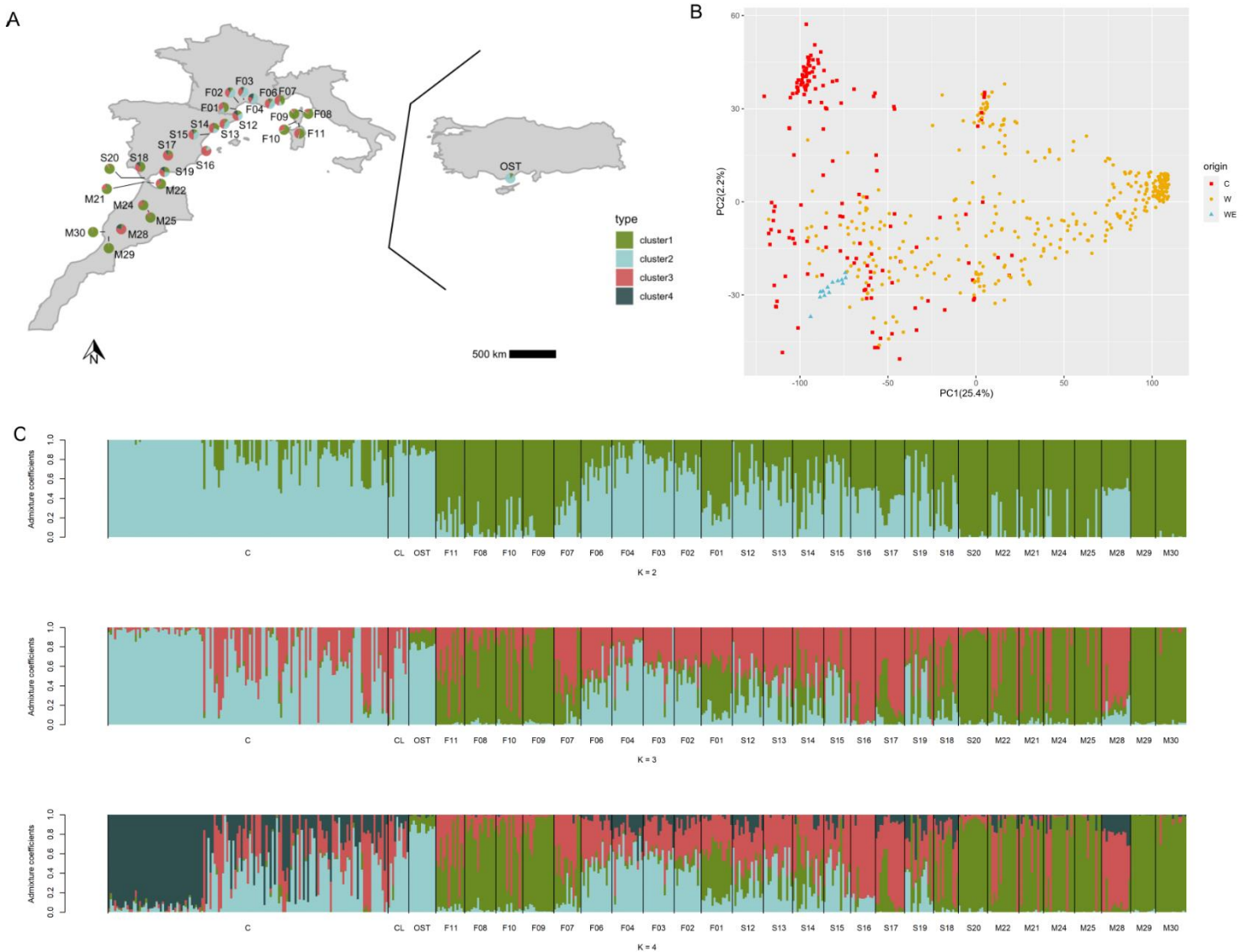


Fig 1. Result of the population structure analyses performed on the genome-wide SNPs diversity of natural populations of *O. europaea L.* collected in France (143), Spain (123), Morocco (96) and Turkey (13) and cultivated *O. europaea L.* from the western Mediterranean Basin (145), using 142,060 SNPs. (A) Geographical distribution of the populations and proportion of genetic cluster assigning. Pie chart at each location represents the fraction of individuals belonging to each genetic cluster as inferred by sNMF (K = 4). (B) PCA inferred with LEA. C: Cultivated, W: Western wild, WE: Eastern wild. (C) Genetic structure inferred by

sNMF, each horizontal bar indicates individual assignment to a genetic cluster with K being the number of genetic clusters (K from 2 to 4).

2.2 | Reference set of cultivated and eastern wild accessions

In addition to the sample sites described above, leaves from 10 cultivated olive varieties were added to the sampling. These varieties were selected because of their significant presence in the French sampling area of natural populations [33], which can be a potential source of introgression. Fifteen individual wild olive trees from Turkey in the eastern MB [30] were also added to create a genetically distinct group which will be considered as an outgroup (Fig 1A). Moreover, 135 cultivated varieties from the WOGBM, representative of the genetic diversity of olive resources in the western MB [32] were considered as reference varieties to assess the introgressions from cultivated olive into wild populations. This last dataset was developed in a parallel study by our group that is focusing on cultivated olive (S2 Table). Overall, the experiment included 561 individuals.

2.3 | Bait design

The cultivated *Olea europaea* var. *europaea* (cv. Farga) Oe9 genome assembly [29] was used as a reference to design target sequencing probes. This genome is 1.38 Gb. Baits were designed according to the following parameters: place 80 bp probes with 0.5x tiling targeting the first 640 bp of each of the 55,595 annotated genes available. For each gene, 1 to 4 baits were designed depending on its length. After quality filtration, this resulted in a total set of 210,367 baits representing 55,452 unique loci and a captured length of 16.8 Mb. The probes were designed and synthesized by Daicel Arbor Biosciences, Ann Arbor, Michigan, USA. For this study, we only retained

CHAPTER 1

sequencing data targeted on assembled chromosomes. This subset represented 102,126 baits with a captured length of 8.2 Mb (S1 Table).

2.4 | Library preparation and sequencing

DNA was extracted from leaves using a mixed alkyl trimethylammonium bromide buffer (MATAB) and NucleoMag Plant Kit (Macherey-Nagel, Düren, Germany) as already described by [35] (S1 File). Individual genomic libraries for the NGS experiments were constructed with the NEBNext® Ultra™ II FS DNA Library Prep Kit (New England Biolabs, Ipswich, MA) with inputs ≤ 100 ng (S1 File). DNA was enzymatically sheared at an average 160 bp length before being tagged with the Unique Dual Index. Enrichment by capture was performed with biotinylated RNA probes (80 bp) as recommended by the provider using myBaits kits (Arbor Biosciences). A single dose of bait was used on a bulk of 48 normalized libraries. The sequencing was performed by MGX-Montpellier GenomiX on an Illumina® Novaseq™ 6000 (Illumina Inc., San Diego, CA, USA) platform with an S4 flow cell. In addition to the target sequencing data set, we also sequenced four whole genomes (OES_E13_09, OES_F10_03, Picholine and Picholine Marocaine) to calculate the enrichment rate of the target sequencing method.

2.4.1 | SNP calling

Raw sequencing reads were first trimmed with FastP version 0.20.1 [36]. The resulting data were then mapped on the reference genome Farga Oe9 genome assembly [29] using bwa-mem2 version 2.0 [37]. The mapped reads were sorted with samtools version 1.10 [38]. Only primary alignment, properly paired and unique reads were kept. Duplicates were removed using picard-tools version 2.24.0 [39]. From this clean alignment, GATK version 4.2.0.0 [40] was used for the SNP calling according to

GATK4 best practices (<https://gatk.broadinstitute.org/hc/en-us/sections/360007226651-Best-Practices-Workflows>).

2.4.2 | SNP filtering

After sequencing, we obtained 27,275,679 raw data. These were filtered with VCFtools version 0.1.16 [41]. The following filters were sequentially applied. With vcf-annotate, we first removed SNPs with a quality below 200 and clusters of 3 SNPs in 10 bases. VCFtools was used to remove indels, to only keep biallelic SNPs, to select SNPs with a minimum depth per site of 8 and a maximum mean depth per site of 400. Sites with >15% missing data were removed, then individuals with >20% missing data were also removed (S4 Table). After filtration, 35 individuals were removed (1 cultivated, 2 eastern wild plants and 32 western wild plants). Only sampling sites with at least 12 individuals were considered in this study. M23 had only 6 individuals left and was therefore removed from the data-set. Filtering was carried out to exclude positions with fixed heterozygosity (>85%) and the final filtering was done to keep at least one minor allele count per site. After all filtering steps, we obtained 142,060 SNPs in the final data set, these were located on all chromosomes (S3 Table).

2.5 | Genomic analysis

2.5.1 | Genetic diversity and genetic structure

Genetic diversity measure was examined for each of sampled sites, considered as distinct populations. We calculated diversity measures as expected heterozygosity (H_E), observed heterozygosity (H_o) and inbreeding coefficient (F_{IS}) using Hierfstat version 05-11 [42].

CHAPTER 1

Pairwise between population genetic differentiations were estimated with Weir and Cockeman fixation index (F_{ST}) using pairwise.WCfst function from the Hierfstat package version 05-11 [42]. Support values were calculated per locus, for each pair of population, based on bootstraps procedure (S5 Table).

Genetic structure analyses were conducted using sNMF and Principal Component Analysis (PCA) from the LEA package version 3.11.3 [43]. Both analyses were performed on the dataset. It included 135 individuals from the WOGBM, originating from Spain, France and Morocco according to their passport data, 10 cultivated samples from southern France, 362 western MB wild individuals and 13 eastern MB wild samples. For sNMF, five repetitions per clusters (K) considered, were performed with K ranging from 1 to 10.

2.5.2 | Admixture assessment

Inference of the population history, with the admixture and split pattern were done using TreeMix version 1.13 [44]. This software constructs admixture graphs using allele frequencies of current genetic populations to infer a graph of all ancestral populations related to a common ancestor. For this analysis, we grouped cultivated olives in four different genetic groups, depending on their sNMF assignment to genetic ancestral clusters: C0, C1, C3 and C4 (S1 File; S6 Table). With these clusters, the 27 populations collected in western MB and the Turkish population, we did 100 TreeMix runs with a random SNP block size between 100 and 1000, from 1 to 10 migrations each when considering the M29 population as an outgroup. We inferred the optimum number of migrations with multiple linear models and the Evanno method implemented in the *OptM* package version 0.1.6 (S1 and S2 Figs) [45,46]. TreeMix analysis was performed with 500 bootstrap replicates, which were used to build a consensus tree

with Phylip version 3.697 [47]. We used BITE packages version 1.2.0008 [48] to display the trees.

3 | Results

In this study, we analyzed 520 individuals, including 145 cultivars from Spain, France and Morocco, representing the MB olive diversity of cultivated olive trees (S2 Table), a set of 362 wild trees from France, Spain and Morocco collected in 27 natural sites and 13 wild trees from southern Turkey.

3.1 | Target sequencing efficiency

The average enrichment rate in the target sequencing experiment was 34 times higher than expected with whole-genome sequencing (S7 Table). Moreover, for the bait on the chromosome annotated genes, 63.5% of the filtered SNPs (90,157) were on-target SNPs. The remaining was off-target SNPs. The on-target SNPs corresponded to sequences targeted by the baits. Conversely, the off-target corresponded to nonspecific and unintended sequences that can arise through sequencing (S7 Table).

3.2 | Genetic diversity and genetic structure

The average inbreeding coefficient (F_{IS}) calculated on the different populations was on average 0. This is in accordance with the outcrossing mating system of olive tree. For 3 populations we detected F_{IS} values ranging from -0.086 to -0.101 (F03, F11, S13) while 3 other ones F_{IS} values ranging from -0.192 to -0.206 (S16, S17 and M28) (Table 2). All of these populations might be resulting from admixture event.

Table 2. Summary information of genetic diversity for sampling sites of naturally occurring olive trees in the western and eastern Mediterranean Basin.

	n	H_o	H_E	F_{IS}
F01	15	0.142 (0-0.337)	0.141 (0-0.319)	-0.003±0.263
F02	13	0.147 (0-0.346)	0.144 (0-0.324)	-0.014±0.265
F03	15	0.145 (0-0.358)	0.130 (0-0.307)	-0.086±0.234
F04	15	0.134 (0-0.336)	0.124 (0-0.298)	-0.056±0.247
F06	15	0.148 (0-0.350)	0.142 (0-0.322)	-0.032±0.249
F07	13	0.154 (0-0.372)	0.141 (0-0.325)	-0.070±0.262
F08	15	0.127 (0-0.324)	0.121 (0-0.294)	-0.029±0.262
F09	15	0.121 (0-0.315)	0.118 (0-0.290)	-0.017±0.273
F10	13	0.135 (0-0.340)	0.126 (0-0.302)	-0.053±0.271
F11	14	0.147 (0-0.369)	0.129 (0-0.310)	-0.101±0.252
S12	15	0.158 (0-0.372)	0.146 (0-0.330)	-0.057±0.242
S13	14	0.155 (0-0.380)	0.136 (0-0.318)	-0.098±0.246
S14	15	0.154 (0-0.358)	0.149 (0-0.333)	-0.026±0.254
S15	13	0.148 (0-0.353)	0.142 (0-0.324)	-0.024±0.264
S16	12	0.187 (0-0.461)	0.144 (0-0.331)	-0.206±0.319
S17	14	0.168 (0-0.432)	0.131 (0-0.318)	-0.195±0.314
S18	12	0.147 (0-0.351)	0.141 (0-0.323)	-0.028±0.271
S19	14	0.140 (0-0.339)	0.140 (0-0.324)	0.001±0.280
S20	14	0.123 (0-0.324)	0.115 (0-0.289)	-0.043±0.269
M21	12	0.139 (0-0.345)	0.129 (0-0.304)	-0.051±0.262
M22	15	0.141 (0-0.337)	0.135 (0-0.308)	-0.030±0.246
M24	15	0.139 (0-0.341)	0.131 (0-0.305)	-0.045±0.247
M25	13	0.127 (0-0.324)	0.123 (0-0.297)	-0.020±0.270
M28	14	0.175 (0-0.437)	0.139 (0-0.328)	-0.192±0.295
M29	12	0.114 (0-0.312)	0.111 (0-0.287)	-0.018±0.307
M30	15	0.121 (0-0.319)	0.115 (0-0.289)	-0.031±0.268
OST	13	0.117 (0-0.300)	0.116 (0-0.282)	-0.007±0.271

n, number of genotypes; H_E , expected heterozygosity; H_o , observed heterozygosity; F_{IS} , population level deviation from Hardy-Weinberg heterozygosity

The pairwise differentiation values between the studied populations (F_{ST}) ranged from 0.01 to 0.42 (Fig 2). The Turkish eastern wild sampling site was the most genetically differentiated from the western MB sites, such as Corsica, F01 and F07 in Continental France, S14, S17, S18 and S20 in Spain, and all sites in Morocco (all above 0.2). Compared to the Turkish wild population, M29 from the southern limit of the olive

distribution is the most differentiated population ($F_{ST} = 0.42$), while F04 from France was the least differentiated population ($F_{ST} = 0.01$). We revealed a high genetic differentiation between western MB wild populations and the eastern MB wild population.

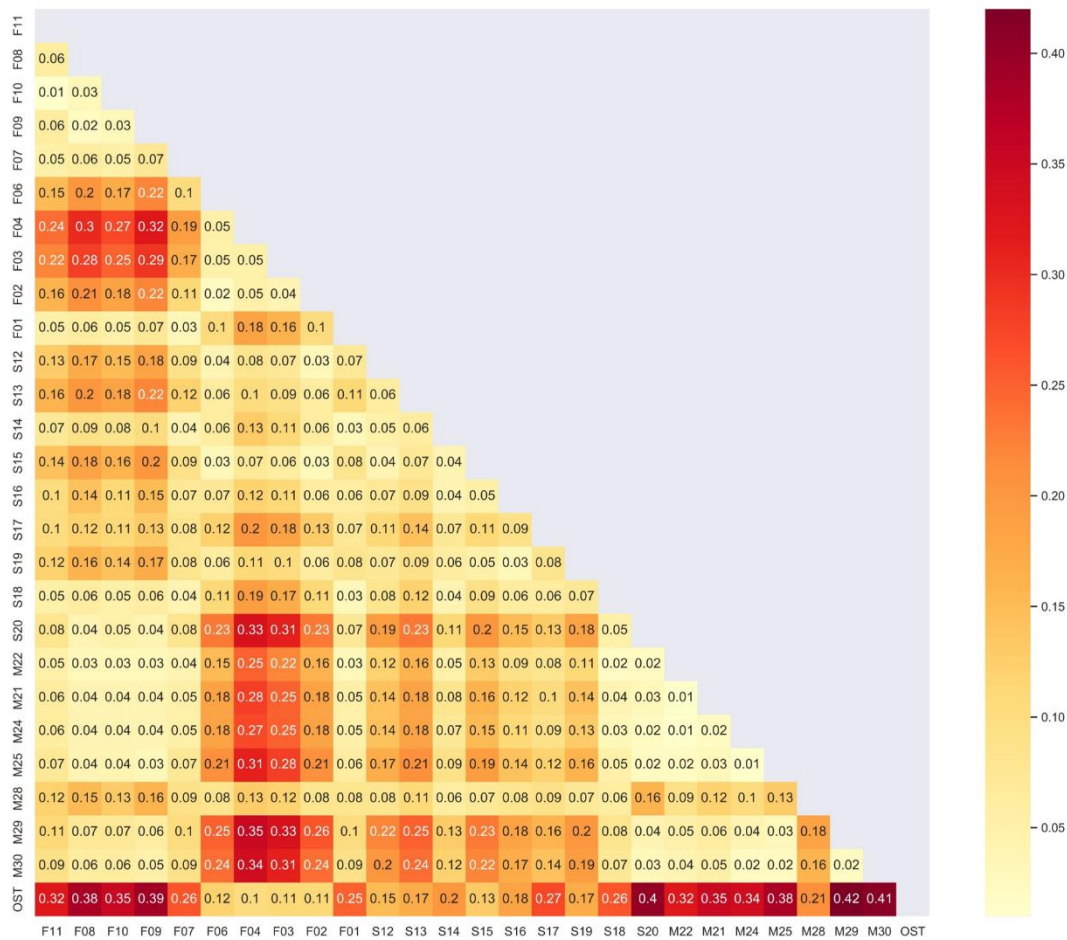


Fig 2. Heatmap of pairwise F_{ST} performed on the genome-wide SNPs diversity of 27 natural populations of *O. europaea* L. collected in western Mediterranean Basin in France (143), Spain (123), Morocco (96) and in the eastern Mediterranean Basin in Turkey (13).

In the principal component analysis (PCA), the first axis PC1 accounted for 25.4% of the variation and revealed an eastern-western genetic structure between western MB

CHAPTER 1

wild olive trees and eastern MB wild olive trees, with cultivated accessions mainly related to the eastern MB wild populations (Figs 1B and 3). On the first axis, we observed accessions from sampling sites in Corsica (F08 to F11, Fig 3B), S20, a large part of S18 from Spain (Fig 3C) and from all the sites in Morocco (Fig 3D), with the notable exception of M28, were clearly separated from the cultivated accessions and eastern MB wild accessions (Fig 3). All individuals collected in central southern France (F02, F03, F04 and F06; Fig 3A), one from Morocco (M28, Fig 3D) and some from north-central Spain (S12, S13, S15 and S16; Fig 3C) clustered with cultivated accessions as shown in the Figs 1B and 3 (left side of the PCA). This profile suggests admixture events. Several other individuals collected in eastern and western France (F01 and F07; Fig 3A), in Morocco (M21, M22, M24 and M25; Fig 3D) and in Spain (S14, S17, S18 and S19; Fig 3C) also exhibited a pattern of admixture with the cultivated accessions. The second axis, i.e. PC2, accounted for 2.2% of the observed variability, highlighting two subgroups within cultivated trees. The first subgroup includes cultivated genotypes mostly from Spanish varieties such as “Picual” and three Moroccan varieties including “Picholine Marocaine”. The other cultivated group included several varieties from Spain, Morocco and France (Fig 1B).

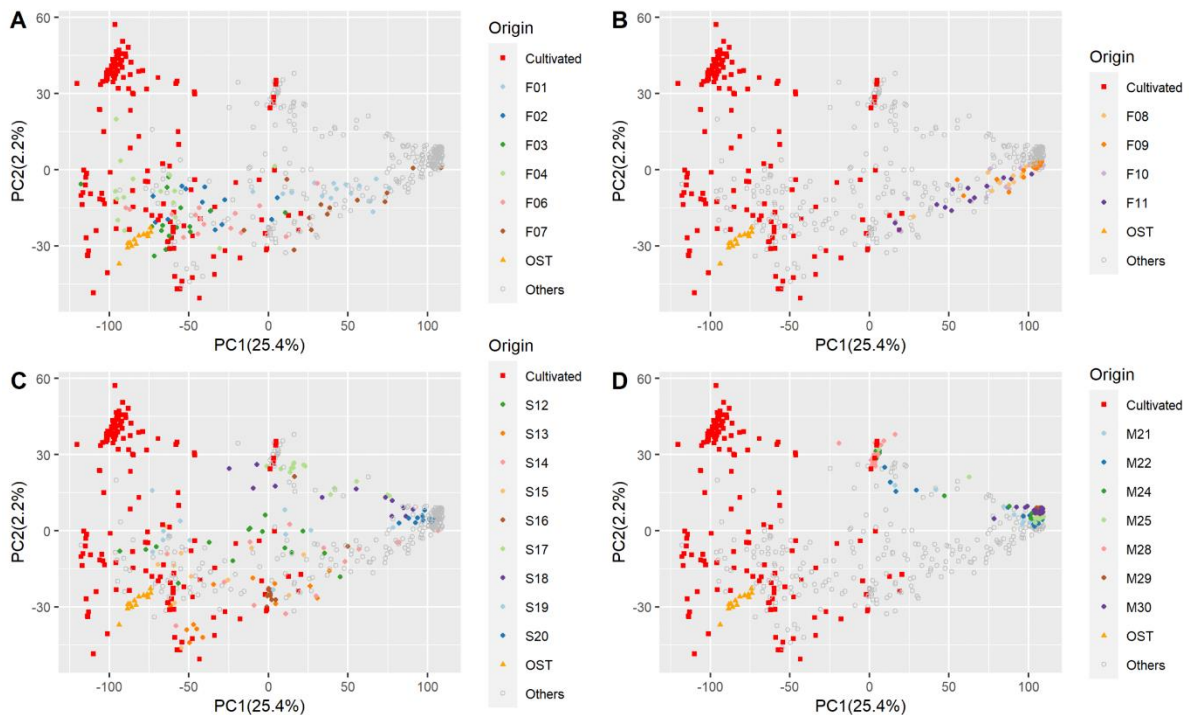


Fig 3. Detailed genetic structure of *O. europaea* L. populations using PCA analysis. Red boxes represent cultivated individuals, orange triangles represent Turkish population with the (A) French continental populations, (B) Corsican French populations, (C) Spain populations, (D) Moroccan populations.

A similar pattern was supported by the sNMF analyses. According to the cross-entropy criterion, only K from 2 to 4 were considered suitable to explain the western MB natural olive tree genetic pattern (Fig 1C; S3 Fig). The wild Turkish olive population (OST) was assigned to a specific cluster from $K=2$ to $K=4$ (in blue) regardless of the admixture model examined, thereby supporting the existence of a structure between western MB and eastern MB natural populations. At $K = 2$, the cultivated and eastern MB olive trees collected in natural sites were assigned to cluster 1 (in blue) and the western MB wild natural olive trees in Morocco, Corsica, South Spain, France (F01 and F07 sites) were mainly assigned to cluster 2 (in green). Olive trees from the F02, F03, F04 and F06 sites in France were mainly assigned to cluster 1 (in blue). Trees

CHAPTER 1

collected from S12, S13, S15, S16, S17 S19 and M28 sites spread in the two clusters. At $K = 3$, these sampling sites were mostly assigned to cluster 3 (in red), particularly for individuals from populations M28, S17, and S16. With $K = 4$, a fourth cluster (in dark blue) was noted within the cultivated cluster, corresponding to the same first subgroup described in the PCA results above (Figs 1B and 3).

By combining three analyses (i.e. PCA, sNMF and pairwise F_{ST} ; Figs 1 and 3), particularly by considering the left and central part of the PCA (Figs 1B and 3) and the cluster 3 (in red; Fig 1C) from the sNMF analyses, we have several arguments strongly suggesting admixtures between natural olive trees and cultivated ones. It seems to be consistent with F_{ST} values, with much lower levels of differentiation between cultivated olive accessions and wild olive accessions (Fig 2). Accordingly, all olive trees mapped in the right side of the PCA (Figs 1B and 3) and assigned to the cluster (in green) regardless of the admixture model examined (Fig 1C) and considered as genuine wild.

3.3 | Inference of population admixture and gene flow

The tree inferred by TreeMix was ranked using M29 population. This population was chosen because all of the individuals collected in this site belong to the same cluster, referred as the western MB wild cluster (in green; Fig 1C). This genuinely western MB wild olive also appeared to be the most genetically distinct from eastern MB wild olive ($F_{ST} = 0.42$) and from the cultivated accessions (Fig 1).

The TreeMix analysis revealed the highest divergence between M29 and C3 (cultivated) and OST (Turkish population). The lowest divergence (below 0.015) from M29 were found for with almost all the Moroccan sites (except for M28), with S20, F08, F09, F10 and F11. A second group of sampling sites was found with a divergence from M29 of 0.017 to 0.024, including S18, M28, F07, F01 and S17. Accessions from

sampling sites in Spain, except for S20, had a genetic divergence of >0.032 from M29 and were closer to cultivated groups (<0.012 genetic divergence between S16 and C3). Accessions from the French F06, F02, F03 and F04 sites were found to be grouped with the cultivated clusters C3 and C0. TreeMix inferred a low divergence between C4, C1 and OST (around 0.005). Two gene flow events were inferred (S1 and S2 Figs), with the first one being from cultivated and French populations from the center to M28, with a weight of ($w = 0.425$). The second one was from M22 to northern Spain sites ($w = 0.485$) (Fig 4).

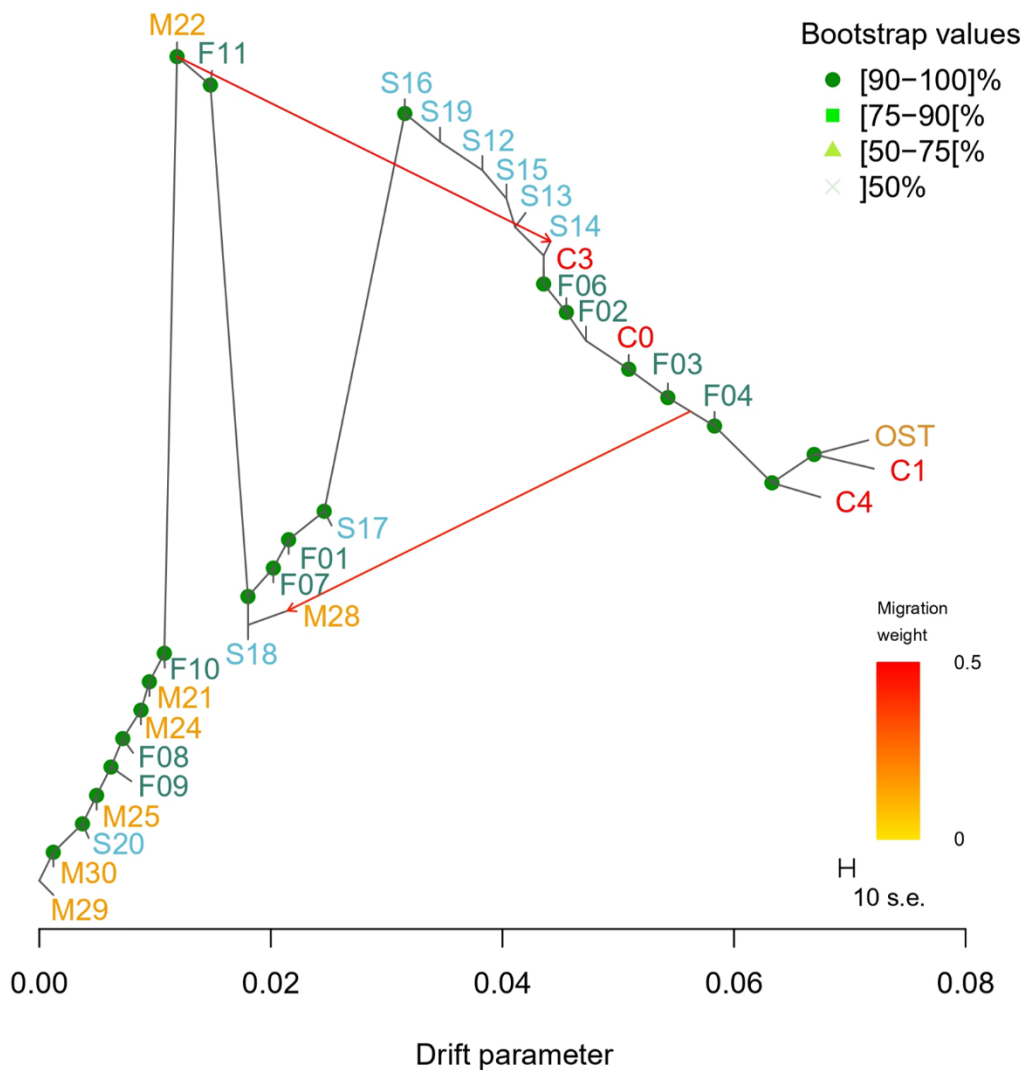


Fig 4. Tree inferred by TreeMix analysis on natural *O. europaea* L. populations from the western and eastern Mediterranean Basin and cultivated accessions. Genetic divergence is represented by the horizontal difference between populations. The vertical bars are only graphical representations and are not taken into account in the analysis. C0, C1, C3 and C4 are groups of cultivated accessions (S1 File).

4 | Discussion

Wild olive molecular identification and characterization are important to assess the genetic diversity of this species, in addition, to evolutionary history and to investigate local adaptation. This is particularly important for wild relative crops such as olive, where wild olives and cultivated olives coexist in the same area [14,15] and cannot be clearly distinguish with morphological traits [26,27] but only using genetic markers [19,21,28–31]. As crop-to-wild gene flows likely occur, deciphering their genetic relationships can help to understand the impact of crop on natural populations, their demographic histories and to explore new sources of genetic diversity. In this study, our aim was to identify genuine wild olive populations by investigating the genetic structure and diversity patterns of olive trees evolving in the natural environment in the western MB. We sampled allegedly wild olive trees according to past study [21] over a large geographic area in natural environment within the western MB. We hypothesis that all these populations are genuine wild olives. In addition, we included eastern MB genuine wild olive trees and cultivated olive trees. We analyzed this large panel of population using SNPs from target sequencing. This if the first time target sequencing has been used to study genomic variation in olive tree—it enabled a genome scan of many individuals while accessing more than 140,000 SNPs distributed throughout the genome. This represents a major advance over previous methods in similar studies using SSR-based molecular analyses [21,28] or SNPs analyses from RNAseq [30]. We documented the presence of genuine wild olive tree populations in the western MB, in southern France, Spain and Morocco. However, contrary to our assumption, our analysis suggested that admixed populations are more frequent than genuine ones.

4.1 | Genetic variation patterns in natural and cultivated olive trees highlight the persistence of genuine wild olive populations in the western Mediterranean Basin

Based on a large set of SNP markers, located all over the genome, we identify a very strong genetic differentiation between the natural populations collected in the western MB and those in the eastern MB. This confirmed the results of previous studies using microsatellite markers highlighted two distinct gene pools in the eastern and western MB. In these previous studies, the western/central MB gene pool clustered wild populations from Greece to Morocco [21,28] while the eastern MB gene pool clustered wild populations from Greece to the Levant. The eastern MB gene pool was represented here by one Turkish population (OST), i.e. a genuine wild population previously revealed using plastid DNA polymorphism [16,21], SSR markers [28] and SNPs from RNAseq [30]. Our results reflects the typical long evolutionary history of Mediterranean species as described by [49].

Our study confirmed that the wild eastern MB olive population from Turkey is very closely related to cultivated olives, as shown by different approaches (PCA, sNMF, F_{ST} and TreeMix). This finding is consistent with the olive domestication history which was likely the results of genetic selection of eastern MB wild trees [16,28,30]. Moreover, we combined several approaches, i.e. PCA, sNMF, pairwise F_{ST} estimation and inference of splits in mixture in populations and clearly identified a genetic group of olive trees strongly differentiated from the eastern wild and cultivated olive as previously found by [30] which probably constitute genuine wild olive trees. Finally, within cultivated accessions, we identified two genetic groups. The first one similar to OST and which might be composed of varieties likely issued from two processes, primary selection in the east and secondary diversification in the central and western

Mediterranean areas as proposed by Khadari & El Bakkali (2018). The second group essentially consisted of Spanish and Moroccan varieties that were highly differentiated from OST and which might be composed of western MB cultivated olives through a secondary diversification process mainly via selection involving crossing between ancient varieties such as Gordal Sevillana and Lechin de Granada, as shown by [19].

4.2 | Wild versus admixed olive trees: an evolutionary history impacted by domestication

As discussed above, we identified genuine wild olive trees in the western Mediterranean basin in Spain, Corsica (France) and Morocco as already shown by [21]. The genetic diversity observed in Corsican and Moroccan populations were very close suggesting a common ancestral history for these populations. We also discovered, for the first time, the presence of genuine wild olive trees in continental France at Mont-Boron (eastern-south; F07) and near Leucate (western-south; F01). The abundance of wild olives trees in the western MB has been well characterized [16,20,28,30,31]. However, populations initially characterized as genuine wilds according to plastid DNA polymorphism and SSR markers, for instance, in Spain [21] were found admixed in our study. These sites were in habitats considered to be little or not at all impacted by human activities, e.g. in natural reserves, remote from urban centres or areas with olive orchards. Even within some sites, from a genetic viewpoint, several individuals were considered genetically to be wild, while others were very admixed with cultivated olives. This high genetic admixture intensity was unexpected in the western MB based on previous studies [16,21,28,30,50]. This differing results compared to past study [21,28] may arise from variations in the sampling distribution area and the size of the samples at each site, particularly with the presence of heterogeneous populations, including admixed and wild individuals. The SSRs should

CHAPTER 1

be able to detect a similar pattern of admixture. Our findings have provided new insights into the evolutionary history of olive trees in natural western MB habitat.

4.3 | What factors could influence the prevalence of admixed populations in natural environments?

We obtained clear evidence in this study on the substantial presence of admixed populations within the natural olive populations which were previously reported to be little impacted by crop-to-wild gene flow [21,28]. Recent phylogenomic and population structure investigations revealed genetic admixtures during olive domestication thus highlighting the impact of domesticated alleles on two wild olive trees for the western MB [29]. However, these authors analyzed very limited sampling of wild olive trees (7 olive trees from the western MB), while in our investigations, we analyzed 362 trees sampled from 27 natural sites in Spain, Morocco and France—this sampling covered the olive natural distribution range in the western MB. Through TreeMix analysis and comparing genomic variation comparison in these 362 olive trees to the genuine Turkish natural population and to cultivated olive trees, we, therefore, were able to depict gene flow between wild and cultivated olives (see Fig 4). The observed patterns highlighted the two kinds of gene flow events: from cultivated to wild populations in Morocco, France and Spain and from wild to cultivated or already admixed populations in Spain.

These gene flows could have been driven by several factors. First, the mating system of olive, which, is allogamous, pollination mostly depends on wind (anemophilous pollination) and seed dissemination relies on birds (zoochorous dissemination). These forms of dissemination may occur over long geographic distances (>50 km) [51]. Second, cultivated and wild olive trees share the same climatic and ecological niches,

the geographic proximity between them increases the possibility of gene flow and events of admixture [50]. Third, there could be cultivated versus wild olive tree pollen competition: monocultures and single-varietal olive orchards are responsible for broad dissemination of pollen from orchards (thousands of trees), whereas wild populations are often composed of few individuals. Wild olive pollen is thus less abundant. Fourth, gene flow between cultivated and wild olive trees may increase genetic diversity in admixed populations. Associated new variants or combination might be better adapted to the local environment, promoting an acceleration of local adaptation of a species to an environment is a recognized evolutionary force explaining the occurrence of admixed populations [8]. Evolutionary factors such as allogamy [52] and cultivated versus wild olive tree pollen competition (see above) may not be sufficient to explain the large frequency of admixed populations versus genuine wild populations. Here we assume that admixed olive trees could have a better adaptive potential to their natural environment, as this has been previously demonstrated in several short-lived and annual crops [3,4,10]. This assumption is supported by the findings of genetic investigations on natural olive trees in Australia [53]. These authors hypothesized that hybridization between two introduced *Olea* species, *Olea europaea* subsp. *europaea* and *Olea europaea* subsp. *cuspidata*, overcame the lack of diversity after their introduction bottleneck, thereby facilitating their establishment. Strikingly, to our knowledge, this assumption has yet to be investigated in long-lived woody plants such as olive trees, even though knowledge on the impact of admixture on the evolution of natural populations could help guide appropriate conservation strategies in forest areas and other natural ecosystems.

4.4 | Consequences of extensive hybridization of wild olive via domesticated olive introgression and conservation recommendations

The future of genuine wild genotypes might be threatened by the gene flow we highlighted here. For instance, extensive gene flow could ultimately lead to complete replacement of wild populations by admixed genotype [8]. However, in global change context, this gene flow could enhance adaptation to a changing environment. Our study offers new opportunities for more in-depth studies on this introgression process. We identified three different compartments, i.e. a genuine wild olive compartment, a cultivated one and an admixed one, that could be study to address this long-standing question. Conservation policies on wild olive trees should take into account the risk of introgression from cultivated alleles and by the impact of climate change. The naturally occurring olive trees sampled here were positioned on a north-south gradient with different environmental conditions, which could facilitate studies on their potential local adaptation to changing climatic conditions. Our study finding may provide a basis for designing new conservation measures to protect genuine wild genotypes, *in-situ* and *ex-situ*, including repositories of wild genetic diversity not impacted by artificial selection.

5 | Conclusion

In this study we assessed the genetic structure of natural olive populations from the western Mediterranean Basin. We confirmed that the western MB genuine wild olive is genetically well differentiated from eastern MB wild olive as well as cultivated forms. We detected its presence in France, Spain and Morocco. We also found many admixed populations resulting from strong crop-to-wild gene flow. The presence of admixed olive populations in the same distribution area as genuine wild populations

raises questions on the reasons for their predominance in the natural environment and on designing conservation strategies for both compartments. Finally, the two genetic patterns revealed by our investigations could be considered as a suitable model for investigating two core questions, the first on the admixture nature, i.e. what domesticated genomic alleles/regions would be suitable for introgression in wild genomes? The second question is related to local adaptation: are wild better locally adapted than admixed olive trees?

Acknowledgements

We thank Guillaume Perez, Magalie Delalande, Karine Loridon, Juliette Arzul, Sylvia Lochon-Menseau for helping us in wild olive sampling from in situ areas; Christian Pinatel for the sampling of French cultivated olive trees; Bénédicte Rhoné for her advices in the target sequencing design. Anaïs Fossot, Hélène Vignes and Ronan Rivallan for the laboratory help. We acknowledge the helpful comments from Joëlle Ronfort and the English corrections from David Manley.

A part of a field work has been supported by the Conservatoire Botanique National Méditerranéen de Porquerolles, l'Institut Agro Montpellier and INRA Morocco. This work has been realized with the support of MESO@LR-Platform at the University of Montpellier, the CIRAD, and the technical support of the bioinformatics group of the UMR AGAP Institute, member of the French Institute of Bioinformatics (IFB) - South Green Bioinformatics Platform

References

1. Mary N, Iannuccelli N, Petit G, Bonnet N, Pinton A, Barasc H, et al. Genome-wide analysis of hybridization in wild boar populations reveals adaptive introgression from domestic pig. *Evol Appl.* 2022;15(7):1115–28.
2. Anderson D, Negishi Y, Ishiniwa H, Okuda K, Hinton TG, Toma R, et al. Introgression dynamics from invasive pigs into wild boar following the March 2011 natural and anthropogenic disasters at Fukushima. *Proc Biol Sci.* 2021;288(1953):20210874.
3. Johnsson M, Gering E, Willis P, Lopez S, Van Dorp L, Hellenthal G, et al. Feralisation targets different genomic loci to domestication in the chicken. *Nat Commun.* 2016;7:12950.
4. Le Corre V, Siol M, Vigouroux Y, Tenaillon MI, Délye C. Adaptive introgression from maize has facilitated the establishment of teosinte as a noxious weed in Europe. *Proc Natl Acad Sci U S A.* 2020;117(41):25618–27.
5. Hauser TP, Shim SI. Survival and flowering of hybrids between cultivated and wild carrots (*Daucus carota*) in Danish grasslands. *Environ Biosafety Res.* 2007;6(4):237–47.
6. Gering E, Incorvaia D, Henriksen R, Conner J, Getty T, Wright D. Getting Back to Nature: Feralization in Animals and Plants. *Trends Ecol Evol.* 2019;34(12):1137–51.
7. Keller M, Kollmann J, Edwards PJ. Genetic introgression from distant provenances reduces fitness in local weed populations. *J Appl Ecol.* 2000;37(4):647–59.
8. Ellstrand NC, Prentice HC, Hancock JF. Gene Flow and Introgression from Domesticated Plants into their Wild Relatives. *Annu Rev Ecol Syst.* 1999;30(1):539–63.
9. Cornille A, Antolín F, Garcia E, Vernesi C, Fietta A, Brinkkemper O, et al. A Multifaceted Overview of Apple Tree Domestication. *Trends Plant Sci.* 2019;24(8):770–82.
10. Feurtey A, Cornille A, Shykoff JA, Snirc A, Giraud T. Crop-to-wild gene flow and its fitness consequences for a wild fruit tree: Towards a comprehensive conservation strategy of the wild apple in Europe. *Evol Appl.* 2017;10(2):180–8.

11. Chenault N, Arnaud-Haond S, Juteau M, Valade R, Almeida JL, Villar M, et al. SSR-based analysis of clonality, spatial genetic structure and introgression from the Lombardy poplar into a natural population of *Populus nigra* L. along the Loire River. *Tree Genet Genomes*. 2011;7(6):1249–62.
12. Jelić M, Patenković A, Skorić M, Mišić D, Kurbalija Novičić Z, Bordács S, et al. Indigenous forests of European black poplar along the Danube River: genetic structure and reliable detection of introgression. *Tree Genet Genomes*. 2015;11(5):89.
13. Nishio S, Takada N, Terakami S, Takeuchi Y, Kimura MK, Isoda K, et al. Genetic structure analysis of cultivated and wild chestnut populations reveals gene flow from cultivars to natural stands. *Sci Rep*. 2021;11(1):240.
14. Zohary D, Spiegel-roy P. Beginnings of Fruit Growing in the Old World. *American Association for the Advancement of Science*. 1975;187(4174):319–27.
15. Rubio de Casas R, Besnard G, Schönswetter P, Balaguer L, Vargas P. Extensive gene flow blurs phylogeographic but not phylogenetic signal in *Olea europaea* L. *Theor Appl Genet*. 2006;113(4):575–83.
16. Besnard G, Khadari B, Navascués M, Fernández-Mazuecos M, Bakkali AE, Arrigo N, et al. The complex history of the olive tree: From late quaternary diversification of mediterranean lineages to primary domestication in the northern Levant. *Proceedings of the Royal Society B: Biological Sciences*. 2013;280:1–7.
17. Clark PU, Dyke AS, Shakun JD, Carlson AE, Clark J, Wohlfarth B, et al. The Last Glacial Maximum. *Science*. 2009;325(5941):710–4.
18. Carrión Y, Ntinou M, Badal E. *Olea europaea* L. in the North Mediterranean Basin during the Pleniglacial and the Early-Middle Holocene. *Quat Sci Rev*. 2010;29(7–8):952–68.
19. Diez CM, Trujillo I, Martinez-Urdiroz N, Barranco D, Rallo L, Marfil P, et al. Olive domestication and diversification in the Mediterranean Basin. *New Phytol*. 2015;206(1):436–47.
20. Besnard G, Baradat P, Breton C, Khadari B, Bervillé A. Olive domestication from structure of oleasters and cultivars using nuclear RAPDs and mitochondrial RFLPs. *Genet Sel Evol*. 2001;33:S251.
21. Besnard G, Bakkali AE, Haouane H, Baali-Cherif D, Moukhli A, Khadari B. Population genetics of Mediterranean and Saharan olives: Geographic patterns of differentiation and evidence for early generations of admixture. *Ann Bot*. 2013;112(7):1293–302.

CHAPTER 1

22. Rodríguez-Sánchez F, Guzmán B, Valido A, Vargas P, Arroyo J. Late Neogene history of the laurel tree (*Laurus* L., Lauraceae) based on phylogeographical analyses of Mediterranean and Macaronesian populations. *J Biogeogr.* 2009;36(7):1270–81.
23. Arroyo-García R, Ruiz-García L, Bolling L, Ocete R, López MA, Arnold C, et al. Multiple origins of cultivated grapevine (*Vitis vinifera* L. ssp. *sativa*) based on chloroplast DNA polymorphisms. *Mol Ecol.* 2006;15(12):3707–14.
24. Kaniewski D, Van Campo E, Boiy T, Terral JF, Khadari B, Besnard G. Primary domestication and early uses of the emblematic olive tree: Palaeobotanical, historical and molecular evidence from the Middle East. *Biol Rev Camb Philos Soc.* 2012;87(4):885–99.
25. Newton C, Lorre C, Sauvage C, Ivorra S, Terral JF. On the origins and spread of *Olea europaea* L. (olive) domestication: Evidence for shape variation of olive stones at Ugarit, Late Bronze Age, Syria—a window on the Mediterranean Basin and on the westward diffusion of olive varieties. *Veg Hist Archaeobot.* 2014;23(5):567–75.
26. Terral, Alonso N, Buxó I, Capdevila R, Chatti N, Fabre L, Fiorentino G, et al. Historical biogeography of olive domestication (*Olea europaea* L.) as revealed by geometrical morphometry applied to biological and archaeological material. *J Biogeogr.* 2004;31(1):63–77.
27. Terral JF, Bonhomme V, Pagnoux C, Ivorra S, Newton C, Paradis L, et al. The Shape Diversity of Olive Stones Resulting from Domestication and Diversification Unveils Traits of the Oldest Known 6500-Years-Old Table Olives from Hishuley Carmel Site (Israel). *Agronomy.* 2021;11(11):2187.
28. Khadari B, El Bakkali A. Primary Selection and Secondary Diversification: Two Key Processes in the History of Olive Domestication. *International Journal of Agronomy* [Internet]. 2018;2018. Available from: <http://dx.doi.org/10.1155/2018/5607903>
29. Julca I, Marcet-Houben M, Cruz F, Gómez-Garrido J, Gaut BS, Díez CM, et al. Genomic evidence for recurrent genetic admixture during the domestication of Mediterranean olive trees (*Olea europaea* L.). *BMC Biol.* 2020;18(1):1–25.
30. Gros-Balthazard M, Besnard G, Sarah G, Holtz Y, Leclercq J, Santoni S, et al. Evolutionary transcriptomics reveals the origins of olives and the genomic changes associated with their domestication. *Plant J.* 2019;100(1):143–57.
31. Lumaret R, Ouazzani N. Ancient wild olives in Mediterranean forests. *Nature.* 2001;413(6857):700–700.

32. El Bakkali A, Essalouh L, Tollon C, Rivallan R, Mournet P, Moukhli A, et al. Characterization of Worldwide Olive Germplasm Banks of Marrakech (Morocco) and Córdoba (Spain): Towards management and use of olive germplasm in breeding programs. *PLoS One*. 2019;14(10):e0223716.
33. Khadari B, El Bakkali A, Essalouh L, Tollon C, Pinatel C, Besnard G. Cultivated Olive Diversification at Local and Regional Scales: Evidence From the Genetic Characterization of French Genetic Resources. *Front Plant Sci*. 2019;10:1593.
34. Copernicus Land Monitoring Service. CORINE Land Cover [Internet]. Land Copernicus Eu. 2020 [cited 2021 Mar 24]. Available from: <https://land.copernicus.eu/pan-european/corine-land-cover/clc2018>
35. Cormier F, Lawac F, Maledon E, Gravillon MC, Nudol E, Mournet P, et al. A reference high-density genetic map of greater yam (*Dioscorea alata* L.). *Theor Appl Genet*. 2019;132(6):1733–44.
36. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 2018;34(17):i884–90.
37. Vasimuddin M, Misra S, Li H, Aluru S. Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems. In: 2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS). ieeexplore.ieee.org; 2019. p. 314–24.
38. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078–9.
39. Broad Institute. Picard toolkit. [Internet]. 2019. Available from: <https://broadinstitute.github.io/picard/>
40. Poplin R, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, Van der Auwera GA, et al. Scaling accurate genetic variant discovery to tens of thousands of samples [Internet]. *bioRxiv*. 2017 [cited 2023 Mar 28]. p. 201178. Available from: <https://www.biorxiv.org/content/10.1101/201178v2>
41. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics*. 2011;27(15):2156–8.
42. Goudet J, Jombart T. Hierfstat: Estimation and tests of hierarchical F-statistics. R package version 0.5-11. [Internet] 2022. Available from: <https://CRAN.R-project.org/package=hierfstat>
43. Frichot E, François O. LEA: An R package for landscape and ecological association studies. *Methods Ecol Evol*. 2015;6(8):925–9.

CHAPTER 1

44. Pickrell JK, Pritchard JK. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* 2012;8(11):e1002967.
45. Fitak RR. OptM: estimating the optimal number of migration edges on population trees using Treemix. *Biol Methods Protoc.* 2021;6(1):bpab017.
46. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol.* 2005;14(8):2611–20.
47. Felsenstein J. PHYLIP (phylogeny inference package), version 3.5 c. Joseph Felsenstein. [Internet] 2005. Available from: <https://phylipweb.github.io/phylip/>
48. Milanese M, Capomaccio S, Vajana E, Bomba L, Garcia JF, Ajmone-Marsan P, et al. BITE: an R package for biodiversity analyses [Internet]. *bioRxiv.* 2017 [cited 2023 Mar 2]. p. 181610. Available from: <https://www.biorxiv.org/content/10.1101/181610v1>
49. Blondel J, Aronson J, Bodiou JY, Boeuf G. *The Mediterranean Region: Biological Diversity in Space and Time.* Oxford University Press; 2010. 376 p.
50. Besnard G, Terral JF, Cornille A. On the origins and domestication of the olive: A review and perspectives. *Ann Bot.* 2018;121(3):385–403.
51. Spennemann DHR, Richard Allen L. The avian dispersal of olives *Olea europaea*: Implications for Australia. *Emu.* 2000;100(4):264–73.
52. Saumitou-Laprade P, Vernet P, Vekemans X, Billiard S, Gallina S, Essalouh L, et al. Elucidation of the genetic architecture of self-incompatibility in olive: Evolutionary consequences and perspectives for orchard management. *Evol Appl.* 2017;10(9):867–80.
53. Besnard G, Dupuy J, Larter M, Cuneo P, Cooke D, Chikhi L. History of the invasive African olive tree in Australia and Hawaii: evidence for sequential bottlenecks and hybridization with the Mediterranean olive. *Evol Appl.* 2014;7(2):195–211.

Chapter 2



Ex

Selection footprints in *Olea europaea* subsp. *europaea*: Detection in wild, admixed and cultivated olive trees from the Western Mediterranean Basin - Clues to their evolutionary history?

In preparation

Lison Zunino et al

Abstract

Selective sweeps are genetic variations that have undergone selection due to various pressures. In wild perennial plants, these selective sweeps often occur as a result of adaptations, shedding light on the evolutionary history of species within their natural habitat. In the case of cultivated species, selective sweeps may arise from artificial selection, leading to their domestication. *Olea europaea* subsp. *europaea* L., the olive tree, is a perennial species emblematic of the Mediterranean Basin. Both wild and cultivated olive trees share the same geographical range. In natural settings, distinguishing genuine wild from domesticated olive trees is challenging due to the absence of easily distinguishable morphological features. The prevalence in naturally occurring populations of admixed olives, resulting from the hybridization of wild and cultivated trees in the western Mediterranean Basin, raises questions about the adaptive potential of natural olive populations and the selective pressures that have allowed all three genetic groups to coexist in the same area. In this study, we investigated the evolutionary history and genetic selection patterns in *Olea europaea* subsp. *europaea* across three groups: wild, admixed, and cultivated. We analysed previously collected genomic sequencing data from 504 individuals, which included 26 naturally occurring populations categorized as wild or admixed, as well as 145 cultivated olive trees from the worldwide Olive Germplasm Bank of Marrakech and French collections. Our analysis involved examining genetic diversity, utilizing PCAdapt and D_{FS} analysis to investigate their evolutionary history. Our findings revealed distinct selection patterns in cultivated, wild, and admixed olive trees in the western Mediterranean Basin. The positive selection observed in both the wild and cultivated varieties reflects a genetic signature of their ancient evolutionary history in the Mediterranean Basin.

Key-words: Selective sweep detection, population genomics, target sequencing, natural selection, crop-to-wild gene flow, admixture, *Olea europaea* L.

1 | Introduction

Selective sweeps are genomic variations that have undergone selection due to selective pressures. These pressures can encompass diverse environmental factors, including abiotic elements such as temperature and precipitation, as well as biotic factors like pathogens (Mitchell-Olds *et al.* 2007; Jordan *et al.* 2017; Capblancq *et al.* 2023). In addition, selective sweeps may result from artificial selection, eventually leading to the domestication of species. In such cases, individuals with specific phenotypic traits are deliberately chosen for their suitability for cultivation uses (Wright *et al.* 2005; Duan *et al.* 2017).

In wild perennial plants, selective sweeps often result from local adaptations, providing information into the evolutionary history of species within their natural range. Genome scan methodologies enable the study of these selection patterns, particularly in the case of perennial species with long generation times in natural environments (Mitchell-Olds *et al.* 2007; Neale & Kremer 2011). Recent advancements in analytical methodologies, together with the development of Next Generation Sequencing (NGS), make it possible to conduct such studies by examining a large amount of genomic variations (Steane *et al.* 2014; Capblancq *et al.* 2018).

The olive tree *Olea europaea* subsp. *europaea* is a perennial species with a long generation time. It stands as an iconic tree in the Mediterranean Basin (MB; Kaniewski *et al.* 2012). In natural areas, it is complicated to distinguish between genuine wild and domesticated olive trees because of the absence of easily measurable discriminating morphological traits in the field. However, they can be distinguished looking at genotypic markers (Lumaret & Ouazzani 2001; Besnard *et al.* 2013a; Diez *et al.* 2015; Khadari & El Bakkali 2018; Gros-Balthazard *et al.* 2019; Julca *et al.* 2020; Zunino *et al.* 2023), cultivated and wild lineages bearing each a distinct evolutionary history but also

intricate interconnections (Zohary & Spiegel-roy 1975; Rubio de Casas *et al.* 2006). Wild olive trees *O. europaea* subsp. *europaea* var. *sylvestris* have an ancient evolutionary history in the MB (Besnard *et al.* 2013b). Olive trees arrived between 5.59 to 5.33 million years ago within the MB, from north Africa to the south Mediterranean, through two pathways, leading to the apparition of two gene pools in the eastern and western MB respectively (Terral *et al.* 2009; Besnard *et al.* 2013b). Pollen studies showed they survived multiple glaciation events within favourable climatic refuges, with for example the Iberian Peninsula in the western MB or the Levant in the eastern MB (Carrión *et al.* 2010; Langgut *et al.* 2019). This isolation led to fragmentation of populations inside the MB. Favourable climatic conditions in the western MB during the early/mid-Holocene has led to an expansion of wild olive trees within the region, allowing an expansion outside the thermo-Mediterranean climate zone and potential reconnection between populations from refuges (Planchais 1984; Terral *et al.* 2009; Carrión *et al.* 2010; Langgut *et al.* 2019). Cultivated olives appeared more recently, with a domestication event supposedly from the wild in the eastern MB, around 6,000 years BP (Zohary & Spiegel-roy 1975; Kaniewski *et al.* 2012; Newton *et al.* 2014). The question of how olive trees spread within the rest of the MB remains a topic of debate. Study argues in a dispersion after this single event of domestication in the Levant (Besnard *et al.* 2013b). In contrast, other studies argues in favour of a second domestication event in the central MB (Diez *et al.* 2015). A more recent study by Julca *et al.* (2020) presents an intermediate scenario, suggesting a concept of domestication continuum, with a primary domestication event in the Levant, followed by instances of hybridization with local wild genetic populations throughout the MB. Inside the western MB, at least two events of introduction would be at the origin of the cultivated presence, with a first one in the southern Iberian peninsula 2,500 BP years ago (Langgut *et al.*

CHAPTER 2

2019) and a second one in the southern France, arrived with Roman from Greece during first century BC (Leveau *et al.* 1991).

The olive tree is pollinated by wind (anemophilous) and seeds are disseminated by birds (zoochoria), it is an allogamous species with a self-incompatibility system (Spennemann & Richard Allen 2000; Saumitou-Laprade *et al.* 2017). Cultivated olives are genetically compatible with wild olives, resulting in extensive gene flow between both genetic groups. This has led to the emergence of numerous admixed genotypes within the natural environment, making it challenging to distinguish them from the wild or cultivated forms in the western MB without the use of molecular markers (Zunino *et al.* 2023). These hybridizations events might have steadily increased over time due to the expansion of olive cultivation based on clonally-propagated varieties (Lumaret *et al.* 1997, 2004; Breton *et al.* 2006).

The large occurrence of admixed genotypes, resulting from hybridization between wild and cultivated genotypes, in the natural environment of western MB (Zunino *et al.* 2023) raises questions about the adaptive potential of natural olive populations and the selective pressures that have allowed the coexistence of all three genetic groups in sympatry. Here, we address the following three questions: (1) Are there genetic differences between the three genetic groups? (2) What are the selective sweeps found within each genetic group? (3) Can we distinguish different patterns of evolutionary history within western MB through the identified selective sweeps? To investigate these questions, we used target sequencing data from over 500 individuals as reported in Zunino *et al.* (2023), supplemented with sequencing data for unplaced scaffolds in Farga v2 reference genome (Julca *et al.* 2020) and sequencing data from *Olea europaea* subsp. *cuspidata* (Julca *et al.* 2020) that were used as an outgroup. Our analysis encompassed assessments of genetic diversity, population structure,

selective sweeps, and admixture across the genome, to shed light on the evolutionary history of each genetic group.

2 | Methods

2.1 | Sampling and sequencing

We analysed a total of 504 accessions including 359 naturally occurring olive trees *O. europaea* subsp. *europaea* var. *sylvestris* and 145 cultivated olive varieties *O. europaea* subsp. *europaea* var. *europaea*. The sampling and sequencing of accessions are described in Zunino *et al.* (2023). The naturally occurring populations came from 26 sites from western Mediterranean Basin, in France, Spain and Morocco. These sites contained 12 to 15 trees considered as wild or admixed olive trees as the result of hybridisation between wild and cultivated varieties as shown in (Zunino *et al.* 2023) (Figure 1). The analysed cultivated olive varieties, from the Worldwide Olive Germplasm Bank of Marrakech (WOGBM) are representative of the genetic diversity of olive resources in the western MB (Khadari *et al.* 2019). An additional set of 10 cultivated French varieties corresponding to varieties significantly present in the sampling area of natural populations were added. The olive trees were sequenced using a NovaSeq™ sequencing platform, with a target sequencing experiment targeting the 55,595 annotated genes of the reference genome ‘Farga’ v2 (Julca *et al.* 2020). In addition to 102,126 baits targeting the assembled chromosomes studied in Zunino *et al.* (2023), we retained the sequencing data targeted on scaffolds, which in total represent 210,367 (108,246 on scaffolds) baits with a captured length of 16.8Mb (8.2Mb on chromosomes and 8.7Mb on scaffolds). The raw sequence reads have been deposited in the European Nucleotide Archive (PRJEB61410; Zunino *et al.* 2023). We

CHAPTER 2

also add to the dataset, after sequencing, the whole genome data of *Olea europaea* subsp. *cuspidata* available from Julca *et al.* (2020) as an outgroup for the study.

2.2 | SNP filtering

The SNP calling as well as the SNP filtering were performed as described in Zunino *et al.* (2023). After sequencing, we obtained 89,322,843 genetic and structural variants. We filtered the raw data with VCFtools version 0.1.16 (Danecek *et al.* 2011). We removed indels, kept only SNPs with a quality superior to 200 and removed clusters of 3 SNPs in 10 bases. We kept only biallelic SNPs, with a minimum mean depth of 8 and maximum depth of 400. We removed sites with more than 15% missing data, then we removed individuals with more than 20% of missing data. Only sampling sites with at least 12 individuals were considered in this study. Filtering was carried out to exclude positions with fixed heterozygosity (>85%). At this step, we obtained 316,639 SNPs located on chromosomes and scaffolds.

We separated the data according to their assignment to an ancestral genetic cluster with sNMF from LEA packages version 3.11.3 (Frichot & François 2015). The genotypes with more than 70% assignment to the wild ancestral cluster defined by Zunino *et al.* (2023) are considered as genuine wild olive trees. Other genotypes, whether assigned to other groups or unassigned, which do not reach the 70% threshold are considered admixed. We obtained 135 genuine wild genotypes and 224 admixed genotypes (Supplementary Table S1). We separated the genotypes according to their belonging to groups and applied a minor allele count of one filter with VCFtools and we reapplied a filter to exclude positions with high heterozygosity (>85%). We obtained 187,979 SNPs for genuine wild olives, 191,565 SNPs for cultivated olives and 202,832 SNPs for admixed olives. We also generated a file containing the 504 genotypes which

was filtered with a minor allele count of one and removed high heterozygosity (>85%). It resulted 30,462 SNPs.

2.3 | Genetic diversity and genetic structure

Genetic measure was examined for each locus of each genetic group (wild, admixed and cultivated). We calculated descriptive statistics measures of diversity including expected heterozygosity (H_E), observed heterozygosity (H_O) and inbreeding coefficient (F_{IS}) using Hierfstat version 05-11 (Goudet & Jombart 2022). The significance of differences between groups were tested with Mann-Whitely and Kruskal-Wallis test. We computed the genetic differentiation (F_{ST}) of each loci for each pairs of genetic groups using the same package. The linkage disequilibrium decay was estimated for each genetic group using PopLDdecay version 3.42 (Zhang *et al.* 2018).

2.4 | Detecting traces of selection

The detection of selection traces was performed with PCAdapt 4.3.3 (Privé *et al.* 2020) on the three genetic groups: wild, admixed and cultivated. The number of principal components K was chosen according to the percentage of variance explained by each PC for each group. The significance threshold for outlier consideration was set using either a Bonferroni correction of the *p-values* (alpha = 0.05) or a false discovery rate (FDR = 0.1) approach. As our FDR set up appeared less conservative, the SNPs retained according to the FDR are considered as minor outliers, among which the SNPs exhibiting *p-value* above Bonferroni correction threshold are considered as major outliers.

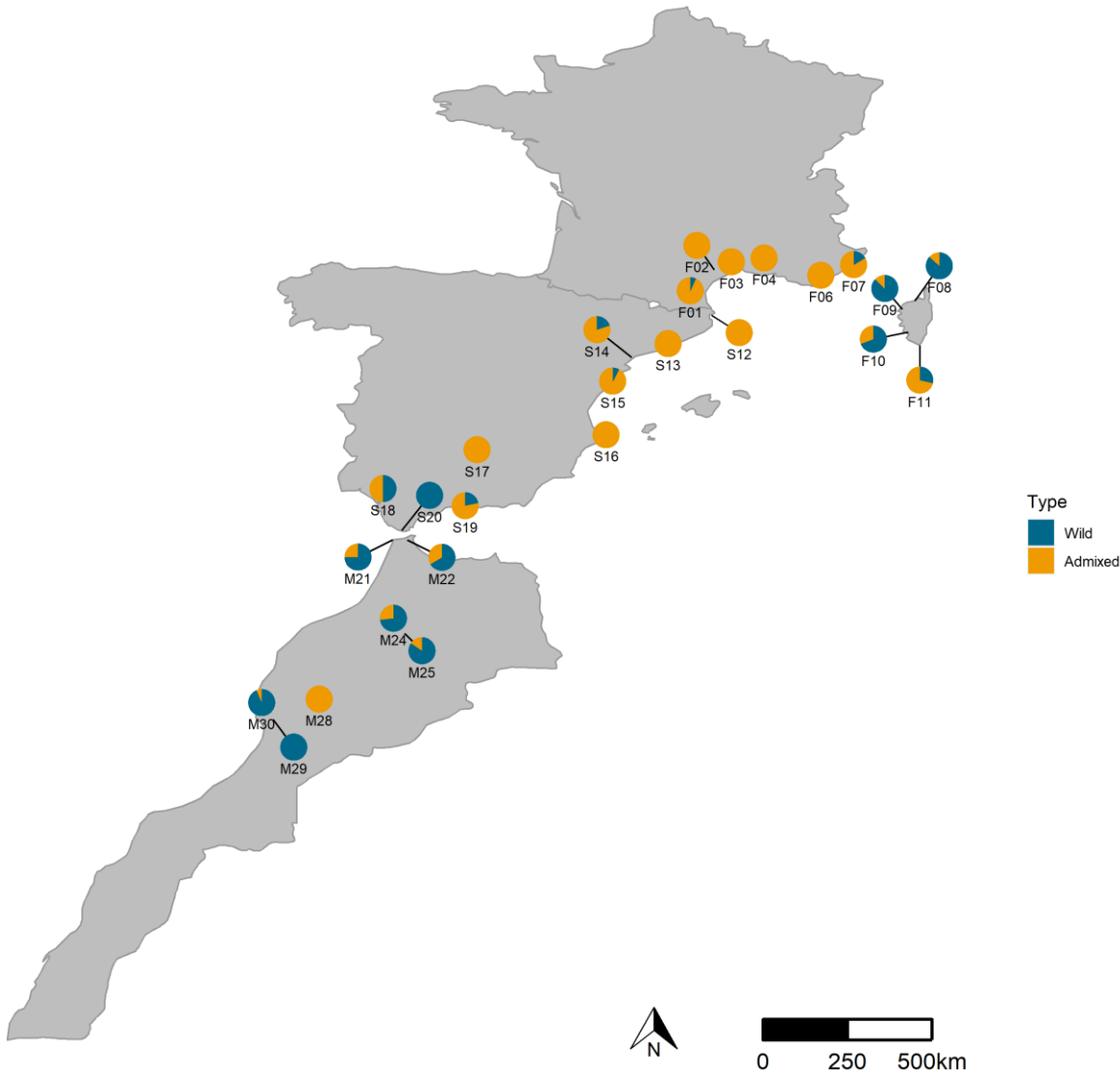


Figure 1. Geographical distribution of the naturally occurring western Mediterranean Basin of *O. europaea* with wild individuals (in blue) and admixed individuals (in yellow).

2.5 | Characterisation of outlier loci

We conducted an analysis of gene functional annotations within the genomic regions containing outlier loci. To identify these outliers within gene regions, we used the bed window function from bedtools version 2.30.0 (Quinlan & Hall 2010) to intersect them with the annotated reference genome Farga (Julca *et al.* 2020) with a window of 500 bp upstream and downstream. In cases where certain SNPs were associated with two genes due to the annotation on the reference genome, we chose to retain both gene assignments.

For each identified gene, we proceeded to determine their potential functions by assigning specific gene ontology (GO) terms and functions using GO.db version 3.16.0 (Carlson 2022) and AnnotationDbi version 1.60.2 (Pagès & Carlson 2023). Subsequently, we performed GO term enrichment tests on this dataset for each group of outliers examined, employing clusterProfiler version 4.6.2 (Wu *et al.* 2021). We then selected only those GO terms that displayed enrichment, characterised by an adjusted p-value below 0.05 and a q-value below 0.1. The outcomes of this analysis were visualised using enrichplot version 1.18.4 (Yu 2023).

SnEff version 4.3 was used to annotate and predict the effects of the outlier SNPs on genes and protein (Cingolani *et al.* 2012). We used the information to estimate the putative impact/deleteriousness of each outlier (high, modifier, moderate, low).

2.6 | Admixture patterns detection and potential geneflows

We utilised the previously calculated F_{ST} values to compute the difference, $F_{ST} CA - F_{ST} WA$, where $F_{ST} CA$ represents the differentiation between admixed and cultivated genotypes, and $F_{ST} WA$ represents the differentiation between admixed and wild

CHAPTER 2

genotypes. This difference is expected to exhibit a negative value in cases where gene flow occurs between admixed and wild genotypes. To assess whether the F_{ST} difference in SNPs detected as selected in admixed genotypes is significantly different than a random distribution, we randomly resampled 1,000 sets of the same size as the number of detected SNPs from the admixed genotype dataset. The significance was evaluated using a Wilcoxon-Mann-Whitney test.

To detect introgression patterns, we employed D-statistics (ABBA-BABA test) for each genetic group (wild, cultivated and admixed). This analysis was conducted exclusively on chromosomes dataset with *cuspidata* as an outgroup. Subsequently, we calculated the site frequency spectrum (SFS) for each SNP within each genotype group, using *cuspidata* as the reference allele. This approach allowed to compute the D frequency spectrum (D_{FS}), a metrics that enables the assessment of introgression signatures across the allele frequency spectrum (Martin & Amos 2020).

3 | Results

3.1 | Target sequencing characterization

The target sequencing efficiency, extended to scaffolds as compared to the data presented in Zunino *et al.* (2023), showed mainly SNPs in intergenic regions (more than 67% for all, wild, admixed or cultivated genotypes). SNPs present in gene exons represent more than 9.7% of the total, while the SNPs situated in introns represent more than 2.5%. Lastly, regions with the least SNPs are splice sites (acceptor and donor) with less than 0.02% SNPs of the dataset (Supplementary Figure S1).

3.2 | Genomic diversity, LD decay and genomic structure

The average inbreeding coefficient (F_{IS}) calculated on the different genetic groups was significantly different between all groups (Kruskal-Wallis, p -value < 0.001) (Figure 2). For the cultivated genotypes, the F_{IS} per loci is on average -0.06, with a median of -0.03. The group is almost at equilibrium with a small excess of heterozygotes. For the wild genotypes, the average F_{IS} per loci is 0.02, with a median of 0. It is almost at equilibrium with a small excess of homozygotes. For the admixed genotypes, the average F_{IS} is -0.013 with a median of -0.014. It is very close to equilibrium with a small excess of heterozygotes (Figure 2).

The measures of expected heterozygosity (H_S) and observed heterozygosity (H_O) for the three genetic groups are both significantly different between groups (Kruskal-Wallis – p -value < 0.001). For H_O the average is higher for admixed (0.1950) than cultivated (0.1828) and wild (0.1499) (Supplementary Figure S2). For H_S , the same pattern is observed with an average higher for admixed (0.1880) than cultivated (0.1595) and wild (0.1495) (Supplementary Figure S2).

The linkage disequilibrium estimation showed a rapid decay for each genetic group (Figure 3, Supplementary Figure S3). The admixed is dropping faster than wild and cultivated genotypes with at 0.2, mean R^2 of 200 bp for wild and cultivated and 40 bp for admixed group.

CHAPTER 2

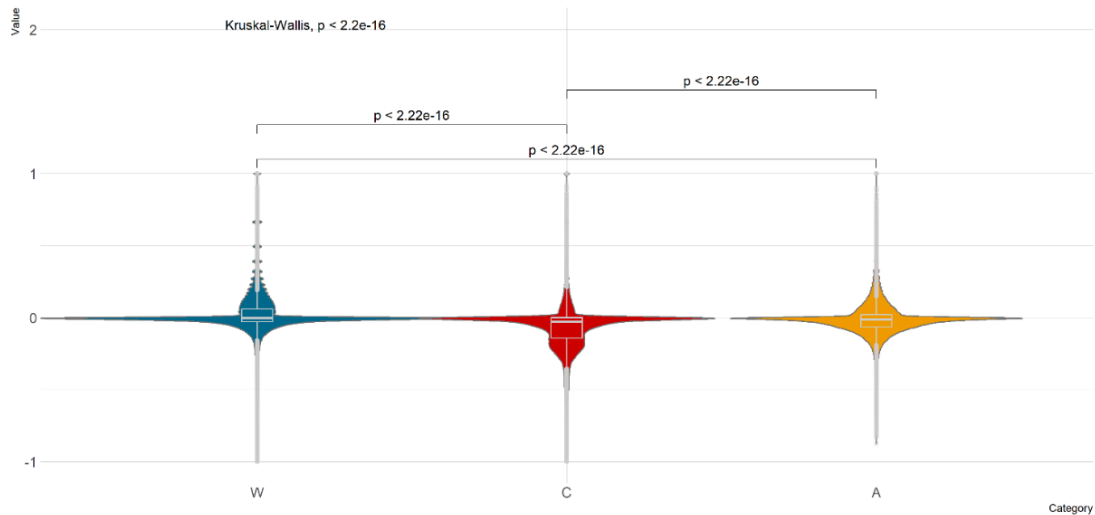


Figure 2. Measure of the inbreeding coefficient, F_{IS} , per loci of *O. europaea* for wild (W; in blue), cultivated (C; in red) and admixed genetic groups (A; in yellow). The comparisons between groups were statistically tested with the Mann-Whitney, with a significant difference between Wild and Admixed, Admixed and Cultivated, Wild and Cultivated F_{IS} . The distribution of all groups was tested using Kruskal-Wallis.

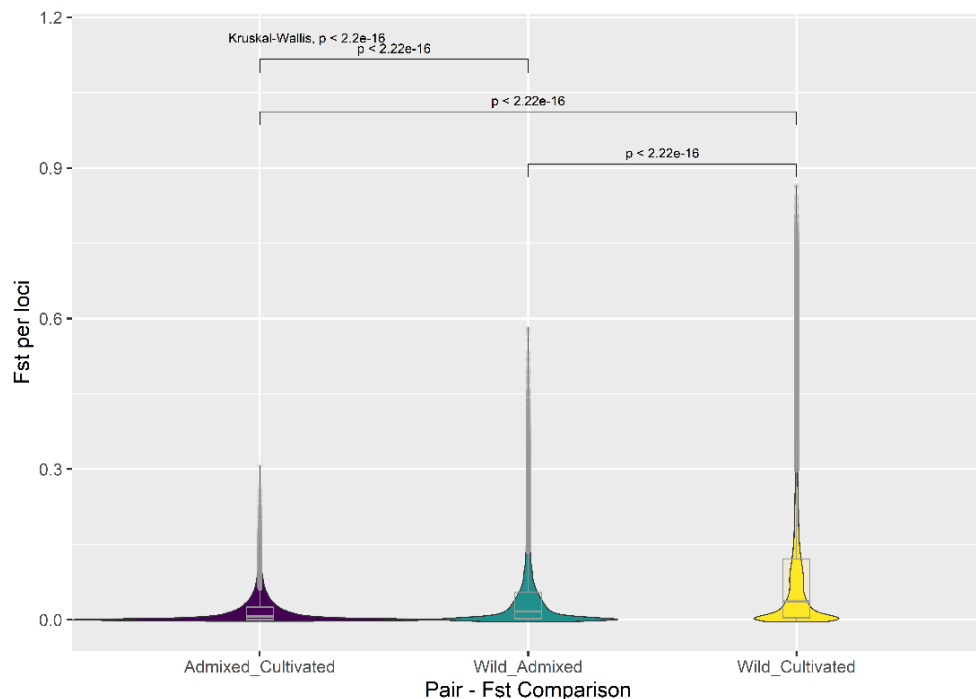


Figure 3. Pairwise locus-based F_{ST} between genotype groups of *Olea europaea*. The F_{ST} per locus for Admixed-Cultivated (in purple), Wild-Admixed (in green) and Wild-Cultivated (in yellow) distribution are reported as violin-plots. The distribution between each group were statistically tested with Mann-Whitney test. The distribution of all groups was tested using Kruskal-Wallis.

The pairwise F_{ST} estimation per loci was significantly different between all genetic group's comparison (Kruskal-Wallis, p -value < 0.001) (Figure 4). The differentiation between the wild genotypes and the cultivated genotypes is the highest with on average 0.097 (median of 0.036). The second highest was the differentiation between the wild genotypes and the admixed genotypes with on average 0.045 (median of 0.017). The lowest differentiation was observed between admixed genotypes and cultivated genotypes with a mean of 0.019 (median of 0.007).

The mean F_{ST} , calculated in a window size of 100 SNPs per chromosome, between admixed-cultivated and admixed-wild does not reveal any discernible pattern in the admixed genome of high differentiation. Cultivated exhibited lower differentiation compared to wild genotypes at the same loci within the admixed chromosome genomes on a global scale (Supplementary Figure S4).

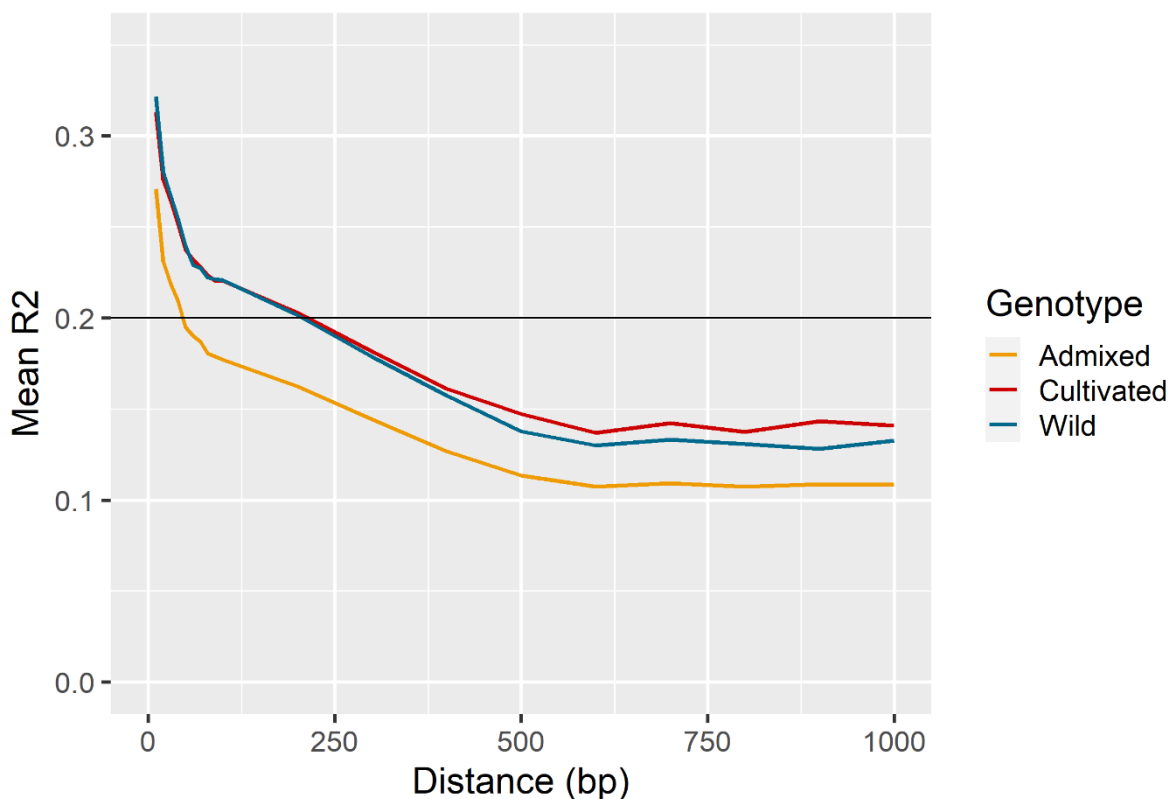


Figure 4. Linkage disequilibrium decay in different groups of *O. europaea* based on SNPs. It corresponds to values of mean r^2 according to distance in base pairs between

SNPs for three genotypic groups: wild (in blue), admixed (in yellow) and cultivated (in red).

3.3 | Detection of selection footprints

The footprints of selection analysis detected 2,102 unique SNPs inside the three genetic groups (Figure 5, Table 1). These results take into account the detection of minor selected SNPs using a *q-value* threshold of < 0.1 and significant major selected SNPs with the Bonferroni correction ($\alpha = 0.05$). The wild genetic group included the highest count of candidate SNPs for selection comprising a total of 1,168 SNPs, out of which 124 were identified as major outliers (Figure 5A, Table 1). The cultivated genetic group revealed 904 selected SNPs, including eight major outliers (Figure 5B, Table 1). Finally, the admixed genetic group exhibited 36 selected SNPs, of which four are considered as major outliers (Figure 5C, Table 1). No common SNPs were found across all genetic groups, and only three SNPs are shared between the wild genetic group and the cultivated genetic group (Figure 6).

Table 1. Number of candidate SNPs for selection resulting from PCAdapt analysis for admixed, cultivated and wild groups.

	Minor	Major	Total
Admixed	32	4	36
Cultivated	896	8	904
Wild	1044	124	1168

The minor outliers are detected above a threshold of *q-value* < 0.1 and the major outliers detected above *p-value adjusted* < 0.05 with Bonferroni correction ($\alpha = 0.05$).

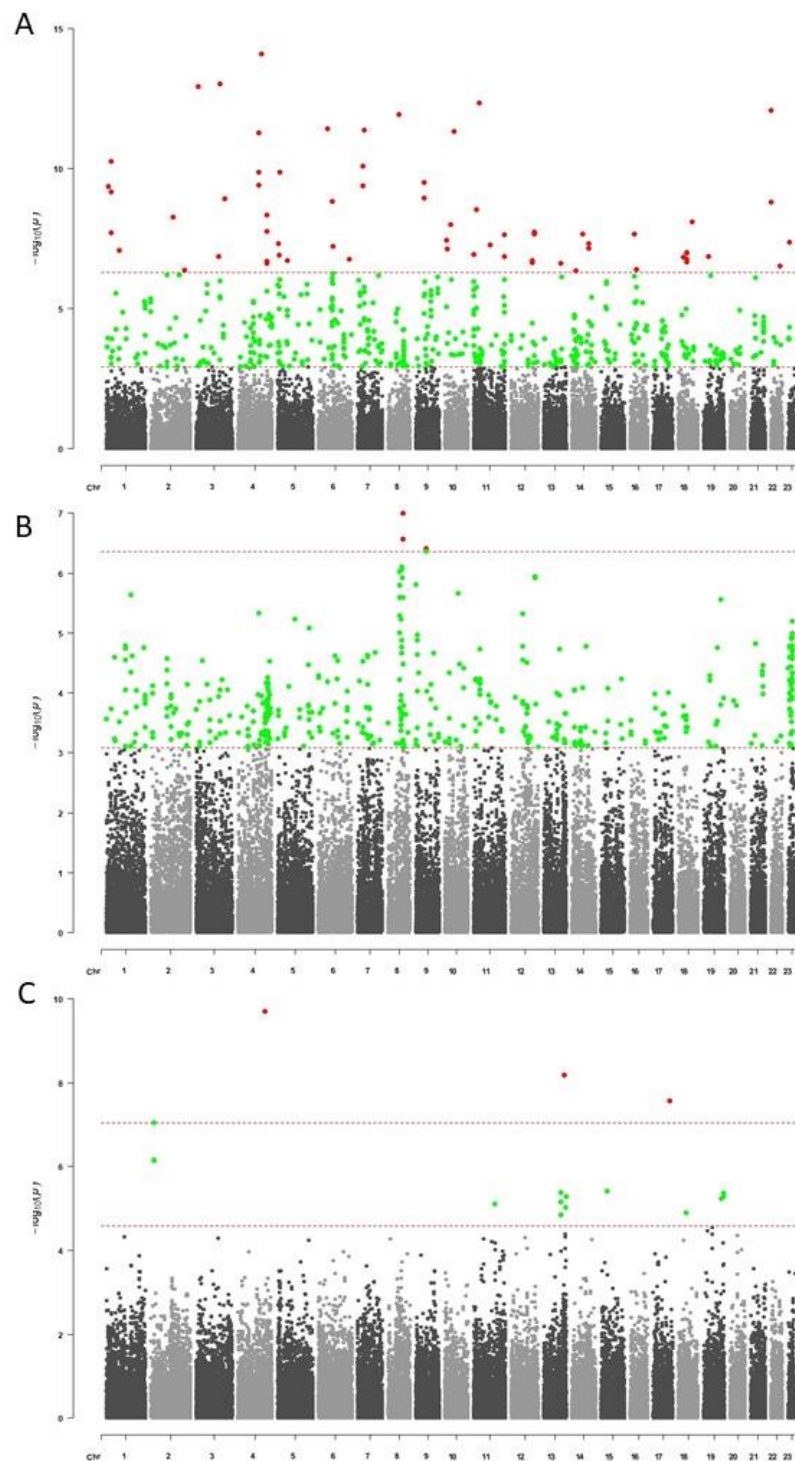


Figure 5. Manhattan plots of candidate SNPs on chromosomes for selection resulting from PCAdapt analysis for (A) wild, (B) cultivated and (C) admixed genetic groups. Minor outliers are detected above a threshold of q -value < 0.1 (in green) and the major outliers detected above p -value adjusted < 0.05 with Bonferroni correction ($\alpha = 0.05$; in red).

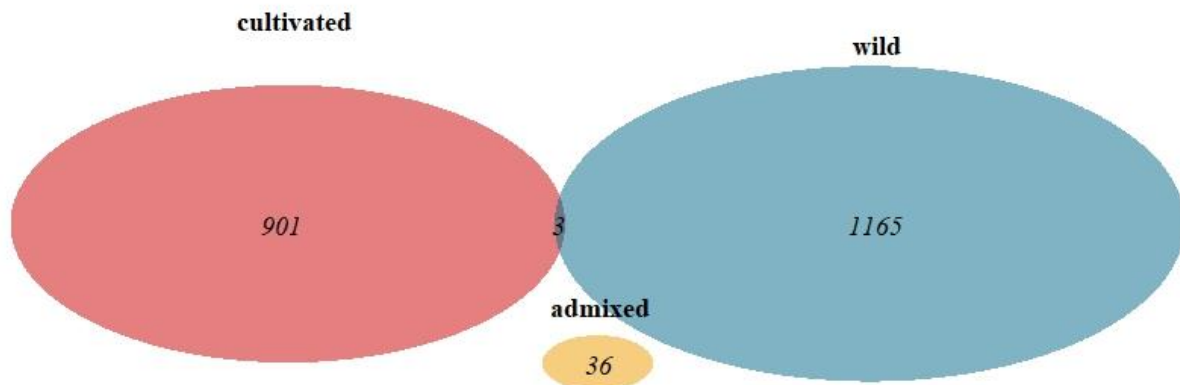


Figure 6. Venn diagram of candidate SNPs for selection resulting from PCAadapt analysis in each genetic group: wild in blue, cultivated in red and admixed in yellow.

3.4 | Biological significance of outlier SNPs

Among the 2,102 SNPs detected through the selection trace analysis within the three genetic groups, 1,955 are linked to 1,308 distinct genes. Some SNPs are found in multiple genes, because of the structural genome annotation of the reference genome "Farga" (Julca *et al.* 2020). Using AnnotationDbi, 1043 of these genes have been functionally annotated (Supplementary Table S2). The enrichment analysis demonstrates statistical significance exclusively within the dataset of admixed individuals. Out of the 36 detected SNPs in this dataset, a substantial 30 are considered significantly enriched within our dataset (p -value <0.5, q -value <0.1). For the 30 enriched GO, 45.9% are involved in Biological Process (BP), 37.8% in Molecular Function (MF) and 16.2% in Cellular Component (CC). The five most significant GO terms (p -value <0.006) are in order GO:0048226, GO:0046274, GO:0052716, GO:0008878 and GO:0005507. These terms are encountered at least twice within the dataset, except for GO:0008878, which is encountered once. They are respectively found to be involved for *Arabidopsis thaliana* in "Casparian strip", "lignin

catabolic process”, “hydroquinone:oxygen oxidoreductase activity”, “glucose-1-phosphate adenylyltransferase activity” and “copper ion binding” (Figure 7).

Using SNPeff, 89% of the detected outliers for wild genotypes were predicted to have a modifier impact (e.g. downstream gene variant or exon variant), as well as 91% for the cultivated outliers and 95% for the admixed outliers. The only instance of a high-impact prediction was found among wild outliers, with one SNP resulting in a ‘stop gained’ prediction, associated with a function related to microtubule cytoskeleton organisation (Supplementary Table S3).

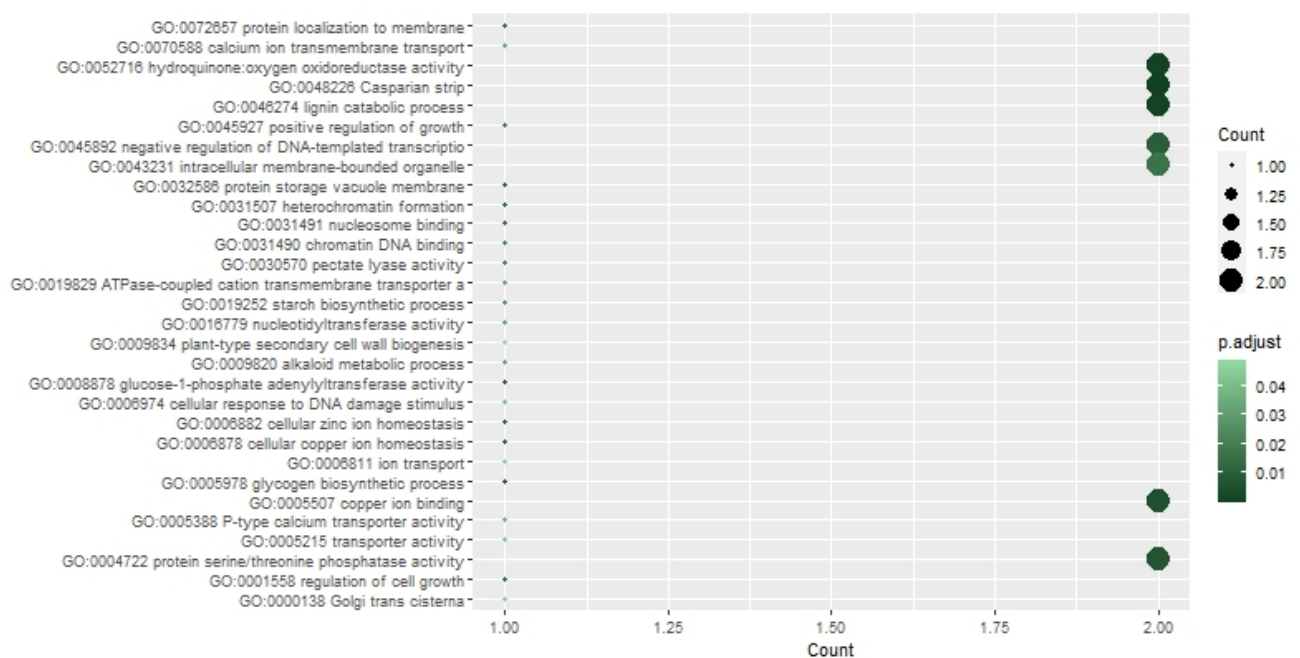


Figure 7. Gene ontology (GO) term enrichment analysis based on candidate SNPs for selection resulting from PCAdapt analysis for admixed olive trees. Out of the 36 detected candidate SNPs, 30 are considered significantly enriched (p -value <0.5 , q -value <0.1).

when P1 = wild or cultivated, P2 = admixed and P3 = wild or cultivated meaning an ABBA pattern indicating an incomplete lineage sorting (ILS) or introgression process between wild and cultivated. D is on average 0 when P1 and P2 correspond to wild or cultivated and P3 = admixed, leading to an ILS conclusion for these genotypes (Table 2).

Table 2. Results of potential introgression (excess of alleles) between wild, admixed and cultivated genetic groups in the western Mediterranean basin, using Patterson's D (or D statistic) test. *O. cuspidata* was used as an outgroup.

Conditions	D statistic
((W,A),C),CUS)	0.287
((C,A),W),CUS)	0.296
((C,W),A),CUS)	-0.001
((W,C),A),CUS)	0.001

The order corresponds to ((P1, P2),P3),Outgroup) with P1: population 1, P2: population 2, P3: population 3 and an outgroup. W: wild, A: Admixed and C: cultivated.

The D_{FS} was tested under three possible scenarios (Figure 9): CAW (P1 = cultivated, P2 = admixed, P3 = wild and outgroup = cuspidata), WAC (P1 = wild, P2 = admixed, P3 = cultivated and outgroup = cuspidata) and CWA (P1 = cultivated, P2 = wild, P3 = admixed and outgroup = cuspidata). Under the CAW scenario, D_{FS} displayed a mean D positive. More specifically, it exhibited a negative D at low derived allele frequencies, followed by a "round" pattern of positive D at intermediate frequencies. This pattern suggests a possible ancient bottleneck in the admixed genotypes, followed by subsequent introgression events from the wild genotypes, with an additive effect of smaller gene flow from wild genotypes to cultivated genotypes (Figure 9A). The

CHAPTER 2

possible scenario is presented in Figure 9D. For the scenario CWA, the D_{FS} demonstrated a similar profile to CAW, indicating a potential ancient bottleneck for the admixed genotypes. However, in this scenario, gene flow appears to occur from the cultivated to the admixed population, along with a smaller gene flow from the cultivated to the wild genotypes (Figure 9B), as illustrated in Figure 9E. The third scenario, CWA, yielded a D value of 0, suggesting a potential bidirectional gene flow between genetic groups, without a distinct introgression pattern or bottleneck. This scenario presents a complex signal that makes it difficult to interpret (Figure 9C).

4 | Discussion

We used target sequencing to perform a genomic scan for selection of different genetic groups within olive trees *O. europaea* subsp. *europaea* in the western Mediterranean Basin (MB). These groups included cultivated, wild, and admixed genotypes. Our dataset, encompassing over 16 Mb of the genome, provides information about genetic diversity, genetic structure, selective sweep events and admixture dynamics among these genetic groups. Here, we were able to infer the outline of the evolutionary history of each group within the western MB. Our findings revealed distinct ancient evolutionary histories for cultivated and wild genotypes, supported by pronounced differences in footprints of selection. These differences are likely linked to environmental selection and genetic drift for both groups, with different environmental conditions, with maybe the participation of domestication in the case of cultivated trees.

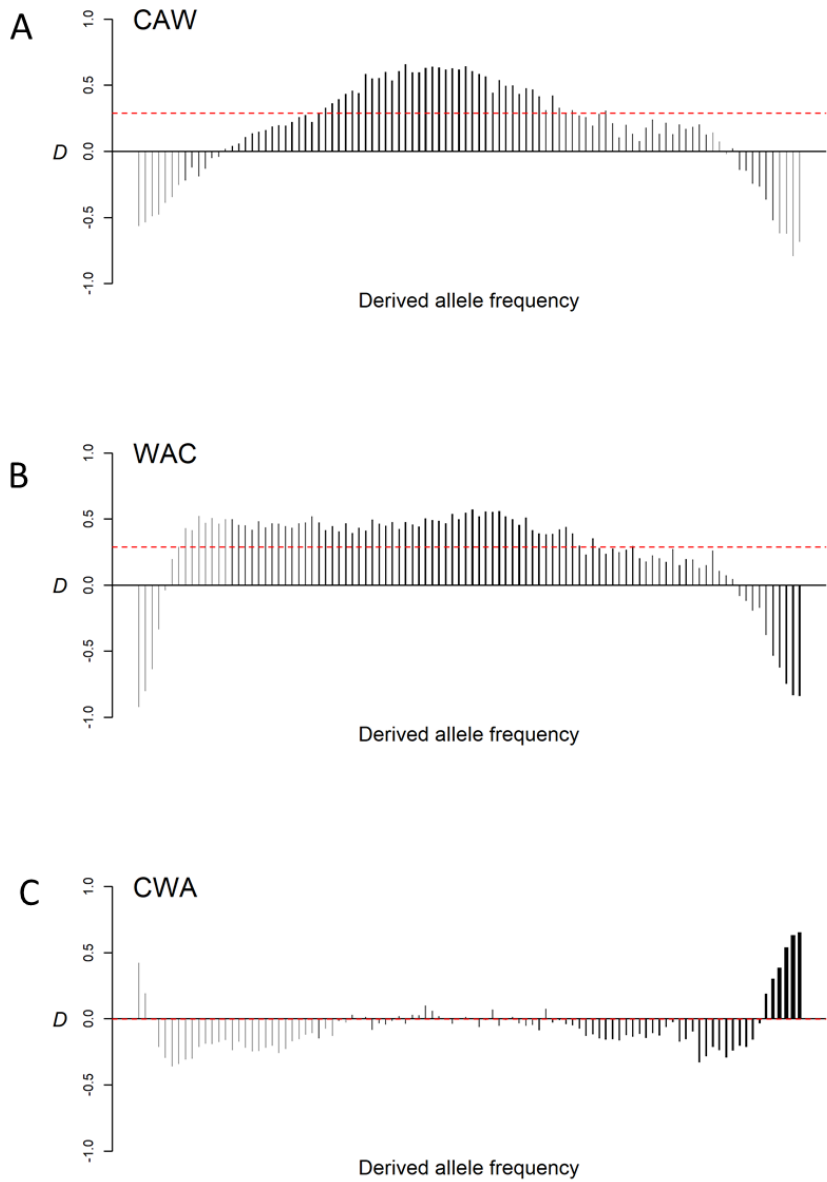
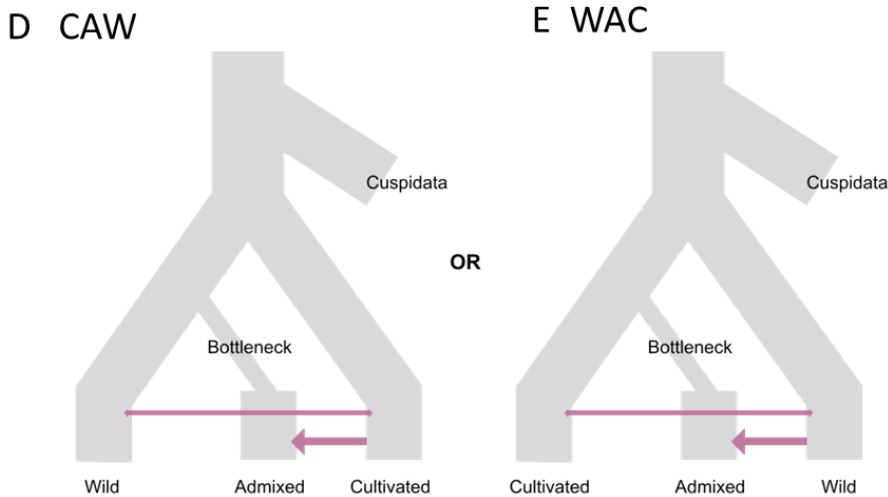


Figure 9. D_{FS} results of the three genotypes of *O. europaea* and scenario of evolutionary history suggested by D_{FS} results. (A) First scenario CAW P1 = cultivated, P2 = admixed, P3 = wild and outgroup = cuspidata (B) WAC P1 = wild, P2 = admixed, P3 = cultivated and outgroup = cuspidata (C) CWA P1 = cultivated, P2 = wild, P3 = admixed and outgroup = cuspidata. (D) hypotheses of demographic history under CAW scenario and (E) hypotheses of demographic history under WAC scenario.



CHAPTER 2

These conclusions align with the evolutionary histories of wild and cultivated genotypes as described in previous studies (Besnard *et al.* 2018; Gros-Balthazard *et al.* 2019; Julca *et al.* 2020). In contrast, admixed genotypes, originating from hybridization events between wild and cultivated genotypes in the western MB, appear to have undergone an ancient bottleneck followed by substantial gene flow. We cannot definitively ascertain whether they descended directly from wild genotypes with gene flow from cultivated or the opposite scenario, and the bottleneck may be detected from their parent evolutionary history than just the admixed evolutionary history. However, our analysis suggests a relatively recent history characterised by limited selection pressure, as indicated by the small number of positively selected SNPs detected.

4.1 | Three genetic patterns (wild, domesticated and admixed) and three evolutionary trajectories in the western Mediterranean Basin

The genetic differences observed among genetic groups revealed distinct patterns. The wild genetic group exhibited a reduction in genetic diversity with an excess of homozygotes, in contrast to the cultivated and admixed groups. This slightly lower genetic diversity from the wild group than from the cultivated one contrasts with the findings of a prior study (Gros-Balthazard *et al.* 2019). The slightly lower genetic diversity observed in wild genotypes in comparison to cultivated and admixed ones may be attributed to several factors. Firstly, an excess of heterozygous in cultivated olive because of the genetic heterogeneous of the group with possible gene flow coming from the natural populations from several place of the MB (Lavee & Zohary 2011). Secondly, the extensive distribution of cultivated olive varieties across the gradient, leading to a homogenization of allelic frequencies due to pervasive gene flow. Thirdly, ancient widespread destruction of the natural environments of olive trees,

contributing to the disappearance of wild populations (Lumaret *et al.* 1997, 2004; Breton *et al.* 2006; Lavee & Zohary 2011). This dynamics has given rise to numerous new populations that have either escaped from cultivation or have become admixed with wild varieties (Lumaret *et al.* 1997, 2004; Breton *et al.* 2006; Lavee & Zohary 2011). These combined factors may have led to the present genetic diversity wild oleaster populations in the western MB.

For both cultivated and admixed genotypes, the excess of heterozygotes suggests an important gene flow. This can be explained by the dataset, which account for two different cultivated lineages present in the range, Q2 (central cultivated) and Q3 (western cultivated) (Besnard *et al.* 2013b; a; Diez *et al.* 2015; Zunino *et al.* 2023), as well as the admixed populations resulting from hybridization with these lineages (Zunino *et al.* 2023). Additionally, for the admixed genotypes, hybridization leads to an excess of heterozygosity.

The genetic differentiation analysis revealed that admixed genotypes are genetically closer to cultivated genotypes. There are three potential reasons for this proximity: 1) a potential direct descendance from the cultivated genotypes with gene flow from the wild, 2) an origin from the wild with extensive gene flow from cultivated genotypes or 3) a mix of the two first reason with some admixed coming from cultivated and other coming from wild, followed by extensive gene flow (Lavee & Zohary 2011). The admixed genotypes exhibit more differentiation with wild than with cultivated ones. While the extensive cultivation of olive trees and the importance of pollen may support the second scenario, at present, there is no conclusive evidence to favour one scenario over the other.

4.2 | Selection footprints as a genetic signature of olive: supportive information of evolutive history?

The footprints of positive selection within the different compartments are very different. Selection footprints are mostly detected in wild (1,168 outliers) then domesticated olive trees (904), while for admixed gene pool we observed very few outliers (36) indicating a weak signal of positive selection. However, these results could be explained by the approach we used to detect the footprints of positive selection, which is less informative for admixed populations due to their allele frequency and complicated structure (de Villemereuil *et al.* 2014; Rellstab *et al.* 2015).

By detecting 1168 outliers including 124 major outliers, genuine wild olive in the western Mediterranean Basin has experienced a positive selection indicating a genetic signature of its known evolutionary history. The wild olive trees have arrived in the MB between 5.59 to 5.33 million years ago, in a significant expansion wave in southern Europe (Planchais 1984; Suc 1984; Terral *et al.* 2004, 2009; Carrión *et al.* 2010; Langgut *et al.* 2019). The populations were then affected by multiple glaciation events leading to population fragmentation and loss (Carrión *et al.* 2010; Langgut *et al.* 2019) and potential fixation of alleles within the wild genotypes restricted to glacial refuges because of increasing genetic drift, leading to differentiation between populations in distant refuges. The conclusion of the glaciation periods, coupled with a highly favourable environment for recolonization, has led to the extension and potential reconnection between populations originating from different refuges (Planchais 1984; Terral *et al.* 2009; Langgut *et al.* 2019). This process may have the potential to mitigate the effect of the genetic drift fixation within the wild genotypes. Wild olives have established a long-term presence in their current environment.

The cultivated olive tree has a more recent history with at least one domestication event occurring approximately 6,000 years ago in the eastern MB (Zohary & Spiegel-roy 1975; Kaniewski *et al.* 2012; Newton *et al.* 2014). In our studied area, the cultivated history is even more recent, marked by at least two probable events of introduction by humans. The first introduction may have taken place in the south Iberian Peninsula around 2,500 BP (Langgut *et al.* 2019), while the second introduction would have occurred in France, likely with the Romans, in the 1st century BC (Leveau *et al.* 1991). The selection footprints detected may be the result of several process, the evolutionary history and different environment pressure from before the domestication in the Levant, in addition with some selection footprints of artificial selection by human during domestication. The detection of slightly fewer associated SNPs than wild group, may be attributed to two factors. First, it could be a result of genetic group heterogeneity, as seen in the admixed group, with the presence of diverse ancestral genetic backgrounds. Second, it might be influenced by multiple gene flows occurring between the cultivated and the naturally occurring populations of the MB.

For the admixed genotypes, the limited evidence of selection suggests a very recent history that likely started with admixture events between the wild and the cultivated (Zunino *et al.* 2023). The absence of SNPs detected as being under selection in either wild or cultivated genotypes implies that the selective pressures on these groups may no longer be present in the current natural environment or these groups have not yet been strongly selected by these selective pressures. There is no evidence of preferential selection of SNPs during hybridization events, even though the majority of detected selected SNPs originate from cultivated, which is consistent with their closer genetic similarities with the cultivated. The study of the admixture patterns and D_{FS} has provided insights about the demographic history of the admixed group, suggesting

CHAPTER 2

two potential scenarios. One scenario involves direct descent from cultivated genotypes and extensive gene flow from the wild, while the other scenario proposed direct descent from the wild and extensive gene flow from cultivated genotypes. Currently, we cannot argue in favour of one scenario or the other. A very important result is the proportion of wild genetic material within admixed genotypes remains important, especially when compared to the limited presence of wild populations, denoted particularly in Spain. This could be attributed to either an extensive gene flow from the admixed populations or to a retention of wild genome parts within the admixed genotypes. Our tests were not conclusive regarding this question. Nonetheless, the examination of adaptive introgression is currently a widespread practice, and there is potential for further analyses on admixed group (Burgarella *et al.* 2019).

The selection footprints we are detecting could potentially be associated with local adaptation within the genetic groups. However, as explained by (Capblancq *et al.* 2018), it does work if the environmental gradient related to adaptation is correlated to population structure. The complex evolutionary histories of each group make it challenging to ascertain the presence of such correlation. Therefore, it might be more appropriate to consider the use of genomic environment association analysis to validate the potential adaptive meaning of these SNPs (Capblancq *et al.* 2018).

5 | Conclusion

We evidenced distinct footprints of selection in cultivated, wild and admixed olive trees in the western Mediterranean Basin, which provides some assumption about their evolutionary history within this region. The positive selection found within the wild and the cultivated are the results of a genetic signature of their ancient evolutionary history within the Mediterranean Basin. Our investigation raises questions concerning the

demographic history of wild, admixed and cultivated genetic groups. To address these questions, coalescent modelling or diffusion approximation methods such as, respectively, FastSimCoal (Excoffier *et al.* 2021) or Moments (Jouganous *et al.* 2017) could be used. Such modelling can help elucidate which scenario is the most probable and provide further insights into the population dynamics of both western wild olives and cultivated olives.

References

- Besnard, G., Bakkali, A. E., Haouane, H., Baali-Cherif, D., Moukhli, A., & Khadari, B. (2013). Population genetics of Mediterranean and Saharan olives: Geographic patterns of differentiation and evidence for early generations of admixture. *Annals of Botany*, *112*(7), 1293–1302.
- Besnard, G., Khadari, B., Navascués, M., Fernández-Mazuecos, M., Bakkali, A. E., Arrigo, N., Baali-Cherif, D., Brunini-Bronzini de Caraffa, V., Santoni, S., Vargas, P., & Savolainen, V. (2013). The complex history of the olive tree: From late quaternary diversification of mediterranean lineages to primary domestication in the northern Levant. *Proceedings of the Royal Society B: Biological Sciences*, *280*, 1–7.
- Besnard, G., Terral, J. F., & Cornille, A. (2018). On the origins and domestication of the olive: A review and perspectives. *Annals of Botany*, *121*(3), 385–403.
- Breton, C., Tersac, M., & Bervillé, A. (2006). Genetic diversity and gene flow between the wild olive (oleaster, *Olea europaea* L.) and the olive: Several Plio-Pleistocene refuge zones in the Mediterranean basin suggested by simple sequence repeats analysis. *Journal of Biogeography*, *33*(11), 1916–1928.
- Burgarella, C., Barnaud, A., Kane, N. A., Jankowski, F., Scarcelli, N., Billot, C., Vigouroux, Y., & Berthouly-Salazar, C. (2019). Adaptive Introgression: An Untapped Evolutionary Mechanism for Crop Adaptation. *Frontiers in Plant Science*, *10*, 4.
- Capblancq, T., Lachmuth, S., Fitzpatrick, M. C., & Keller, S. R. (2023). From common gardens to candidate genes: exploring local adaptation to climate in red spruce. *The New Phytologist*, *237*(5), 1590–1605.
- Capblancq, T., Luu, K., Blum, M. G. B., & Bazin, E. (2018). Evaluation of redundancy analysis to identify signatures of local adaptation. *Molecular Ecology Resources*, *18*(6), 1223–1233.
- Carlson, M. (2022). *GO.db: A set of annotation maps describing the entire Gene Ontology*.
- Carrión, Y., Ntinou, M., & Badal, E. (2010). *Olea europaea* L. in the North Mediterranean Basin during the Pleniglacial and the Early-Middle Holocene. *Quaternary Science Reviews*, *29*(7–8), 952–968.
- Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., & Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly*, *6*(2), 80–92.
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158.
- De Villemereuil, P., Frichot, É., Bazin, É., François, O., & Gaggiotti, O. E. (2014). Genome scan methods against more complex models: When and how much

- should we trust them? *Molecular Ecology*, 23(8), 2006–2019.
- Diez, C. M., Trujillo, I., Martínez-Urdiroz, N., Barranco, D., Rallo, L., Marfil, P., & Gaut, B. S. (2015). Olive domestication and diversification in the Mediterranean Basin. *The New Phytologist*, 206(1), 436–447.
- Duan, N., Bai, Y., Sun, H., Wang, N., Ma, Y., Li, M., Wang, X., Jiao, C., Legall, N., Mao, L., Wan, S., Wang, K., He, T., Feng, S., Zhang, Z., Mao, Z., Shen, X., Chen, X., Jiang, Y., ... Chen, X. (2017). Genome re-sequencing reveals the history of apple and supports a two-stage model for fruit enlargement. *Nature Communications*, 8(1). <https://doi.org/10.1038/s41467-017-00336-7>
- Excoffier, L., Marchi, N., Marques, D. A., Matthey-Doret, R., Gouy, A., & Sousa, V. C. (2021). fastsimcoal2: demographic inference under complex evolutionary scenarios. *Bioinformatics*, 37(24), 4882–4885.
- Frichot, E., & François, O. (2015). LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution / British Ecological Society*, 6(8), 925–929.
- Goudet, J., & Jombart, T. (2022). *Hierfstat: Estimation and tests of hierarchical F-statistics*. R package version 0.5-11. <https://CRAN.R-project.org/package=hierfstat>.
- Gros-Balthazard, M., Besnard, G., Sarah, G., Holtz, Y., Leclercq, J., Santoni, S., Wegmann, D., Glémin, S., & Khadari, B. (2019). Evolutionary transcriptomics reveals the origins of olives and the genomic changes associated with their domestication. *The Plant Journal: For Cell and Molecular Biology*, 100(1), 143–157.
- Jordan, R., Hoffmann, A. A., Dillon, S. K., & Prober, S. M. (2017). Evidence of genomic adaptation to climate in *Eucalyptus microcarpa*: Implications for adaptive potential to projected climate change. *Molecular Ecology*, 26(21), 6002–6020.
- Jouganous, J., Long, W., Ragsdale, A. P., & Gravel, S. (2017). Inferring the joint demographic history of multiple populations: Beyond the diffusion approximation. *Genetics*, 206(3), 1549–1567.
- Julca, I., Marcet-Houben, M., Cruz, F., Gómez-Garrido, J., Gaut, B. S., Díez, C. M., Gut, I. G., Alioto, T. S., Vargas, P., & Gabaldón, T. (2020). Genomic evidence for recurrent genetic admixture during the domestication of Mediterranean olive trees (*Olea europaea* L.). *BMC Biology*, 18(1), 1–25.
- Kaniewski, D., Van Campo, E., Boiy, T., Terral, J. F., Khadari, B., & Besnard, G. (2012). Primary domestication and early uses of the emblematic olive tree: Palaeobotanical, historical and molecular evidence from the Middle East. *Biological Reviews of the Cambridge Philosophical Society*, 87(4), 885–899.
- Khadari, B., & El Bakkali, A. (2018). Primary Selection and Secondary Diversification: Two Key Processes in the History of Olive Domestication. *International Journal of Agronomy*, 2018, 1–9.
- Khadari, B., El Bakkali, A., Essalouh, L., Tollon, C., Pinatel, C., & Besnard, G. (2019). Cultivated Olive Diversification at Local and Regional Scales: Evidence From the Genetic Characterization of French Genetic Resources. *Frontiers in Plant Science*, 10, 1593.

CHAPTER 2

- Langgut, D., Cheddadi, R., Carrión, J. S., Cavanagh, M., Colombaroli, D., Eastwood, W. J., Greenberg, R., Litt, T., Mercuri, A. M., Miebach, A., Roberts, C. N., Woldring, H., & Woodbridge, J. (2019). The origin and spread of olive cultivation in the Mediterranean Basin: The fossil pollen evidence. *Holocene*, 29(5), 902–922.
- Lavee, S., & Zohary, D. (2011). The potential of genetic diversity and the effect of geographically isolated resources in olive breeding. *Israel Journal of Plant Sciences*, 59(1), 3–13.
- Leveau, P., Heinz, C., Laval, H., Marival, P., & Medus, J. (1991). Les origines de l'oléiculture en Gaule du Sud [Données historiques, archéologiques et botaniques]. *Revue d'archéométrie*, 15(1), 83–94.
- Lumaret, R., Ouazzani, N., Michaud, H., & Villemur, P. (1997). Cultivated olive and oleaster: Two closely connected partners of the same species (*Olea europaea* L.): Evidence from allozyme polymorphism. *Boccone*, 7, 39–42.
- Lumaret, R., Ouazzani, N., Michaud, H., Vivier, G., Deguilloux, M.-F., & Di Giusto, F. (2004). Allozyme variation of oleaster populations (wild olive tree) (*Olea europaea* L.) in the Mediterranean Basin. *Heredity*, 92(4), 343–351.
- Lumaret, Roselyne, & Ouazzani, N. (2001). Ancient wild olives in Mediterranean forests. *Nature*, 413(6857), 700–700.
- Martin, S. H., & Amos, W. (2020). Signatures of Introgression across the Allele Frequency Spectrum. *Molecular Biology and Evolution*, 38(2), 716–726.
- Mitchell-Olds, T., Willis, J. H., & Goldstein, D. B. (2007). Which evolutionary processes influence natural genetic variation for phenotypic traits? *Nature Reviews. Genetics*, 8(11), 845–856.
- Neale, D. B., & Kremer, A. (2011). Forest tree genomics: growing resources and applications. *Nature Reviews. Genetics*, 12(2), 111–122.
- Newton, C., Lorre, C., Sauvage, C., Ivorra, S., & Terral, J. F. (2014). On the origins and spread of *Olea europaea* L. (olive) domestication: Evidence for shape variation of olive stones at Ugarit, Late Bronze Age, Syria—a window on the Mediterranean Basin and on the westward diffusion of olive varieties. *Vegetation History and Archaeobotany*, 23(5), 567–575.
- Pagès, H., & Carlson, M. (2023). *AnnotationDbi: Manipulation of SQLite-based annotations in Bioconductor (1.60.2)* [Computer software]. <https://bioconductor.org/packages/AnnotationDbi>
- Planchais, N. (1984). Palynologie lagunaire: l'exemple du Languedoc-Roussillon. *Annales de Géographie*, 93(516), 268–275.
- Privé, F., Luu, K., Vilhjálmsson, B. J., & Blum, M. G. B. (2020). Performing Highly Efficient Genome Scans for Local Adaptation with R Package pcadapt Version 4. *Molecular Biology and Evolution*, 37(7), 2153–2154.
- Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6), 841–842.
- Rellstab, C., Gugerli, F., Eckert, A. J., Hancock, A. M., & Holderegger, R. (2015). A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology*, 24(17), 4348–4370.

- Rubio de Casas, R., Besnard, G., Schönswetter, P., Balaguer, L., & Vargas, P. (2006). Extensive gene flow blurs phylogeographic but not phylogenetic signal in *Olea europaea* L. *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik*, 113(4), 575–583.
- Steane, D. A., Potts, B. M., McLean, E., Prober, S. M., Stock, W. D., Vaillancourt, R. E., & Byrne, M. (2014). Genome-wide scans detect adaptation to aridity in a widespread forest tree species. *Molecular Ecology*, 23(10), 2500–2513.
- Suc, J.-P. (1984). Origin and evolution of the Mediterranean vegetation and climate in Europe. *Nature*, 307, 429–432.
- Terral, Alonso, N., Buxó I Capdevila, R., Chatti, N., Fabre, L., Fiorentino, G., Marival, P., Jordá, G. P., Pradat, B., Rovira, N., & Alibert, P. (2004). Historical biogeography of olive domestication (*Olea europaea* L.) as revealed by geometrical morphometry applied to biological and archaeological material. *Journal of Biogeography*, 31(1), 63–77.
- Terral, J.-F., Durand, A., Newton, C., & Ivorra, S. (2009). Archéo-biologie de la domestication de l'olivier en Méditerranée occidentale : de la remise en cause d'une histoire dogmatique à la révélation de son irrigation médiévale. *Etudes*, 1–19.
- Wright, S. I., Bi, I. V., Schroeder, S. G., Yamasaki, M., Doebley, J. F., McMullen, M. D., & Gaut, B. S. (2005). The effects of artificial selection on the maize genome. *Science*, 308(5726), 1310–1314.
- Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., Fu, X., Liu, S., Bo, X., & Yu, G. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Cambridge (Mass.))*, 2(3), 100141.
- Yu, G. (2023). *enrichplot: Visualization of Functional Enrichment Result*. <https://yulab-smu.top/biomedical-knowledge-mining-book/>
- Zhang, C., Dong, S.-S., Xu, J.-Y., He, W.-M., & Yang, T.-L. (2018). PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics*, 35(10), 1786–1788.
- Zohary, D., & Spiegel-roy, P. (1975). Beginnings of Fruit Growing in the Old World. *American Association for the Advancement of Science*, 187(4174), 319–327.
- Zunino, L., Cubry, P., Sarah, G., Mournet, P., El Bakkali, A., Aqbouch, L., Costes, E., & Khadari, B. (2023). Genomic evidence of genuine wild versus admixed olive populations evolving in the same natural environments in western Mediterranean Basin. In *bioRxiv* (p. 2023.04.18.537296). <https://doi.org/10.1101/2023.04.18.537296>

Chapter 3



Does crop-to-wild gene flow lead to maladaptation of perennial species to their natural environment? Insights from the study of local adaptation of wild olive trees to their environment.

In preparation

Lison Zunino et al

Abstract:

Today, organisms face significant threats due to substantial alterations in their environments, primarily driven by human-mediated global changes. For instance, within the Mediterranean Basin (MB), the average temperature is rising 20% faster than in the rest of the world, which is predicted to raise of at least of 1.5°C. Climate change presents a substantial risk to the survival and distribution of species in this region. Understanding how species adapt in response to its environment can significantly enhance species conservation efforts. Nowadays, the advent of new sequencing methods has provided access to genetic information of species, and the use of genome-environment association make feasible to investigate local adaptation of species using their genome. Here, we investigated the local adaptation of wild olive trees (*Olea europaea* subsp. *europaea* var. *sylvestris*, n = 135) and admixed olive trees (resulting from hybridization between wild and cultivated olive trees, n = 224) to their environments. We used previously generated target sequencing data in combination with climatic and elevation data retrieved from publicly available databases. The study includes 26 populations sampled along a north-south gradient of 13 degrees, covering regions in Southern France, Corsica, Spain, and Morocco. Our analysis revealed 1,400 SNPs associated with climate and elevation for the wild olive group and 1,343 SNPs for the admixed group, with relatively few candidate SNPs being shared between the two groups. Wild olive trees exhibit local adaptation to their environments, with temperature-related variables being the primary selective factor. Conversely, the admixed olive trees do not seem to display strong local adaptation, and the detected signal may be influenced by the genetic heterogeneity and strong population structure within this group. Nevertheless, admixed individuals still retain approximately 80% of the wild candidate SNPs in their genome. This research has the potential to help predicting the future adaptive potential of natural olive tree populations in the western MB in the face of global environmental changes.

Key word: local adaptation, standing genetic variation, admixture, *Olea europaea* L., genomic environment association, latitudinal gradient

1 | Introduction

Adaptation results from the differential selection of genomic variation within populations, enabling the expression of phenotypes with better fitness in their specific environment, while this fitness is being reduced in different environmental conditions (Kawecki & Ebert 2004; Savolainen *et al.* 2013). This pattern, resulting from natural selection, is very commonly found in plants (e.g. 45% of the species investigated) (Leimu & Fischer 2008). It is also prevalent in the majority of tree species, despite a long generation time (Alberto *et al.* 2013). Local adaptation is favoured by standing genetic variation, meaning the presence of different alleles for one locus in a population (Barrett & Schluter 2008; Savolainen *et al.* 2013; Capblancq *et al.* 2023). Local adaptation can be studied through different methods. One of the most used is the reciprocal transplant experiment, which puts individuals from their original environment to another environment, containing other individuals, and respectively, and quantifies the individual fitness. If the fitness of local individuals in its area of origin is higher compared to foreign individuals, and if the fitness of local individuals is reduced when grown in another area, they are considered as locally adapted. The second one is common garden, conceptually similar, but putting individuals from several environment in a single environment, before measuring the difference of fitness between all individuals in the same place (Kawecki & Ebert 2004; Savolainen *et al.* 2013; Alberto *et al.* 2013; Blanquart *et al.* 2013; Lascoux *et al.* 2016). However, the use of reciprocal transplant/common garden experiments is time-consuming and cost-expensive, especially for long generation trees. The advent of new sequencing methods has provided access to genetic information of species, and more particularly within natural populations. Past studies showed that the use of genome scan approach is efficient to study local adaptation, either as a complement to or independently of

common garden/reciprocal transplant experiments (Steane *et al.* 2014; Hoban *et al.* 2016; Sork 2017). Several methods exist for this aim, one of which is the genomic-environment association (GEA), which tries to identify genomic sequences in association with climatic variable variations (Sork *et al.* 2013; Rellstab *et al.* 2015). Nowadays, GEA is performed in numerous studies (Steane *et al.* 2014; Jordan *et al.* 2017; Leroy *et al.* 2020; Capblancq *et al.* 2023). This approach needs several conditions to be used, as the access to numerous populations and their corresponding environmental data (Steane *et al.* 2014; Hoban *et al.* 2016). Several factors can influence association studies, including linkage disequilibrium and population structure, leading to detection of false-positives (Lotterhos & Whitlock 2015; Hoban *et al.* 2016). These factors should be taken into consideration when correcting the tests by adding them as confounding factors in the association model (Lotterhos & Whitlock 2015).

Today, organisms are threatened by important modification of their environment, through human mediated global change. This phenomenon is resulting from human activities since the advent of the industrial era (Calvin *et al.* 2023), and led to a significant change in climate across the world, and more particularly an increase of average global temperature and changes in precipitation regimes. In the Mediterranean Basin (MB), global change has already been evidenced, characterized by a 20% faster increase in average temperatures compared to the global average (Lionello & Scarascia 2018). This region of the world is a biodiversity hotspot, hosting various endemic plant species (Medail & Quezel 1997). Climate change thus poses a substantial risk to their survival and distribution within the MB. To understand how natural perennial populations could respond to environmental constraints due to global changes, it is important to determine the relative strength of the climatic drivers that

CHAPTER 3

have been involved during past climates, which would help to predict their long-term behaviour in response to climate change and implement conservation measures (Steane *et al.* 2014; Hoffmann *et al.* 2015).

Olive tree, *Olea europaea* subsp. *europaea* var. *sylvestris* is a perennial species emblematic of the Mediterranean climate (Ozenda 1975; Rivas Martínez *et al.* 1987; Kaniewski *et al.* 2012). Its long generation time, as well as its presence along a significant 13° latitudinal gradient, make it an interesting model to investigate local adaptation. Olive tree has an anemophilous and zoochorous reproduction system, supplemented by a strict allogamy (Saumitou-Laprade *et al.* 2017), resulting in substantial gene flow across the MB (>50 km) (Spennemann & Richard Allen 2000). It exhibits a genetically homogeneous wild group in several environmental area of western MB encompassing Southern France, Corsica, Spain, and Morocco (Zunino *et al.* 2023a). Olive tree evolutionary history in the MB began with their arrival during the Middle Pleistocene (Clark *et al.* 2009). They were subsequently impacted by several glaciation events leading to fragmentation of populations and persistence through climatic refugia (Carrión *et al.* 2010; Besnard *et al.* 2013b), before expanding significantly in the MB when climatic environment became more favourable (Planchais 1984; Terral *et al.* 2009). Wild olive is living in sympatry with cultivated olive trees *Olea europaea* subsp. *europaea* var. *europaea* (Zohary & Spiegel-roy 1975; Rubio de Casas *et al.* 2006). Numerous hybridization events between these two genetic groups, wild and cultivated, have resulted in the emergence of a genetically heterogeneous new group, the admixed group, extensively present in the natural environment, along the wild olive tree populations (Zunino *et al.* 2023a). The hybridization might have led to an increase of the standing variation within the admixed individual genome, enabling to cope more easily with a changing environment (Barrett & Schluter 2008).

Here, we examine the presence of local adaptation in wild and admixed genetic groups of olive trees from the western MB. Using previously sequenced genetic data from 359 individuals, including 135 wild and 224 admixed olive trees (Zunino *et al.* 2023b; a), we explore the genomic potential associations with climate through genomic-environment association (GEA) analysis, using climatic variables retrieved from the CHELSA database (Karger *et al.* 2017) from 1981 to 2010 and elevation data from GMTED2010 (Danielson & Gesch 2011). Based on the association results of each group, wild and admixed, we investigated the gene annotations to determine potential functions selected by these environmental variables.

2 | Methods

2.1 | Sampling and sequencing

We analysed 359 natural olive trees, *Olea europaea* var. *sylvestris* along a 13° latitudinal gradient in France, Spain and Morocco (Figure 1, Table 1 and Supplementary Table 1). Naturally occurring olive trees were sampled at 27 sites. The sampling details are described in Zunino *et al.* (2023a) (Figure 1, Table 1 and Supplementary Table 1). The sampling sites were selected based on their climatic data. To prevent redundancy in climatic data per population, for each potential site pre-selected according to the criteria given in Zunino *et al.* (2023a) we extracted climatic data from WorldClim version 2.1 (Fick & Hijmans 2017; Harris *et al.* 2020). For each site, we examined the following climatic variables from 1971 to 2000: the average monthly minimum temperature of the coldest quarter, calculated with the monthly data, and the average monthly precipitation of the driest quarter (Bio17). These variables were chosen according to the biology of the olive tree, with the chilling requirement parameters and the precipitation requirements of olive trees

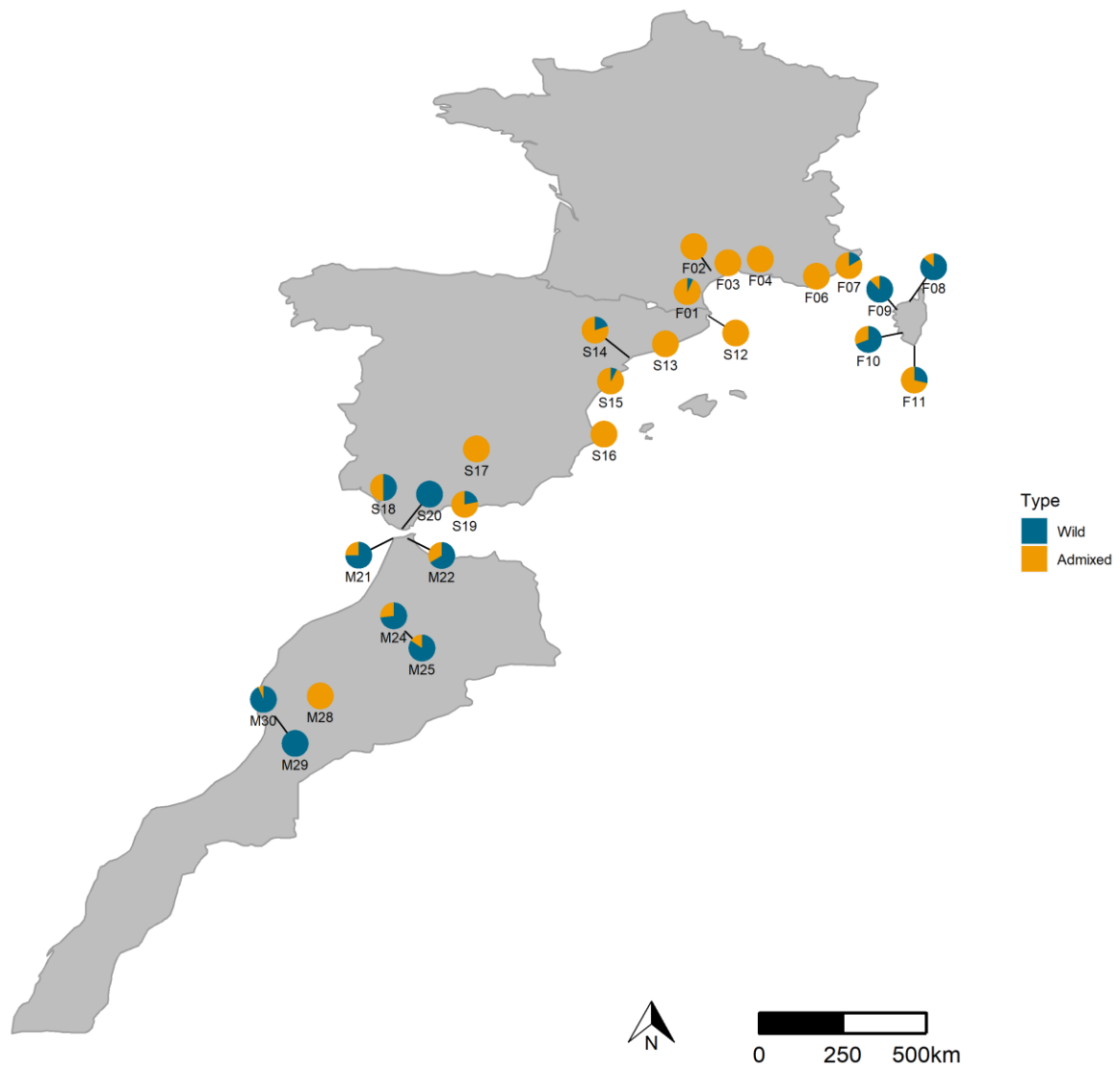


Figure 1. Geographical distribution of the naturally occurring western Mediterranean Basin of *O. europaea* with wild individuals (in blue) and admixed individuals (in yellow). Map from Zunino *et al.* (2023b).

CHAPTER 3

Table 1. Summary information of environmental data uncorrelated of 26 natural populations *O. europaea* subsp. *europaea* var. *sylvestris* used in this study. Climate data extracted from CHELSA and topographic data extracted from GMTED2010.

Sites	N wilds	N admixed	Longitude	Latitude	Country	Temperature						Precipitation				Elevation
						Bio1	Bio2	Bio4	Bio6	Bio8	Bio9	Bio12	Bio14	Bio15	Bio19	gmted
F01	1	14	3.013	42.937	France	15.8	7.7	585	4.65	12.5	23.6	609.2	14.9	40.2	188.6	2.52
F02	0	13	3.303	43.541	France	14.3	8.3	584	3.05	10.9	22.2	866.5	22.6	40.1	247.5	142.85
F03	0	15	3.792	43.771	France	13.3	9.8	607	1.25	13.6	21.5	1069	28.6	42.3	290.6	317.28
F04	0	15	4.733	43.877	France	15	9.9	659	1.75	15.2	23.9	705.1	27.1	40.7	153.5	81.42
F06	0	15	6.361	43.383	France	15.6	9.2	624	2.65	12	24	846.4	17.9	43.1	232	101.79
F07	2	10	7.303	43.689	France	15.8	5.4	538	5.55	13.2	22.8	798.1	16.2	50	224.1	71.99
F08	13	2	9.064	42.654	France	16.7	5.3	545	6.85	14.3	23.9	574.8	9.6	42.7	153.6	33.63
F09	13	2	8.699	42.404	France	16.6	4.4	522	7.45	14.6	23.4	1215	13.3	51.6	376.8	49.70
F10	9	4	8.869	41.750	France	16.2	6.2	525	6.45	14	23	1158	15.9	48.4	354.2	63.19
F11	4	10	9.202	41.372	France	16.8	2.7	471	9.35	15.4	22.9	600.7	7.3	50.5	162.3	45.58
S12	0	15	3.219	42.235	Spain	15.6	3.7	479	7.75	17	21.9	644.3	18.2	38.3	173.9	36.83
S13	0	14	1.976	41.421	Spain	14.9	7.4	545	3.75	16.2	22	640.1	25.2	32.2	129.4	241.22
S14	3	12	0.935	41.020	Spain	16.6	6.3	535	6.45	18	23.6	561.1	14.8	43.2	116.7	40.93
S15	1	12	0.386	40.341	Spain	17.8	8.1	551	6.15	18.8	25	604	12.5	41	147.7	9.99
S16	0	12	0.195	38.803	Spain	17.9	2.9	463	10.7	19.8	23.8	776.6	6.7	50.7	218.2	48.26
S17	0	14	-3.507	38.374	Spain	16.8	10.8	719	2.35	12.3	26.5	534.3	4.9	51.2	183.3	606.55
S18	6	6	-6.209	37.256	Spain	18.7	11.1	571	5.95	15.4	26.2	612	2	68.8	248.6	37.31
S19	3	11	-3.850	36.766	Spain	17.8	3.8	428	10.3	13.7	23.4	464	1.2	68.1	189.5	200.24
S20	13	0	-5.669	36.062	Spain	18	5.3	368	10.5	14.5	22.6	710.3	1.9	76.5	326.8	40.58
M21	9	3	-5.925	35.790	Morocco	17.5	5	366	10.4	14.1	22.1	733.9	2.4	75.6	323.6	72.11
M22	10	5	-5.515	35.783	Morocco	17.6	5.9	381	9.75	14.1	22.5	639.7	1.6	74.2	288.9	157.56
M24	11	4	-5.908	33.534	Morocco	16.9	12.8	514	3.55	11.3	23.6	512.5	4.2	58.2	205.1	703.84
M25	11	2	-5.588	33.100	Morocco	16.4	12.7	632	1.65	8.75	24.8	719.1	7.6	62.7	319.3	1098.03
M28	0	13	-8.040	31.210	Morocco	17.2	12.4	545	3.65	13.4	24.5	431.1	8.5	43.7	130.4	983.07
M29	12	0	-9.370	30.631	Morocco	18.3	12.4	539	5.35	11.6	25.4	360	1.7	82.8	182.4	905.89
M30	14	1	-9.691	31.111	Morocco	19.1	8.1	350	10.3	15.6	23.6	379.2	1.3	85.3	195.2	234.78

N= number of individuals, longitude and latitude in wgs84, Bio1: Mean annual air temperature (°C), Bio2: Mean diurnal temperature arrange (°C), Bio4: Temperature seasonality (°C/100), Bio6: Mean daily minimum air temperature of the coldest month (°C), Bio8: Mean daily air temperatures of the wettest quarter (°C), Bio9: Mean daily air temperatures of the driest quarter (°C), Bio12: Annual precipitation amount (Kg m-2 year -1), Bio14: Precipitation amount of the driest month (kg m-2 month-1), Precipitation seasonality (kg m-2), Bio19: Mean monthly precipitation amount of the coldest quarter (kg m-2 month-1), gmted: elevation (m).

CHAPTER 3

(Haberman *et al.* 2017; Alfieri *et al.* 2019). Then, the sites were grouped in six clusters based on these data using a kmeans clustering approach, a machine learning method whose aim is to regroup similar data in clusters, implemented with the kmeans function from R version 4.2.0, (R Core Team 2022). Within each cluster, we selected one to two potential sites to avoid redundancy of climatic data between geographically closed populations. After sampling all populations, we extracted the 19 Bioclimatic variables of each site from the CHELSA database, which contains recent climatic variables from 1981 to 2010 (Karger *et al.* 2017; Brun *et al.* 2022). A final assessment of climatic data heterogeneity at the sites was conducted using a Principal Component Analysis (PCA; Supplementary Figure S1). We also included 135 cultivated accessions of *Olea europaea* var. *europaea* from the Worldwide Olive Germplasm Bank of Marrakech (WOGBM) and 10 cultivated French varieties corresponding to varieties significantly present in the sampling area of natural populations, as previously analysed by Zunino *et al.* (2023a) (Supplementary Table S1). DNA extraction, library preparation and target sequencing methods are described in Zunino *et al.* (2023a). The raw sequence reads have been deposited in the European Nucleotide Archive (PRJEB61410).

2.2 | SNP calling and SNP filtering conducted on the basis of genetic groups: Wild, admixed and cultivated olive trees.

After sequencing, we used a bioinformatic pipeline for variant calling (Zunino *et al.* 2023a) and followed several filtration steps as described in (Zunino *et al.* 2023b). Except for cultivars which were kept as one group (n= 145), samples were regrouped according to their genetic assignment to an ancestral genetic cluster using sNMF from LEA package (Frichot & François 2015) and described in (Zunino *et al.* 2023b). They were split in two groups, wild and admixed olive, with respectively 135 and 224 trees (Zunino *et al.* 2023b) (Supplementary Table S1). Within each of the three groups, we

subsequently applied a minor allele count of one using VCFtools (Danecek *et al.* 2011), followed by a minor allele frequency (MAF) of 5% and a final filter to exclude positions with fixed heterozygosity (>85%). This filtering process yielded to final sets of 96,648 SNPs for genuine wild olives, 134,664 SNPs for admixed olives, and 109,428 SNPs for cultivated olives. For the next steps, we focused on the wild and admixed datasets, sampled in the natural environments. To mitigate the effects of linkage disequilibrium (LD), which could lead to the detection of false positives in the data (Lotterhos & Whitlock 2015), we also generated a second dataset for each of these groups with pruned genotypes. To this aim, we filtered the variants of each dataset with PLINK version 2.0 (Purcell & Chang 2015; Chang *et al.* 2015) to remove link variants with a $r^2 > 0.2$ in an sliding window of 50 base pairs, resulting in 71,632 unlinked SNPs for genuine wild olive trees and 102,930 unlinked SNPs for admixed olive trees.

2.3 | Selection of environmental variables

The environmental variables for each site were extracted from the 19 Bioclimatic variables of CHELSA from 1981 to 2010. This dataset has a spatial resolution of 30 arc seconds, corresponding to one square km (Karger *et al.* 2017) at global elevation. Soil elevation data for each site was extracted using the GMTED2010 database, with the same spatial resolution as climatic data (Danielson & Gesch 2011) (Supplementary Table S2).

We examined the correlation between the environmental variables in three steps: first, a principal component analysis (PCA) was performed on R version 4.2.0 (Supplementary Figure S2). Subsequently, a correlation matrix with a Pearson test was generated (Supplementary Figure S3). When the correlation between two environmental variables was found to be less than 75%, only one of the two was

CHAPTER 3

retained for the study. To ensure the remaining environmental variables had minimal correlation, a second PCA was performed (Supplementary Figure S4).

2.4 | Genomic environment association analysis

Genomic environment analysis (GEA) was conducted using a latent factor mixed-model analysis (function LFMM2 from LEA v3 package; Caye *et al.* (2019)) for both wild and admixed groups separately). For each dataset, an initial genetic structure analysis was carried out using sNMF from the LEA package version 3.11.3 (Frichot & François 2015) to determine the number of ancestral genetic clusters based on cross-entropy criterion. This allowed imputation of missing data in each dataset according to the best practices of LEA package for lfmm2 analysis. This lfmm2 analysis uses a two-step approach to estimate association between genetic and environmental variables. In a first step, the estimation function (lfmm2), build an object containing the latent factors (confounding factors e.g. genetic structure) which will be included as correction factors, then in a second step, the test function (lfmm2.test) estimates the significance of associations between the allelic frequencies at each SNP and the environmental variables chosen, using linear model tests. The environmental variables can be tested one by one or being computed together for the test. We used a pruned dataset for the estimation of latent factors, while the test function was performed using all data. This allowed to reduce the possible false positive association due to high LD in some regions. For both datasets (admixed and wild olives), the tests were conducted 1) on all environmental variables tested together (call "*All*", containing Bio1, Bio2, Bio4, Bio 6, Bio8, Bio9, Bio 12, Bio 14, Bio 15, Bio19, gmted), 2) on group of temperature variables (call "*Temperature*", containing Bio1, Bio2, Bio4, Bio 6, Bio8, Bio9), 3) on group of precipitation variables (call "*Precipitation*" containing Bio 12, Bio 14, Bio 15, Bio19), 4) on the "*elevation*" variables using gmted data, then 5) on all the "*Bio*"

separately (Table 1). Significant outliers were assessed using either Bonferroni correction ($\alpha = 0.05$, $n_{\text{Wild}} = 96,648$, $n_{\text{Admixed}} = 134,664$) or false discovery rate (FDR = 10%) with *q-value* test package *q-value* version 2.30.0 (Storey *et al.* 2022). SNPs with *p-value* above the *q-value* test threshold but below Bonferroni threshold were considered as minor outliers while those only above the Bonferroni correction threshold were considered as major outliers.

2.5 | Gene functional annotation, prediction of mutation and gene-ontologies enrichment analysis

We conducted a gene functional annotation analysis for genomic regions containing outlier loci. Each gene containing associated SNPs were identified using bed windows from bedtools version 2.30.0 (Quinlan & Hall 2010) to intersect the outliers with the annotated reference genome Farga (Julca *et al.* 2020). In cases where a SNP was hitting two different genes due to geneID overlapping in the annotation reference genome Farga V2 (Julca *et al.* 2020), we chose to retain both genes. For each identified gene, we looked for their potential function by retrieving their specific gene ontology (GO) term and function with GO.db version 3.16.0 (Carlson 2022) and AnnotationDbi version 1.60.2 (Pagès & Carlson 2023). For each results of association analysis with a climatic variable, we performed a GO term enrichment test with clusterProfiler version 4.6.2 (Wu *et al.* 2021). Only enriched Go terms with an adjusted enrichment *p-value* below 0.05 combined with a *q-value* under 0.1 were considered as significant. The results were graphically represented with enrichplot version 1.18.4 (Yu 2023).

SnEff version 4.3 was used to annotate and predict the effects of the outlier SNPs on genes and proteins. The database for SNPEff was created using the reference genome annotation, the DNA sequences, the protein sequences and the transcripts

CHAPTER 3

(Cingolani *et al.* 2012; Julca *et al.* 2020). We used the information to estimate the putative impact/deleteriousness of each outlier which can be either high (e.g. stop gained), moderate (e.g. missense-variant), modifier (e.g. downstream gene variant) or low (e.g. synonym variant).

2.6 | Characterisation of the candidate SNPs for association to the environment according to genotypic groups

The frequency of each candidate SNPs for association to the environment per population was determined using PLINK 2.0 (Purcell & Chang 2015). We performed PCA on each frequency dataset of outliers, depending on the tested variables, and on PCA on the same environmental variables tested. We did a procrustes analysis to compare the shape of allelic frequencies and climatic variables based PCA using vegan packages 2.6.4 (Oksanen *et al.* 2022) to determine if a convergence was observed within populations, according to their allelic frequencies of SNPs attributed to climatic data and impact on shaping their geographical distribution.

We performed genetic structure analysis based on the candidate SNPs identified by GEA analysis using sparse non-negative matrix factorisation approach (sNMF function from LEA packages version 3.11.3; Frichot & François (2015)) to determine the optimal number of ancestral genetic clusters. To verify if the genetic structure resulting from SNPs significantly associated with climate differed from random, we repeated the same structure analysis based on ten random draws of SNPs the size of candidate SNPs matrix in the wild data set, i.e. we ended up with ten structure analyses based on ten sets of randomly drawn SNP matrixes the size of the candidate matrix. We also verified the presence/absence of these SNPs in admixed and cultivated data set as well as for the SNPs detected as associated to climate in the wild and cultivated dataset.

Using candidate SNPs detected within the wild group, we calculated the F_{ST} and D_{ST} locus-based values between the admixed and the wild groups with Hierfstat package version 0.5.11 (Goudet & Jombart 2022). Additionally, we randomly selected SNPs matrixes the size of the candidate matrix from the wild individuals ten times and performed the same F_{ST} and D_{ST} calculations. Statistical differences between the F_{ST} and D_{ST} values were calculated from the candidate SNPs and those obtained from the random samplings using a Wilcoxon test from rstatix package (Kassambara 2023).

3 | Results

3.1 | Identification of candidate SNPs in wild and admixed populations

The GEA was performed using the following selected uncorrelated environmental variables : *mean annual air temperature* (Bio1), *mean diurnal air temperature range* (Bio2), *temperature seasonality* (Bio4), *mean daily minimum air temperature of the coldest month* (Bio6), *mean daily mean air temperatures of the wettest quarter* (Bio8), *mean daily mean air temperature of the driest quarter* (Bio9), *annual precipitation amount* (Bio12), *precipitation amount of the driest month* (Bio14), *precipitation seasonality* (Bio15), *mean monthly precipitation amount of the coldest quarter* (Bio19) and *elevation* (gmted) (Table 1). The results were considered for wild olive as significant and major outliers when p -values < 5.17e-07 and for admixed olives when p -values < 3.71e-07. The FDR threshold for considering significant association varied for each tested environmental variable, ranging between q -value = 1.40e-07 and q -value = 1.57e-04 for admixed olive and between 1.29e-08 and 1.13e-03 for wild olives (Supplementary Table S3).

The wild groups revealed 1096 SNPs selected by “All” including 55 major outliers. For the admixed group, 1316 SNPs are obtained, including 51 major outliers (Figure 2,

CHAPTER 3

Table 2). There are 665 minor outliers related to “*Temperature*” in the wild group, including 28 major outliers, while we detected only 92 minor outliers in the admixed group, including five major outliers. Regarding the group of variables related to “*Precipitation*”, there are 245 associated SNPs in the wild group, including 22 major outliers, while 213 are found in the admixed group, including seven major outliers. SNPs related to “*Elevation*” are exclusively present in the wild group, consisting of 39 SNPs, including seven major outliers (Figure 2, Table 2). By looking for individual environmental variables separately, the wild group has SNPs selected by five environmental variables. Three temperature variables have associated outliers: 12 minor outliers for “*Bio2*” including one major outlier, 16 for “*Bio6*” including three major outliers and 53 for “*Bio8*” including five major outliers. For precipitation, there are also two variables with associated SNPs: two major outliers for “*Bio14*” and 52 for “*Bio15*”, including six major outliers (Table 2, Supplementary Figure S5). In the admixed group, only two climate variables have associated SNPs, the climatic variable “*Bio8*” related to temperature included one minor outlier, and climatic variable “*Bio12*” related to precipitation contains one major outlier (Table 2, Supplementary Figure S6). All QQplot and histogram of *p-value* for wild and admixed are in the supplementary (Supplementary Figures S7 to S10).

Table 2. Number of SNPs associated for each environmental variable for wild and admixed genotype groups.

Environmental variables	Wild		Admixed	
	Minor	Major	Minor	Major
All (Bio1, Bio2, Bio4, Bio6, Bio8, Bio9, Bio12, Bio14, Bio15, Bio19, gmted)	1096	55	1316	51
Temperatures (Bio1, Bio2, Bio4, Bio6, Bio8, Bio9)	665	28	92	5
Precipitations (Bio12, Bio14, Bio15, Bio19)	245	22	213	7
Elevation (m) - gmted	39	7	0	0
Mean annual air temperature (°C) - Bio1	0	0	0	0
Mean diurnal air temperature range (°C) - Bio2	12	1	0	0
Temperature seasonality (°C/100) - Bio4	0	0	0	0
Mean daily minimum air temperature of the coldest month (°C) - Bio6	16	3	0	0
Mean daily mean air temperatures of the wettest quarter (°C) - Bio8	53	5	1	0
Mean daily mean air temperatures of the driest quarter (°C) - Bio9	0	0	0	0
Annual precipitation amount (kg m ⁻² year ⁻¹) - Bio12	0	0	1	1
Precipitation amount of the driest month (kg m ⁻² month ⁻¹) - Bio14	2	2	0	0
Precipitation seasonality (kg m ⁻²) - Bio15	52	6	0	0
Mean monthly precipitation amount of the coldest quarter (kg m ⁻² month ⁻¹) - Bio19	0	0	0	0

Major outliers are SNPs with association score above p -value < 0.05 after Bonferroni correction, the minor outliers are SNPs with association scores above q value > 0.1 .

Among all the datasets of SNPs associated with environmental variables using “*All*”, only 14 are common to wild and admixed groups, none are major outliers (Figure 3A). However, by comparing groups using genes rather than SNPs, there are 34 common genes between admixed and wild groups (Supplementary Figure S11). Regarding SNPs selected by “*Temperature*” and “*Precipitation*”, neither of them had common SNPs between wild and admixed genetic groups (Figure 3B, Figure 3C). When looking

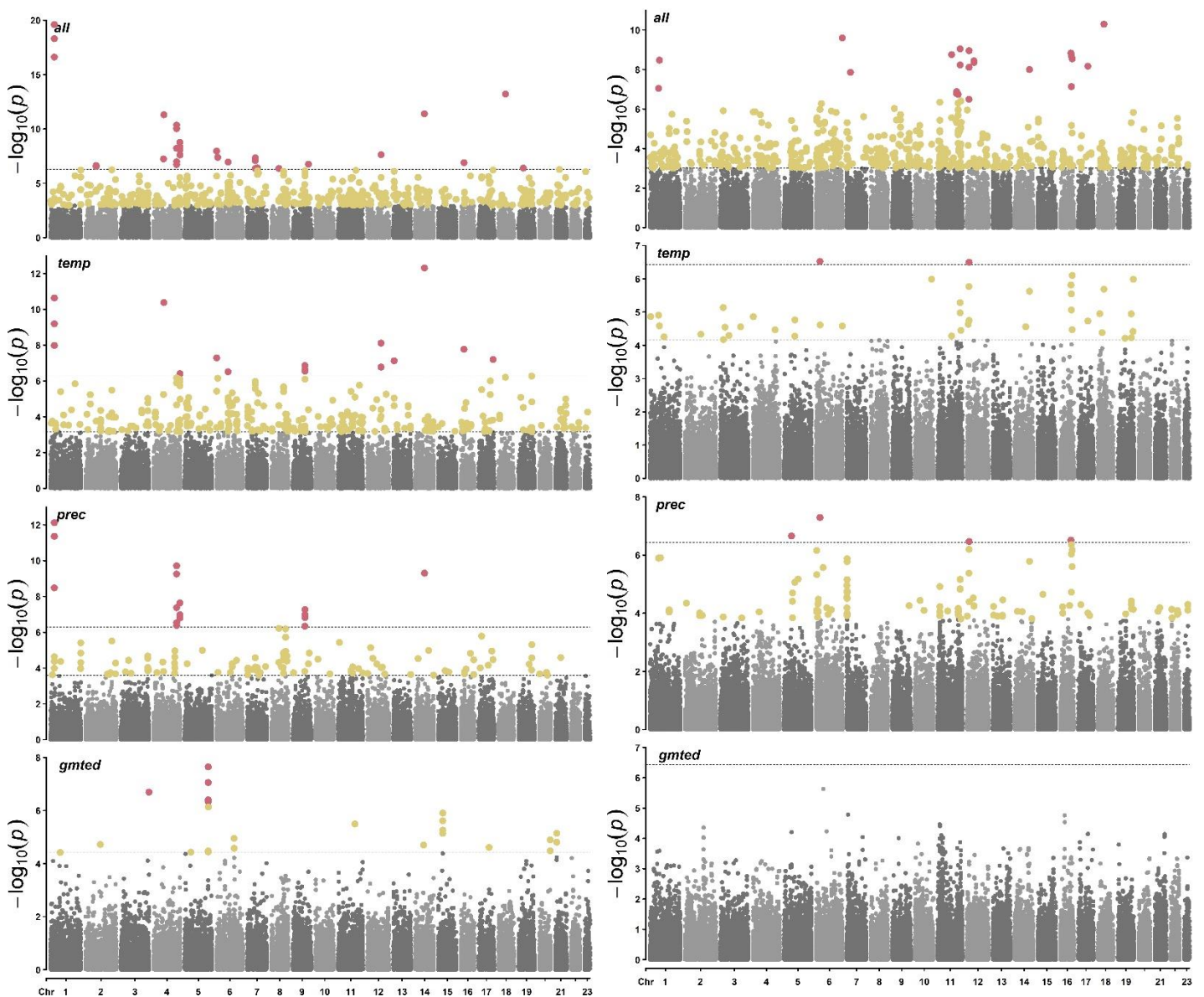


Figure 2. Genome-wide scans mapping of different genomic-environment associations (GEA) for naturally occurring western Mediterranean Basin of *O. europaea* populations. Each point represents a SNP. The yellow points represent minor outlier SNPs, with association score q -value < 0.1 . The pink points represent major outlier SNPs, with association score above p -value < 0.05 after Bonferroni correction with (A) Wild genotype associations to environmental variables in order “All”, “Temperature”, “Precipitations” and “Elevation”. (B) Admixed genotype associations to environmental variables in order “All”, “Temperature”, “Precipitations” and “Elevation”.

at genes, there are only two common genes for “*Temperature*” and none for “*Precipitation*” (Supplementary Figure S11).

Within the wild genotype group, 139 SNPs have an environmental association with at least “*Precipitation*”, “*Temperature*” and “*All*”, and at least 495 SNPs were associated both with “*Temperature*” and “*All*”. For the “*Precipitation*” and “*All*”, 218 SNPs were unique to them, 147 unique SNPs common to “*Precipitation*” and “*Temperature*”. Regarding “*Elevation*”, only one is common to precipitation and “*All*”, eight are exclusively associated with “*All*” (Figure 3D). For the admixed group, we obtained 42 SNPs associated with “*All*”, “*Temperature*” and “*Precipitation*”. “*Temperature*” and “*Precipitation*” also have 89 common SNPs. None are common to “*Precipitation*” and “*Temperature*” only (Figure 3E).

3.2 | Description of associated SNPs and genes in wild and admixed populations

Within the 1,400 SNPs detected for the wild groups using GEA, 1,343 were associated with genes. For the admixed groups, 1379 SNPs were detected, associated to 1,310 genes. Some SNPs were mapped to multiple genes due to the structural genome annotation of the reference genome Farga (Julca *et al.* 2020). Using the AnnotationDbi package (Pagès & Carlson 2023), we identified genes associated with specific functions for wild and admixed group. A comprehensive list of genes and their orthologous gene functions is provided in Supplementary Table S4.

The result of the enrichment test revealed 16 significantly enriched GO terms with *p-value* ranging from 7.14e-05 to 4.90e-03 for the admixed genotypes, including 14 unique GO terms, and 125 GO terms with *p-values* ranging from 2.68e-05 to 4.95e-02 for the wild genotypes including 105 unique GO terms (Figure 4 and Supplementary Figure S12).

CHAPTER 3

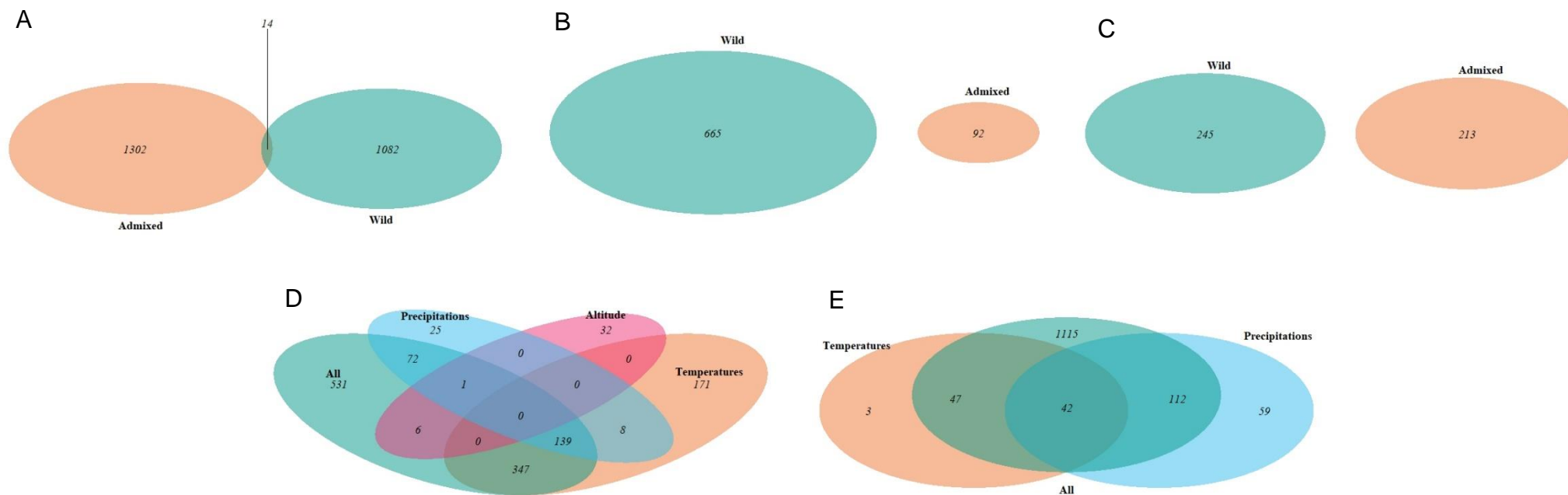
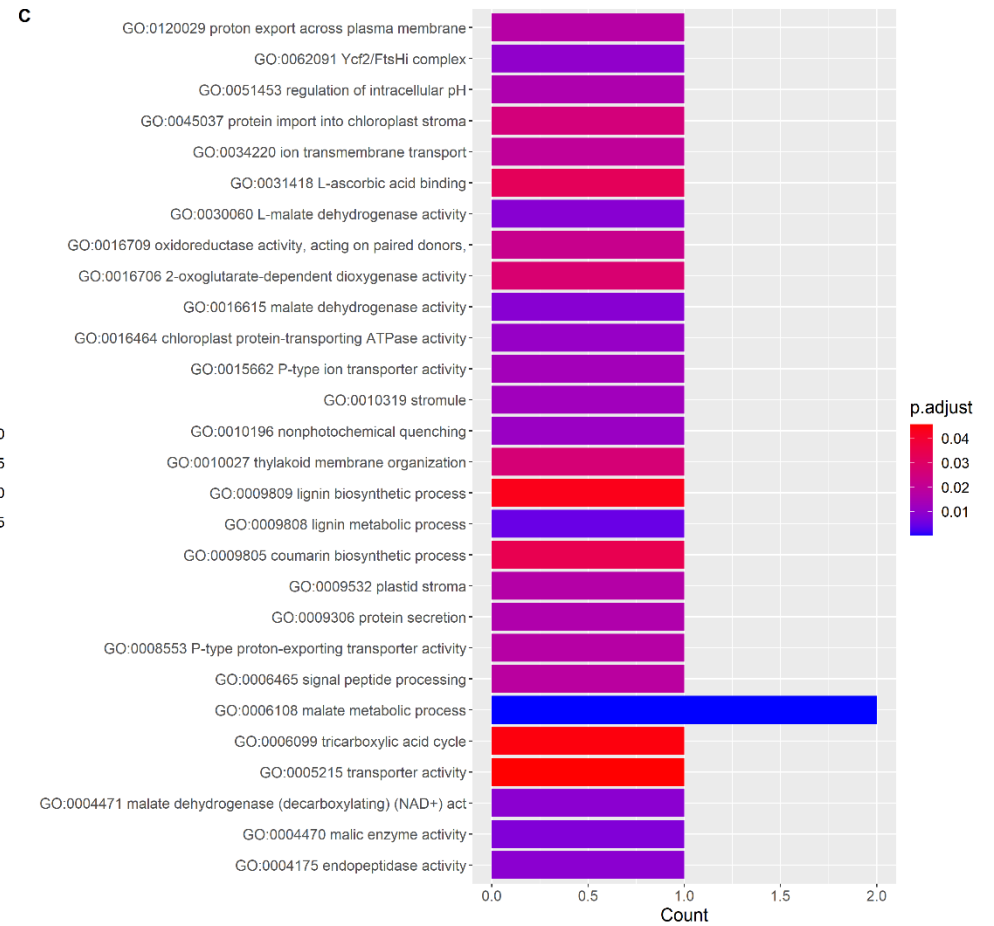
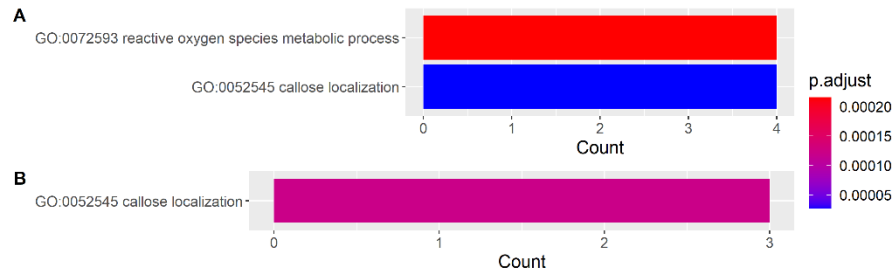


Figure 3. Venn diagram of outlier SNPs associated with environmental variables for wild and admixed genotype groups. (A) Between wild and admixed genotypes for variables group “All” (B) between wild and admixed genotypes for variables group “Temperature” (C) between wild and admixed genotypes for variables group “Precipitations” (D) In wild genotypes, between “All”, “Temperature”, “Precipitation” and “Elevation” (E) In admixed genotypes, between “All”, “Temperature” and “Precipitation”.

For the admixed genotypes, two GO terms, GO:0010629 and GO:0010031, were detected in both the "*Temperature*" and "*Precipitation*" groups, involving "negative regulation of gene expression" and "circumnutation". The most significantly enriched GO term, GO:0010319 (*p-value* 7.14e-05), linked to the "*Precipitation*" group, is involved in "stromule" formation. The second and third most significant enriched GO terms were GO:0012510, related to "trans-Golgi network transport vesicle membrane" (associated *p-value* 6.70e-04), and GO:0010100, involved in "negative regulation of photomorphogenesis" (associated *p-value* 7.80e-04). Several other significantly enriched GO terms were found, including two related to defense against pathogens, GO:1900150, involved in the "regulation of defense response to fungus" (*p-value* 3.37e-03), and GO:1900424, involved in the "regulation of defense response to bacterium" (*p-value* 2.72e-03). Additionally, GO:0007623 (*p-value* 2.15e-03) was associated with "circadian rhythm", and GO:0009834 (*p-value* 4.51e-03) was linked to "plant-type secondary cell wall biogenesis".

For the wild genotypes, GO:0010319, related to "stromule", was also found twice in the "*Bio15*" and "*Elevation*" groups (*p-value* 1.66e-02 and 1.28e-02 respectively). Specifically, 16 GO terms were enriched at least twice, while one GO term, GO:0010196, involved in "photochemical quenching", was enriched three times in the "*Bio6*" (*p-value* 4.00e-03), "*Bio15*" (*p-value* 1.78e-02), and "*Elevation*" (*p-value* 1.14e-02) groups. Among the 16 GO terms enriched twice, five were associated with malate formation or activity (GO:0004470; *p-value* "*Bio15*" 9.86e-03 and "*Elevation*" 7.59e-03, GO:0004471; *p-value* "*Bio15*" 1.17e-02 and "*Elevation*" 9.01e-03, GO:0006108; *p-value* "*Bio15*" 2.04e-04 and "*Elevation*" 1.20e-04, GO:0016615; *p-value* "*Bio15*" 1.11e-02 and "*Elevation*" 8.53e-03, GO:0030060; *p-value* "*Bio15*" 1.11e-02 and "*Elevation*"

CHAPTER 3



CHAPTER 3

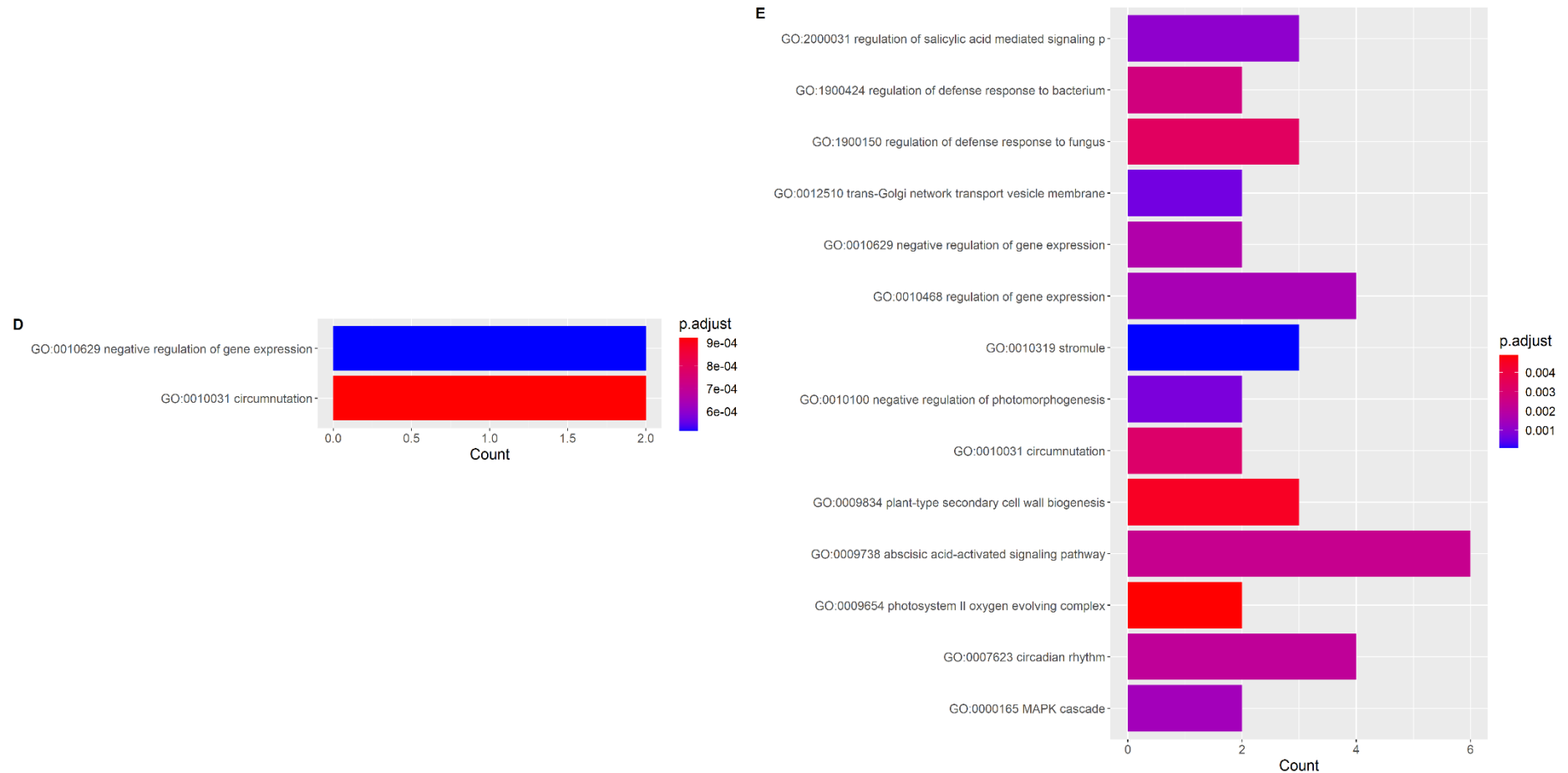


Figure 4. Gene ontology (GO) term enrichment analysis based on SNPs associated with environmental variables for wild and admixed genotype groups. (A) SNPs associated to “*All*” for wild, (B) SNPs associated to “*Temperature*” for wild, (C) SNPs associated to “*Elevation*” for wild, (D) SNPs associated to “*Temperature*” for admixed and (E) SNPs associated to “*Precipitation*” for admixed.

CHAPTER 3

8.53e-03). Additionally, we found GO:0052545 involved in "callose localization" (*p-value* "Temperature" 1.46e-04 and "All" 2.68e-05). GO:0031418, involved in "L-ascorbic acid binding" (*p-value* "Bio15" 4.18e-02 and "Elevation" 3.23e-02), the GO:0009532 corresponding to "plastid stroma" (*p-value* "Bio15" 2.20e-02 and "Elevation" 1.70e-02) and GO:0009735 (*p-value* "Bio6" 1.21e-02 and "Bio15" 4.42e-02), involved in the response to cytokinin.

3.3 | SNP characterisation and localisation along the geographic gradient according to wild and admixed groups

Comparing PCAs of outlier SNPs frequency per population to PCAs of environmental variables tested per population, there are no significant correlations except for "Temperature" and "All" tests (Symmetric procrustes associated *p-value* < 0.05, 999 permutations).

The profile (heterozygous, reference homozygote or alternative homozygote) and the genetic structure of each associated SNPs to environmental variables for wild and admixed groups according to latitudinal gradient (from southern France to Morocco) is represented in Figure 5. Within the wild associated SNPs to environment (Figure 5A), we observed a genetic profile pattern along the latitudinal gradient, ranging from northern to southern regions. This pattern exhibits homozygosity for the alternative allele in the north, heterozygosity in the central regions and homozygosity for the reference allele in the south. The opposite pattern was observed as well. Furthermore, distinct genetic patterns were evidenced for populations further south (M29 and M30), northern Spain, continental France (F01, F07, and S14), and Corsica (F11, F08, F09, F10). In the structure analysis using associated SNPs (Figure 5b, Supplementary Figure S13), three ancestral genetic clusters were found, corresponding for a first one to populations in Corsica, continental France and north Spain (S14), a second cluster

composed of populations in Spain and north Morocco, then a third one mostly of populations in the south of Morocco (M29 and M30) but also present in populations from the centre of Morocco (M24 and M25). We found a similar pattern by testing only the structure of major associated SNPs, exhibiting in addition a more discernible delineation among southern populations (M29 and M30) (Supplementary Figure S14 and S15). The analysis of random SNPs data set led to a similar pattern but not as distinct as in outlier SNPs. The cross-entropy suggested a most likely value for K of 1 or 2 rather than 3 (Supplementary Figure S16 to S18). Conversely, within the admixed associated SNPs to the environment (Figure 5C), the genetic profile is mostly uniform. There are only two different clusters, a northern one from 43.88°N to 38.80°N and a southern one from 38.37°N to 30.63°N. In figure 5D, the genetic structure exhibits a similar configuration with two distinct ancestral genetic clusters. Despite pronounced intra- population admixture, there is a dominant presence of the first genetic cluster within Corsica, Continental France, and northern Spain. At 38°N of latitude there is a shift between the first genetic cluster and the second, subsequently yielding to a dominant presence in the south of the second genetic cluster encompassing the M28 population in Morocco. Among the outlier SNPs found in the admixed genetic groups (admixed outlier SNPs) to be associated to environment, 44.1% (608/1379) are found in wild SNPs dataset while 88.0% (1214/1379) are found in cultivated SNPs dataset. Among the outlier SNPs found in the wild genetic groups (wild outlier SNPs) to be associated to environment, 48.8% (682/1400) are found in cultivated SNPs dataset while 80.4% (1125/1400) are encountered in admixed SNPs data set. The structure of each wild outlier SNPs found in admixed genotypes is represented in Supplementary Figure S19 to S20. Three genomic ancestral groups of admixed genotypes are defined according to structure along the latitudinal gradient.

CHAPTER 3

The genomic differentiation between the wild and the admixed groups were calculated. Using wild associated SNPs to climate ($n = 1400$) and ten randomly sampled set of SNPs within the wild dataset ($10 \times n = 1400$), we calculated the F_{ST} for each dataset and compared the results using Wilcoxon-Mann-Whitely with Bonferroni correction (Supplementary Figure S21). Two datasets of random SNPs were significantly less differentiated than outliers SNPs (Wilcoxon-Mann-Whitely – $p\text{-value} < 0.05$), the rest of the test were not significant (Wilcoxon-Mann-Whitely– $p\text{-value}$ between $5.43e-02$ and 1).

4 | Discussion

We investigated the genomic associations of naturally occurring populations of *O. europaea* in the western Mediterranean Basin. Our study used target sequencing data to conduct a genome scan involving 224 admixed olive trees and 135 genuine wild olive trees across 26 populations sampled along a latitudinal gradient of 13° (from 43.9°N to 30.6°N), encompassing regions in southern France, Corsica, Spain and Morocco, as previously described by Zunino *et al.* (2023a). Using climatic and elevation data retrieved for each site, we performed genome-environment association analyses for both genotype groups, leading to the identification of 2,763 unique SNPs associated with environmental factors. These findings allowed us to characterize local adaptation of genuine wild and admixed genotypes, elucidate potential gene functions, and understand the climatic drivers involved in their selection. Additionally, we conducted a comparative analysis between the two genotypic groups to assess the important presence of admixed genotypes within the natural environment.

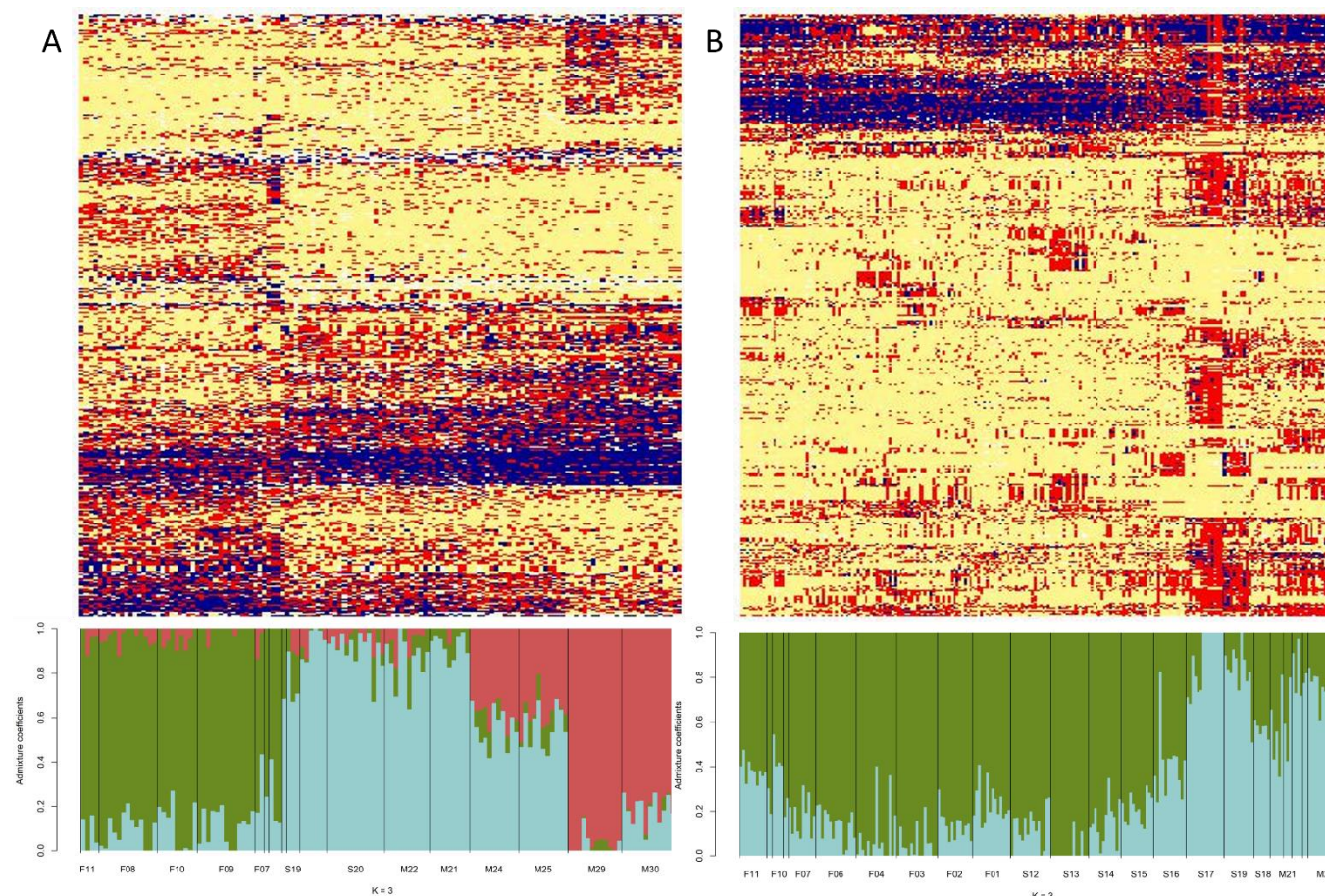


Figure 5. Heatmap of the allelic doses and genetic structure (sNMF) of all SNPs associated with environmental data according to the geographic localisation of their genotype along the latitudinal gradient of genotypes (From Corsica, French continent, Spain to Morocco). For the heatmap, the SNPs were ordered by hierarchical clustering based on local similarity. SNPs are bi-allelic with in yellow reference homozygous, in red, heterozygous, in blue alternative homozygous and white, missing data. (A) Associated SNPs of wild genotype group ($n = 1400$) and (B) Associated SNPs of admixed genotype group ($n = 1379$ SNPs).

4.1 | Local adaptation of western wild olive trees to climate?

Genome-environment association analysis using wild genotypes of *O. europaea* revealed that various SNPs were found to be correlated with environmental variables. When considering all environmental variables collectively, a larger number of associated SNPs was detected, followed by the use of temperature-related variables. These results are supported with the procrustes analysis, demonstrating the significant influence of temperature on the spatial distribution of populations based on their allelic frequencies. These variables, in addition to the precipitation group, have been previously highlighted in other species, such as eucalyptus, spruce, oak and conifer (Grivet *et al.* 2011; Jordan *et al.* 2017; Ramírez-Valiente *et al.* 2022; Capblancq *et al.* 2023).

This important impact of temperature appears to be strongly supportive of the temperature requirements of olive trees. Indeed, wild olive trees are found in the Mediterranean climate, characterized by specific temperature and precipitation patterns. This climate is defined by a dry summer season, which can vary from warm to hot depending on the region (Köppen 1900; Lionello *et al.* 2012). Olive tree is considered as one of the main species representatives of this climate (Medail & Quezel 1997). When considering olive tree biology, the impact of temperature seems very supportive, temperatures variations within the range of 0 to -3°C result in slight damage on the young tree whereas temperatures falling below -10°C can cause severe damage (Fiorino & Mancuso 2000; Petruccelli *et al.* 2022). The impact of temperature tolerance has been assessed in cultivated individuals (Fiorino & Mancuso 2000), but similar evaluations in wild populations have not been conducted. However, significant variations in cold tolerance have been observed among cultivated varieties,

implying the potential for substantial variation among wild populations as well, which we may observe in our results.

Regarding precipitation-related variables, the key drivers for SNP selection are primarily "precipitation seasonality" (indicating irregularity of rainfall across months within years) and precipitation during the driest months. These environmental drivers of selection in the olive tree, temperature and precipitation, in line with those of other perennial species in the MB as conifer (Grivet *et al.* 2011) or oleaceous species such as weeping forsythia, the most important environmental drivers in adaptive differentiation were waterlogging, cold and drought (Li *et al.* 2020).

Various SNPs detected as being under selection in the wild group are associated with functions involved in the response of organisms to environmental factors such as drought resistance and heat shock. The different functional enrichment highlighted different functions, known as involve in environment adaptation: malate formation, for example, means the gene is implicated in the generation cycle of malate, which is a molecule involved in various biological processes, including the Krebs cycle, and is known to play a role in responding to temperature and aluminum resistance in several plants (Sun *et al.* 2019). The "L-ascorbic acid binding" has been found in citrus plants to help in salty soils adaptation (Kostopoulou *et al.* 2015). The "plastid stroma" is a structure associated with adaptation to abiotic temperature stress (Lütz & Engel 2007; Venzhik *et al.* 2019) and the cytokinin is known to be related to various responses to environmental changes (Argueso *et al.* 2009).

Furthermore, the genetic structure analysis of SNPs associated with the environment enabled the delineation of three possible ancestral genetic clusters along the latitudinal gradient: the first cluster encompassing South France, Corsica, and Northern Spain (from F11 to S14); the second cluster extending from S15 in Spain to

CHAPTER 3

M21 in Morocco; a mixture of clusters 2 and 3 for the central Moroccan populations; and a third cluster for the southern populations of Morocco. This delineation was not possible previously when considering all SNPs. Our results argue in favour of adaptation of wild olive to their environment.

4.2 | Local adaptation of admixed olives trees?

Genome-environment association analysis using admixed genotypes revealed that various SNPs were found to be correlated with environmental variables. There is a greater prevalence of candidate SNPs associated with the environment when considering “All” variables compared to wild genotypes. However, when examining temperature-related variables, very few SNPs are detected; and for the variables analyzed separately, only two revealed associated SNPs, with just one SNP outlier each. This suggests that admixed olive trees may not exhibit clear local adaptation to their environment. Notably, no clear patterns emerged from population structure and heatmap analysis, except for a distinct division between populations in southern France and northern Spain compared to those in southern Spain and Morocco. This pattern could be attributed to the distribution of olive cultivation, with two distinct groups of cultivated olives identified in this region: the western cultivated (Q3) in Southern Spain and Morocco and the central cultivated (Q2) in France and North Spain, as reported by Diez et al. (2015) and observed by Zunino et al (2023), or to an isolation by distance of the sampling sites. The detection of SNPs in this study may indeed be influenced by the complex genetic structure observed in the admixed genotypes. This complexity may not have been fully resolved due to their recent admixture history. The admixed genotypes may not be adapted or alternatively, the detection of selection based on environmental cues is blurred by a recent admixture

event. These events might have disrupted or obscured the signals of selection due to the mixing of genetic backgrounds (de Villemereuil *et al.* 2014; Rellstab *et al.* 2015; Zunino *et al.* 2023b). This complex genetic structure is a significant confounding factor in the identification of local adaptation (de Villemereuil *et al.* 2014; Rellstab *et al.* 2015). To improve the analysis, we could study the pattern of introgression as well as the precise parent groups at the origin of hybridization of admixed individuals.

Few functional enrichment of SNPs detected as associated to environment for admixed genotypes were in line with a climatic response, except the circadian rhythm, found to be a strong environmental selective pressure for *Arabidopsis* (Yerushalmi *et al.* 2011) and the plant-type secondary cell wall process. This process is involved in the integrity of the cell wall plants which help respond to exposure to biotic and abiotic stresses (Vaahtera *et al.* 2019). This was found to help with the resistance of pathogens (Miedes *et al.* 2014). Additionally, two other significant enrichments were found to help resistance of pathogens with the regulation of defence to bacterium and the regulation of defence in response to fungus. The importance of these functions of resistance is well documented in other species impacted by crop-to-wild gene flow (Presotto *et al.* 2012; Le Corre *et al.* 2020). A remarkable case is the teosinte, a wild relative of maize that became resistant to herbicide because of crop-to-wild-gene flow, leading to rapid invasion to crop and thereby becoming a noxious weed (Le Corre *et al.* 2020). Admixed local adaptation pattern evidenced is very different from the wild one, in the number of SNPs detected or in the enrichment GO term results, supporting the assumption that they experienced two different evolutionary trajectories.

4.3 | Differential local adaptation between wild and admixed olive trees: a potential help to cope with climate change?

Interestingly, only 14 outliers were found to be common between wild and admixed olive trees and some were even observed within the same environmental or geographical locations. This raises questions about the existence of different selection pressures between these groups. It is also possible that what we observe in the wild may reflect a significant selective pressure from a longer evolutionary history that allowed a more pronounced response to local selective pressures. The common genes detected as associated with the environment between both groups were more abundant than outliers with 34 genes shared, arguing in favour of a same selection pressure but just not impacting in the same part of genes or a different impact of the disequilibrium linkage. Those results support the assumption of Lavee & Zohary (2011) that wild olive trees, because of gene flow with the cultivated group, are led to the loss of their specific characteristic, developed over decades by natural selection, with a reduction of wild genetic diversity.

However, we do not observe particular selection on the outlier SNPs from the wild population within the admixed individuals, but we do know that they retain over 80% of this information. Considering that hybridization events occurred with closely related populations from the wild, that likely carried alleles adapted to the environment, they might be indirectly adapted. However, we cannot currently ascertain this due to the very recent hybridization. Furthermore, the potential presence of two groups of admixed individuals further complicates the situation, as it introduces additional structure, as evidenced by the LFMM analysis. This genetic diversity contribution might be beneficial for the adaptive potential of admixed group. The admixed genetic group, by keeping a genetic information from the wild and the cultivated, exhibit a potential higher standing genetic variation, with the presence of multiple alleles at a locus, which could enable a quicker response to a changing environment (Barrett & Schluter 2008). Gene flow can facilitate adaptive evolution in new environments by

increasing genetic variation or fitness (Kremer *et al.* 2012). These results align with the assumption that long-lived woody perennials species tend to hybridise more frequently, contributing to the high levels of genetic diversity and helping to colonize new habitat (Petit *et al.* 2004; Petit & Hampe 2006).

The gene of resistance to pathogens in admixed groups, could help to explain their important presence in the natural environment. The pathogens lead to a strong selective pressure, involving rapid adaptive change from the trees, possibly thanks to their increased level of genetic diversity (Petit & Hampe 2006). This assumption is supported with the arrival of *Xylella fastidiosa* in Europe, which may lead to a strong selective pressure on the natural occurring populations, for the admixed as well as the wild genetic groups. This bacterial disease affects the xylem and causes various diseases in multiple species such as citrus, grapes, and notably, the olive quick decline syndrome in olive trees. Originally from America, it was first reported in Europe, in the southern part of Italy, in 2013, affecting olive trees, before spreading across Europe, especially in Mediterranean regions such as Corsica, France, Spain, Portugal. (Saponari *et al.* 2013; Trkulja *et al.* 2022). *Xylella* has been observed in cultivated olive trees in Italy, France, Spain, and Portugal, while its presence in wild olive trees has been confirmed in Spain (Jeger *et al.* 2018; Trkulja *et al.* 2022) and may re shape the current distribution of naturally occurring population.

In the near future, adaptive potential of our population regarding climate change scenarios using genomic offset will be studied, as it was done in pearl millet (Rhoné *et al.* 2020). We could map the wild olive vulnerability based on the current genomic-climate association to predict their adaptive potential. In a second time, we could use the admixed genotypes on the same model to see if they have a different adaptive potential to the wild in the face of the future?

5 | Conclusion

In conclusion, we provided evidence that western genuine wild olive tree, *Olea europaea* var. *sylvestris* is locally adapted to its environment, mainly due to climatic temperature factors. In contrast, we did not find evidence for local adaptation for the admixed olive trees found in the same natural environment. However, admixed individuals still keep 80% of local adaptation information in their genome, the SNPs detected as locally associated to climate in wild genetic groups. In addition, they also bring new genetic diversity from the cultivated with pathogen resistance. The broadening of standing genetic variation resulting from this hybridization could enable a faster response to adaptation to new or changing environments. This study paves the way to the prediction of future mal-adaptation of natural olive trees populations within the western MB, and help in its conservation as well as its cultivation.

References

- Alberto FJ, Aitken SN, Alía R *et al.* (2013) Potential for evolutionary responses to climate change - evidence from tree populations. *Global change biology*, **19**, 1645–1661.
- Alfieri SM, Riccardi M, Menenti M *et al.* (2019) Adaptability of global olive cultivars to water availability under future Mediterranean climate. *Mitigation and Adaptation Strategies for Global Change*, **24**, 435–466.
- Argueso CT, Ferreira FJ, Kieber JJ (2009) Environmental perception avenues: the interaction of cytokinin and environmental response pathways. *Plant, cell & environment*, **32**, 1147–1160.
- Barrett RDH, Schluter D (2008) Adaptation from standing genetic variation. *Trends in ecology & evolution*, **23**, 38–44.
- Besnard G, Khadari B, Navascués M *et al.* (2013) The complex history of the olive tree: From late quaternary diversification of mediterranean lineages to primary domestication in the northern Levant. *Proceedings of the Royal Society B: Biological Sciences*, **280**, 1–7.
- Blanquart F, Kaltz O, Nuismer SL, Gandon S (2013) A practical guide to measuring local adaptation. *Ecology letters*, **16**, 1195–1205.
- Brun, Zimmermann, Hari (2022) Global climate-related predictors at kilometre resolution for the past and future. *Earth system dynamics*.
- Calvin K, Dasgupta D, Krinner G *et al.* (2023) *IPCC, 2023: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth*

- Assessment Report of the Intergovernmental Panel on Climate Change [core writing team, H. lee and J. romero (eds.)]. IPCC, Geneva, Switzerland (P Arias, M Bustamante, I Elgizouli, et al., Eds.). IPCC, Geneva, Switzerland.*
- Capblancq T, Lachmuth S, Fitzpatrick MC, Keller SR (2023) From common gardens to candidate genes: exploring local adaptation to climate in red spruce. *The New phytologist*, **237**, 1590–1605.
- Carlson M (2022) *GO.db: A set of annotation maps describing the entire Gene Ontology*.
- Carrión Y, Ntinou M, Badal E (2010) *Olea europaea* L. in the North Mediterranean Basin during the Pleniglacial and the Early-Middle Holocene. *Quaternary science reviews*, **29**, 952–968.
- Caye K, Jumentier B, Lepeule J, François O (2019) LFMM 2: Fast and Accurate Inference of Gene-Environment Associations in Genome-Wide Studies. *Molecular biology and evolution*, **36**, 852–860.
- Chang CC, Chow CC, Tellier LC *et al.* (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, **4**, 7.
- Cingolani P, Platts A, Wang LL *et al.* (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly*, **6**, 80–92.
- Clark PU, Dyke AS, Shakun JD *et al.* (2009) The Last Glacial Maximum. *Science*, **325**, 710–714.
- Danecek P, Auton A, Abecasis G *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.
- Danielson JJ, Gesch DB (2011) Global multi-resolution terrain elevation data 2010 (GMTED2010).
- Fick SE, Hijmans RJ (2017) WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, **37**, 4302–4315.
- Fiorino P, Mancuso S (2000) Differential thermal analysis, supercooling and cell viability in organs of *Olea europaea* at subzero temperatures. *Advances in horticultural science*, **14**, 23–27.
- Frichot E, François O (2015) LEA: An R package for landscape and ecological association studies. *Methods in ecology and evolution / British Ecological Society*, **6**, 925–929.
- Goudet J, Jombart T (2022) Hierfstat: Estimation and tests of hierarchical F-statistics.
- Grivet D, Sebastiani F, Alía R *et al.* (2011) Molecular footprints of local adaptation in two Mediterranean conifers. *Molecular biology and evolution*, **28**, 101–116.
- Haberman A, Bakhshian O, Cerezo-Medina S *et al.* (2017) A possible role for flowering locus T-encoding genes in interpreting environmental and internal cues affecting olive (*Olea europaea* L.) flower induction. *Plant, cell & environment*, **40**, 1263–1280.
- Harris I, Osborn TJ, Jones P, Lister D (2020) Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Scientific Data*, **7**, 1–18.
- Hoban S, Kelley JL, Lotterhos KE *et al.* (2016) Finding the Genomic Basis of Local Adaptation: Pitfalls, Practical Solutions, and Future Directions. *The American*

- naturalist*, **188**, 379–397.
- Hoffmann A, Griffin P, Dillon S *et al.* (2015) A framework for incorporating evolutionary genomics into biodiversity conservation and management. *Climate Change Responses*, **2**.
- Jeger M, Caffier D, Candresse T *et al.* (2018) Updated pest categorisation of *Xylella fastidiosa*. *EFSA journal*, **16**, e05357.
- Jordan R, Hoffmann AA, Dillon SK, Prober SM (2017) Evidence of genomic adaptation to climate in *Eucalyptus microcarpa*: Implications for adaptive potential to projected climate change. *Molecular ecology*, **26**, 6002–6020.
- Julca I, Marcet-Houben M, Cruz F *et al.* (2020) Genomic evidence for recurrent genetic admixture during the domestication of Mediterranean olive trees (*Olea europaea* L.). *BMC biology*, **18**, 1–25.
- Kaniewski D, Van Campo E, Boiy T *et al.* (2012) Primary domestication and early uses of the emblematic olive tree: Palaeobotanical, historical and molecular evidence from the Middle East. *Biological reviews of the Cambridge Philosophical Society*, **87**, 885–899.
- Karger DN, Conrad O, Böhner J *et al.* (2017) Climatologies at high resolution for the earth's land surface areas. *Scientific Data*, **4**, 1–20.
- Kassambara A (2023) *rstatix: Pipe-Friendly Framework for Basic Statistical Tests*.
- Kawecki TJ, Ebert D (2004) Conceptual issues in local adaptation. *Ecology letters*, **7**, 1225–1241.
- Köppen W (1900) Versuch einer Klassifikation der Klimate, vorzugsweise nach ihren Beziehungen zur Pflanzenwelt. *Geographische Zeitschrift*, **6**, 593–611.
- Kostopoulou Z, Therios I, Roumeliotis E, Kanellis AK, Molassiotis A (2015) Melatonin combined with ascorbic acid provides salt adaptation in *Citrus aurantium* L. seedlings. *Plant physiology and biochemistry: PPB / Societe francaise de physiologie vegetale*, **86**, 155–165.
- Kremer A, Ronce O, Robledo-Arnuncio JJ *et al.* (2012) Long-distance gene flow and adaptation of forest trees to rapid climate change. *Ecology letters*, **15**, 378–392.
- Lascoux M, Glémin S, Savolainen O (2016) Local adaptation in plants. *Elsevier Oceanography Series*, 1–7.
- Lavee S, Zohary D (2011) The potential of genetic diversity and the effect of geographically isolated resources in olive breeding. *Israel journal of plant sciences*, **59**, 3–13.
- Le Corre V, Siol M, Vigouroux Y, Tenaillon MI, Délye C (2020) Adaptive introgression from maize has facilitated the establishment of teosinte as a noxious weed in Europe. *Proceedings of the National Academy of Sciences of the United States of America*, **117**, 25618–25627.
- Leimu R, Fischer M (2008) A meta-analysis of local adaptation in plants. *PloS one*, **3**, e4010.
- Leroy T, Louvet JM, Lalanne C *et al.* (2020) Adaptive introgression as a driver of local adaptation to climate in European white oaks. *The New phytologist*, **226**, 1171–1182.
- Li L-F, Cushman SA, He Y-X, Li Y (2020) Genome sequencing and population

- genomics modeling provide insights into the local adaptation of weeping forsythia. *Horticulture Research*, **7**, 1–12.
- Lionello P, Abrantes F, Congedi L *et al.* (2012) Introduction: Mediterranean Climate—Background Information. In: *The Climate of the Mediterranean Region* (ed Lionello P), pp. xxxv–xc. Elsevier, Oxford.
- Lionello P, Scarascia L (2018) The relation between climate change in the Mediterranean region and global warming. *Regional Environmental Change*, **18**, 1481–1493.
- Lotterhos KE, Whitlock MC (2015) The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Molecular ecology*, **24**, 1031–1046.
- Lütz C, Engel L (2007) Changes in chloroplast ultrastructure in some high-alpine plants: adaptation to metabolic demands and climate? *Protoplasma*, **231**, 183–192.
- Medail F, Quezel P (1997) Hot-Spots Analysis for Conservation of Plant Biodiversity in the Mediterranean Basin. *Annals of the Missouri Botanical Garden. Missouri Botanical Garden*, **84**, 112–127.
- Miedes E, Vanholme R, Boerjan W, Molina A (2014) The role of the secondary cell wall in plant resistance to pathogens. *Frontiers in plant science*, **5**, 358.
- Oksanen J, Simpson GL, Blanchet FG *et al.* (2022) *vegan: Community Ecology Package*.
- Ozenda P (1975) Sur les étages de végétation dans les montagnes du bassin méditerranéen. *Documents de cartographie écologique*, **XVI**, 1–32.
- Pagès H, Carlson M (2023) *AnnotationDbi: Manipulation of SQLite-based annotations in Bioconductor*.
- Petit RJ, Bodénès C, Ducouso A, Roussel G, Kremer A (2004) Hybridization as a mechanism of invasion in oaks. *The New phytologist*, **161**, 151–164.
- Petit RJ, Hampe A (2006) Some evolutionary consequences of being a tree. *Annual review of ecology, evolution, and systematics*, **37**, 187–214.
- Petrucelli R, Bartolini G, Ganino T *et al.* (2022) Cold Stress, Freezing Adaptation, Varietal Susceptibility of *Olea europaea* L.: A Review. *Plants*, **11**.
- Planchais N (1984) Palynologie lagunaire: l'exemple du Languedoc-Roussillon. *Annales de Géographie*, **93**, 268–275.
- Presotto A, Ureta MS, Cantamutto M, Poverene M (2012) Effects of gene flow from IMI resistant sunflower crop to wild *Helianthus annuus* populations. *Agriculture, ecosystems & environment*, **146**, 153–161.
- Purcell S, Chang C (2015) *PLINK*.
- Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.
- R Core Team (2022) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Ramírez-Valiente JA, Santos del Blanco L, Alía R, Robledo-Arnuncio JJ, Climent J (2022) Adaptation of Mediterranean forest species to climate: Lessons from common garden experiments. *The Journal of ecology*, **110**, 1022–1042.

CHAPTER 3

- Rellstab C, Gugerli F, Eckert AJ, Hancock AM, Holderegger R (2015) A practical guide to environmental association analysis in landscape genomics. *Molecular ecology*, **24**, 4348–4370.
- Rhoné B, Defrance D, Berthouly-Salazar C *et al.* (2020) Pearl millet genomic vulnerability to climate change in West Africa highlights the need for regional collaboration. *Nature communications*, **11**, 1–9.
- Rivas Martínez S, Gandullo Gutiérrez JM, Allué Andrade JL, Montero de Burgos JL, González Rebollar JL (1987) *Memoria del Mapa de series de vegetación de España 1:400.000*.
- Rubio de Casas R, Besnard G, Schönswetter P, Balaguer L, Vargas P (2006) Extensive gene flow blurs phylogeographic but not phylogenetic signal in *Olea europaea* L. *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik*, **113**, 575–583.
- Saponari M, Boscia D, Nigro F, Martelli GP, Others (2013) Identification of DNA sequences related to *Xylella fastidiosa* in oleander, almond and olive trees exhibiting leaf scorch symptoms in Apulia (Southern Italy). *Journal of plant pathology: an international journal of the Italian Phytopathological Society*, **95**.
- Saumitou-Laprade P, Vernet P, Vekemans X *et al.* (2017) Elucidation of the genetic architecture of self-incompatibility in olive: Evolutionary consequences and perspectives for orchard management. *Evolutionary applications*, **10**, 867–880.
- Savolainen O, Lascoux M, Merilä J (2013) Ecological genomics of local adaptation. *Nature reviews. Genetics*, **14**, 807–820.
- Sork VL (2018) Genomic Studies of Local Adaptation in Natural Plant Populations. *The Journal of heredity*, **109**, 3–15.
- Sork VL, Aitken SN, Dyer RJ *et al.* (2013) Putting the landscape into the genomics of trees: Approaches for understanding local adaptation and population responses to changing climate. *Tree genetics & genomes*, **9**, 901–911.
- Spennemann DHR, Richard Allen L (2000) The avian dispersal of olives *Olea europaea*: Implications for Australia. *The Emu*, **100**, 264–273.
- Steane DA, Potts BM, McLean E *et al.* (2014) Genome-wide scans detect adaptation to aridity in a widespread forest tree species. *Molecular ecology*, **23**, 2500–2513.
- Storey, Jd B, Aj D, A, Robinson, D (2022) *qvalue: Q-value estimation for false discovery rate control*.
- Sun X, Han G, Meng Z, Lin L, Sui N (2019) Roles of malic enzymes in plant development and stress responses. *Plant signaling & behavior*, **14**, e1644596.
- Terral J-F, Durand A, Newton C, Ivorra S (2009) Archéo-biologie de la domestication de l'olivier en Méditerranée occidentale : de la remise en cause d'une histoire dogmatique à la révélation de son irrigation médiévale. *Etudes*, 1–19.
- Trkulja V, Tomić A, Iličić R, Nožinić M, Milovanović TP (2022) *Xylella fastidiosa* in Europe: From the Introduction to the Current Status. *The plant pathology journal / the Korean Society of Plant Pathology*, **38**, 551–571.
- Vaahtera L, Schulz J, Hamann T (2019) Cell wall integrity maintenance during plant development and interaction with the environment. *Nature plants*, **5**, 924–932.

- Venzhik YV, Shchyogolev SY, Dykman LA (2019) Ultrastructural Reorganization of Chloroplasts during Plant Adaptation to Abiotic Stress Factors. *Russian journal of plant physiology: a comprehensive Russian journal on modern phytophysiology*, **66**, 850–863.
- de Villemereuil P, Frichot É, Bazin É, François O, Gaggiotti OE (2014) Genome scan methods against more complex models: when and how much should we trust them? *Molecular ecology*, **23**, 2006–2019.
- Wu T, Hu E, Xu S *et al.* (2021) clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Cambridge (Mass.))*, **2**, 100141.
- Yerushalmi S, Yakir E, Green RM (2011) Circadian clocks and adaptation in *Arabidopsis*. *Molecular ecology*, **20**, 1155–1165.
- Yu G (2023) *enrichplot: Visualization of Functional Enrichment Result*.
- Zohary D, Spiegel-roy P (1975) Beginnings of Fruit Growing in the Old World. *American Association for the Advancement of Science*, **187**, 319–327.
- Zunino L, Cubry P, Sarah G *et al.* (2023a) Genomic evidence of genuine wild versus admixed olive populations evolving in the same natural environments in western Mediterranean Basin. *bioRxiv*, 2023.04.18.537296.
- Zunino L, Sarah G, Cubry P, Khadari B (2023b) Selection footprints in *Olea europaea* subsp. *europaea*: Detection in Wild, Admixed and Cultivated olive trees from the Western Mediterranean Basin - Clues to their evolutionary history ?

Discussion



The history of the olive tree within the Mediterranean Basin has been studied extensively for wild olives, using various tools such as charcoal, pollen or DNA markers (Carrión *et al.* 2010; Besnard *et al.* 2013b; a, 2018; Diez *et al.* 2015; Khadari & El Bakkali 2018; Gros-Balthazard *et al.* 2019; Langgut *et al.* 2019; Julca *et al.* 2020). These studies have often focused on questions related to olive domestication, using wild olives as its crop wild relative. This PhD aimed to look at the evolutionary history of the wild olive in the western MB, through the study of genetic structure, admixture, detection of footprints of selection, and genome-environment associations analysis.

1 | Confirmation of previous results on olive genetic structure and new insights

Chapter 1 confirmed various results previously highlighted in earlier studies. Indeed, the study of genetic structure using the most extensive genomic dataset to date has brought to light various findings that confirm the previously hypothesized evolutionary history and domestication of the olive tree, *Olea europaea* L., within the Mediterranean Basin.

In the case of the wild olive trees, our study confirmed a strong genetic differentiation between eastern MB and western MB populations, supporting the assumptions of the presence of two gene pools distributed along an east-west longitudinal gradient, as previously observed in earlier research (Besnard *et al.* 2001, 2013a; b; Diez *et al.* 2015), a differentiation pattern already observed in other Mediterranean species (Arroyo-García *et al.* 2006; Rodríguez-Sánchez *et al.* 2009). This pattern may result from the Messinian Crisis, creating new pathways of colonization during the desiccation of the Mediterranean Sea (Terral *et al.* 2009), followed by a long evolutionary history of differentiation within the Mediterranean Basin (Blondel *et al.* 2010).

As demonstrated by previous studies on domestication (Besnard *et al.* 2013b; Khadari & El Bakkali 2018; Gros-Balthazard *et al.* 2019), the wild olive trees in the east are genetically very close to the cultivated olive (Besnard *et al.* 2013b; Khadari & El Bakkali 2018; Gros-Balthazard *et al.* 2019). This supports the hypothesis of domestication likely occurring from the wild olive trees in the eastern part of the MB, specifically in the Levant (Zohary & Spiegel-roy 1975; Kaniewski *et al.* 2012; Besnard *et al.* 2013b).

DISCUSSION

Through this thesis, we have brought novel insights concerning naturally occurring populations of olive in the western MB. Specifically, we identified genuine wild olive populations in the western region. Among the 26 populations analysed, 16 contained wild individuals with 11 of these populations consisting of at least 50% wild individuals. Genuine wild trees were encountered in all the sampled countries, including for the first time in continental France, in two geographical opposite locations of Mont-Boron Nice (east) and Leucate (west). These new findings corroborate some prior research outcomes (Besnard *et al.* 2013a) while simultaneously revealing the existence of wild olives in regions not previously confirmed but only hypothesized from botanical studies (Carrión *et al.* 2010; Lavee & Zohary 2011; Langgut *et al.* 2019).

Contrary to our working hypothesis, our sampling, which was based on prior characterizations of wild olive populations, involving nuclear DNA microsatellites and plastid DNA variation (Besnard *et al.* 2013a), has demonstrated a considerable presence of individuals that have resulted from hybridization (admixed olive trees) with cultivated olive in natural environment. This outcome was unexpected, as these regions were previously reported to be less affected by crop-to-wild gene flow (gene flow from the cultivated olives to the wild relatives), in comparison to the eastern MB (Besnard *et al.* 2013a; b; Khadari & El Bakkali 2018). The coexistence of these three groups, wild, admixed and cultivated olive, within the same distribution area has given rise to numerous inquiries that the three chapters of this PhD have sought to address, through the investigation of local adaptation, combining selection footprints and genome-environment association approaches.

2 | Evolutionary history of the wild, cultivated and admixed olives

The history of olive trees within the MB appears to be complex and highly interconnected between wild and cultivated olives. The chapter 2 of this PhD demonstrates the existence of gene flows between these two compartments. These gene flows appear to exist for years, and it is postulated that gene flow from local wild populations into cultivated ones has accelerated their colonization from the eastern MB to the western MB. This process may also result in genetic structuring within the cultivated groups across the MB (Lumaret *et al.* 2004; Besnard *et al.* 2013b; Julca *et al.* 2020).

Through chapter 1, we highlighted the substantial gene flow between these two compartments, resulting in the emergence of a third category as a consequence of their hybridization: the admixed olive trees. The significant presence of admixed olives in the western MB raises questions about their role in the evolutionary history of olive trees in this region. Without a proper understanding of this group, one might assume that wild olives are a highly heterogeneous group with an evolutionary history very similar to cultivated olives. The identification of this admixed group serves to underscore the marked genetic differentiation between wild and cultivated olives and the emergence of a genetic group in the western MB that is highly distinct from the rest of the olive trees, possessing different genetic resources and appearing to be influenced by the presence of cultivated olives in its environment.

The study of selection footprints has revealed differences in the evolutionary history among the three genetic compartments. In Chapter 2, it became evident that both the wild and cultivated compartments have experienced strong selective pressures, while very few selection signals are detected in the admixed olive group. In contrast, the wild olives display signs of both ancient and more recent evolutionary events, while the cultivated group likely reflects an ancient history rooted in the eastern MB, along with selection signals resulting from human artificial selection and local adaptation, supplemented by local gene flow from the wild compartment, although these signals are likely limited due to the long generation time of olive trees. In practice, Gros-Balthazard *et al.* (2019) evidenced a very weak signal of genomic variation between wild and cultivated olive while they argued in favour of gene expression as key drivers of domestication process. Chapter 3 exhibits similar results as Chapter 2 regarding selection footprints and reveals different patterns of genetic differentiation on a gradient scale. The gradient appears to be structured into two genetic groups for the admixed olives, as opposed to three in the case of the wild olives. This divergence could be attributed to either isolation by distance along the gradient or varying levels of introgression, likely linked to differences in genetic introgression from cultivated olives. The presence of two genetic groups among the admixed olives could indeed reflect the variations in genetic introgression from both the cultivated parent group and the wild parent group (Julca *et al.* 2020). Within our genetic group of cultivated olives, there are two genetic subdivisions within the western MB, which probably correlate with local domestication history with an earlier introduction in the southern Iberian Peninsula (Langgut *et al.* 2019; Khadari *et al.*

DISCUSSION

2019) and at least a probable more recent spread to southern France and northern Spain (Leveau *et al.* 1991; Langgut *et al.* 2019).

In perspective, we proposed to use the results here and the data from our study to do two different studies to help understand the evolutionary history of wild, admixed and cultivated olives in the MB. A first study could undertake a rigorous examination of introgression patterns between cultivated, wild, and admixed olives. The objective is to quantify the extent of gene flow and its implications on their respective genomes. Advanced analytical tools like Local PCA analysis may be used to look at the genetic build-up of each genotype along the genome (Li & Ralph 2018). Defining more homogeneous genetic groups and their extensive gene flow would help to do a second analysis of demographic history, using coalescence modelling software such as FastSimCoal2 (Excoffier *et al.* 2021). Using the genomic dataset generated in our study would provide a better understanding of potential evolutionary history scenarios for each of these three groups.

3 | Selection footprints and genome-environment association

The use of target sequencing on a substantial number of individuals within natural populations along a latitudinal gradient has enabled the investigation of the genomic variation along the gradient. While using whole-genome sequencing could have provided more insights, the targeted sequencing of annotated genes has demonstrated that, even with a relatively small dataset, one can still be able to observe significant genetic variation among individuals of a same species or population (Van Andel *et al.* 2019; Slimp *et al.* 2021). Our dataset is considerably larger and encompasses various genomics regions, much more than in these previous studies (Van Andel *et al.* 2019). This extensive SNP dataset appears to be adequate for examining local adaptation and has significantly expanded the number of populations and individuals studied along the gradient, thereby enhancing the robustness of our analyses.

Chapter 2 et 3 have provided insights into the potential local adaptation of wild olive trees to climate using two distinct approaches. The first approach relied on the detection of selection footprints without a priori assumptions about the source of selection. The second method was based on the identification of genomic variation in association with environmental variations of each population (GEA).

The results of each analysis allowed to detect, among the wild olive trees, 233 SNPs detected by both GEA (Chapter 3, 1,400 SNPs in total) and PCAdapt (Chapter 2, 1168 SNPs in total), with 186 of them associated with temperature related variables according to GEA. As for the admixed group, out of the 1,379 SNPs detected by GEA (Chapter 3), only 13 are in common with the 36 SNPs detected by PCAdapt (Chapter 2). However, a more detailed examination of this intersection between the data sets could potentially reveal the most significant selection factors for each group. It might also identify other potential selection factors that were not tested in this study and allow for a finer-scale assessment of local adaptation in olive trees and to detect the most important drivers related to their environmental adaptation.

The differences observed between the two groups in Chapters 2 and 3 likely provide, in part, from the variation in the methods used. The selection signals detected among the wild and admixed olives in Chapter 2 could indeed signify local adaptation. However, as explained by Capblancq *et al.* (2018), this method may not be effective if the environmental gradient associated with adaptation is correlated with population structure. Given the complex evolutionary histories of each group, it can be challenging to establish the presence of such correlation. Therefore, it may be more appropriate to consider the use of genome-environment association analysis to validate the potential adaptive significance of these SNPs (Capblancq *et al.* 2018).

The use of the GEA approach highlights other selection patterns, where temperature appears to drive the spatial distribution. Each of these methods accounts for population structure as a confounding factors (de Villemereuil *et al.* 2014; Rellstab *et al.* 2015; Hoban *et al.* 2016) . However, in the case of admixed olives, the genetic heterogeneity within the group can introduce substantial bias into the results. The correction for population structure in this group may be either too strong or too weak, potentially leading to false positives or, conversely, a lack of results. In this context, investigating patterns of introgression would provide better corrections for our models and enable the creation of more genetically homogeneous groups. Subsequent analyses of detection of selection footprints and genome-environment association on them could then be performed to determine if there are significant differences compared to the results obtained here.

4 | Local adaptation and adaptive potential in front of climate change

Results of Chapter 2 and 3 showed that local adaptation of wild olive trees is particularly driven by temperature variations along the gradient within the MB. On the opposite, the admixed olive trees, which are prevalent in the distribution area, do not necessarily indicate genuine adaptation to their environment. It is possible that the admixed genotypes either are not locally adapted, or that the detection of selection signals based on environmental cues is obscured by their recent hybridization history. These hybridization events may have disrupted or blurred the signals of selection due to the mixing of different genetic backgrounds detected in the Chapter 1 and 2 (de Villemereuil *et al.* 2014; Rellstab *et al.* 2015). However, a significant finding is that the proportion of wild genetic material detected as locally adapted, remains substantial within admixed genomes, particularly when compared to the limited presence of wild populations, especially in Spain. This could be attributed to either an extensive gene flow from the admixed populations or to a retention of wild genome parts within the admixed genotypes. A closer examination of the wild associated SNPs structure patterns reveals a similar pattern between admixed and wild olive trees along the climatic gradient. This observation further suggests that admixed olive trees may indeed retain some of the information related to local adaptation from their wild heritage.

The expansion of standing genetic variation resulting from this hybridization has the potential to facilitate a more rapid response of adaptation of olives in novel or changing environments (Barrett & Schluter 2008; Kremer *et al.* 2012). In perspective, we intend to investigate the adaptive potential of our populations in response to climate change scenarios using the genomic offset approach (Rhoné *et al.* 2020; Gain *et al.* 2023). We could map the wild olive vulnerability based on the current genome-environment association to predict their adaptive potential. In a second time, we could use the admixed genotypes on the same model to see if they have a different adaptive potential to the wild in the face of global change?

5 | Importance of conservation of wild olive trees and potential impact of this research on agricultural practices

The work presented in this thesis emphasizes the importance of conserving wild olive trees and their genetic legacy. Local adaptation in wild olive trees, as well as their persistence in admixed populations, highlights the significance of their genetic contribution in establishing new populations within the MB. The presence of potential adaptive alleles, for example, to warmer temperatures, could facilitate the establishment of these wild populations through assisted migration in future ecological niches corresponding to their current climatic conditions. Similarly, the creation of an *ex-situ* collection would allow for the preservation of the genetic resources of wild olive trees and concurrently a setting-up of a common garden as a support for future research on adaptive traits. Such a collection is planned to be established by the Conservatoire Botanique National Méditerranéen de Porquerolles, using the identification of wild individuals from this thesis, on the Island of Porquerolles in the south of France. Furthermore, recognizing wild olive trees as a species to protect would enable the implementation of conservation measures for existing populations, which may face the threat of destruction due to urbanization. This second project is also underway with the Conservatoire Botanique National Méditerranéen de Porquerolles, which aims to protect threatened Mediterranean species and their natural habitat such as wild olive and leverage the genetic relatedness of wild olive trees to cultivated ones to promote their protection.

This PhD also provides new insights into olive agricultural practices. Presently, conventional olive agricultural practices often involve the cultivation of a dominant variety, such as the Picholine marocaine variety in Morocco. In recent years, a decrease in agricultural production has been observed, linked to rising temperatures, insufficiently cold winters, and hotter summers. The year 2023 has been particularly concerning for the olive industry, with, for example, a nearly 33% decrease in production for Morocco compared to 2022 (Aublanc 2023). In Spain, droughts have also led to nearly 50% production losses in 2022 (Loiseau 2023).

New intensive harvesting methods necessitate olive tree irrigation during dry periods and promote the spread of pathogens within orchards. These pathogens, in turn, are favoured by increasing temperatures (Loiseau 2023). The significant presence of admixed olive trees, which contain a persistent portion of wild olive

DISCUSSION

adaptation, and the gene flows that have occurred over thousands of years, likely facilitated the establishment of cultivated olives in local regions. This suggests that using different olive varieties in orchards and promoting gene flow from wild to cultivated olives could lead to better adaptation of cultivated varieties to climate change. This, in turn, might reduce the need for irrigation in a context of decreasing water resources, which are essential for life.

General conclusion



GENERAL CONCLUSION

The research conducted in this PhD sheds light on the adaptation mechanisms of a perennial species to its environment in a region undergoing significant environmental changes. The study of the olive tree, *Olea europaea* L. in the western Mediterranean Basin, reveals a complex evolutionary history, divided into different genetic compartments, each with its own distinct evolutionary history. The use of target sequencing on hundreds of olive trees along a north-south climatic gradient in the western Mediterranean Basin has provided insights into these evolutionary histories. Through Chapter 1, we assessed the genetic diversity and structure of naturally occurring populations of olive trees in the western Mediterranean Basin. We were able to confirm previous study results and highlight the presence of wild olive trees in France, Spain, and Morocco. Interestingly, we provided the first genomic evidence for the persistence of olive trees of wild ancestry in Continental France. We also found many admixed populations resulting from strong crop-to-wild gene flow. The presence of admixed olive populations in the same distribution area as genuine wild populations raises questions on the reasons for their predominance in the natural environment. We aimed to address these questions in the following chapters, while also investigating local adaptation in wild olives.

Chapter 2 allowed us to explore the complex evolutionary history of the olive trees from the Mediterranean Basin, associated with three observed genetic groups: wild, cultivated, and admixed. Despite significant differences, these groups have evolved in interconnected ways, raising questions about the importance of these connections between the three clusters over time. These results led to refine hypotheses on demographic histories of this species with the perspective to develop new approaches in the future with demographic history modeling and introgression analysis.

Chapter 3 revealed the local adaptation of wild olive trees to their environment, particularly driven by the temperature variations. In contrast, our finding did not provide evidence for local adaptation of the admixed olive trees found in the same natural environment. However, admixed individuals still retain 80% of the SNPs detected as locally associated to climate in wild genetic groups. In addition, their mosaic genomes also bring genetic diversity acquired from the cultivated within the natural environment. The increase in standing genetic variation in this compartment as compared to the wild one due to hybridization could potentially facilitate a more rapid response to selection in the face of global change.

GENERAL CONCLUSION

All the results of this PhD will provide foundation for new studies on an emblematic perennial species of an environment very impacted by global change. These studies could encompass research on introgression, demographic history, and predictions of adaptation to climate change. The results obtained here, as well as future research, have the potential to help adapt current conservation methods and agricultural practices in the olive sector.

References

- Aitken SN, Yeaman S, Holliday JA, Wang T, Curtis-McLane S (2008) Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evolutionary applications*, **1**, 95–111.
- Alberto FJ, Aitken SN, Alía R *et al.* (2013) Potential for evolutionary responses to climate change - evidence from tree populations. *Global change biology*, **19**, 1645–1661.
- Alfieri SM, Riccardi M, Menenti M *et al.* (2019) Adaptability of global olive cultivars to water availability under future Mediterranean climate. *Mitigation and Adaptation Strategies for Global Change*, **24**, 435–466.
- Anderson D, Negishi Y, Ishiniwa H *et al.* (2021) Introgression dynamics from invasive pigs into wild boar following the March 2011 natural and anthropogenic disasters at Fukushima. *Proceedings. Biological sciences / The Royal Society*, **288**, 20210874.
- Argueso CT, Ferreira FJ, Kieber JJ (2009) Environmental perception avenues: the interaction of cytokinin and environmental response pathways. *Plant, cell & environment*, **32**, 1147–1160.
- Arroyo-García R, Ruiz-García L, Bolling L *et al.* (2006) Multiple origins of cultivated grapevine (*Vitis vinifera* L. ssp. *sativa*) based on chloroplast DNA polymorphisms. *Molecular ecology*, **15**, 3707–3714.
- Aublanc A (Ed.) (2023) *Le Maroc limite l'exportation de son huile d'olive pour enrayer la hausse des prix*. Le Monde Afrique.
- Barrett RDH, Schluter D (2008) Adaptation from standing genetic variation. *Trends in ecology & evolution*, **23**, 38–44.
- Benito Garzón M, Alía R, Robson TM, Zavala MA (2011) Intra-specific variability and plasticity influence potential tree species distributions under climate change. *Global ecology and biogeography: a journal of macroecology*, **20**, 766–778.
- Benton MJ, Twitchett RJ (2003) How to kill (almost) all life: The end-Permian extinction event. *Trends in ecology & evolution*, **18**, 358–365.
- Besnard G, Bakkali AE, Haouane H *et al.* (2013a) Population genetics of Mediterranean and Saharan olives: Geographic patterns of differentiation and evidence for early generations of admixture. *Annals of botany*, **112**, 1293–1302.
- Besnard G, Baradat P, Breton C, Khadari B, Bervillé A (2001) Olive domestication from structure of oleasters and cultivars using nuclear RAPDs and mitochondrial RFLPs. *Genetics, selection, evolution: GSE*, **33**, S251.
- Besnard G, Garcia-Verdugo C, De Casas RR *et al.* (2008) Polyploidy in the olive complex (*Olea europaea*): evidence from flow cytometry and nuclear microsatellite analyses. *Annals of botany*, **101**, 25–30.
- Besnard G, Henry P, Wille L, Cooke D, Chapuis E (2007) On the origin of the invasive olives (*Olea europaea* L., Oleaceae). *Heredity*, **99**, 608–619.
- Besnard G, Khadari B, Navascués M *et al.* (2013b) The complex history of the olive tree: From late quaternary diversification of mediterranean lineages to primary

REFERENCES

- domestication in the northern Levant. *Proceedings of the Royal Society B: Biological Sciences*, **280**, 1–7.
- Besnard G, Terral JF, Cornille A (2018) On the origins and domestication of the olive: A review and perspectives. *Annals of botany*, **121**, 385–403.
- Blanquart F, Kaltz O, Nuismer SL, Gandon S (2013) A practical guide to measuring local adaptation. *Ecology letters*, **16**, 1195–1205.
- Blondel J, Aronson J, Bodiou J-Y, Boeuf G (2010) *The Mediterranean Region: Biological Diversity in Space and Time*. Oxford University Press.
- Bourguiba H, Audergon J-M, Krichen L *et al.* (2012) Loss of genetic diversity as a signature of apricot domestication and diffusion into the Mediterranean Basin. *BMC plant biology*, **12**, 49.
- Brachi B, Villoutreix R, Faure N *et al.* (2013) Investigation of the geographical scale of adaptive phenological variation and its underlying genetics in *Arabidopsis thaliana*. *Molecular ecology*, **22**, 4222–4240.
- Breton C, Tersac M, Bervillé A (2006) Genetic diversity and gene flow between the wild olive (oleaster, *Olea europaea* L.) and the olive: Several Plio-Pleistocene refuge zones in the Mediterranean basin suggested by simple sequence repeats analysis. *Journal of biogeography*, **33**, 1916–1928.
- Brun, Zimmermann, Hari (2022) Global climate-related predictors at kilometre resolution for the past and future. *Earth system dynamics*.
- Burgarella C, Barnaud A, Kane NA *et al.* (2019) Adaptive Introgression: An Untapped Evolutionary Mechanism for Crop Adaptation. *Frontiers in plant science*, **10**, 4.
- Calvin K, Dasgupta D, Krinner G *et al.* (2023) *IPCC, 2023: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [core writing team, H. lee and J. romero (eds.)]. IPCC, Geneva, Switzerland* (P Arias, M Bustamante, I Elgizouli, *et al.*, Eds.). IPCC, Geneva, Switzerland.
- Capblancq T, Lachmuth S, Fitzpatrick MC, Keller SR (2023) From common gardens to candidate genes: exploring local adaptation to climate in red spruce. *The New phytologist*, **237**, 1590–1605.
- Capblancq T, Luu K, Blum MGB, Bazin E (2018) Evaluation of redundancy analysis to identify signatures of local adaptation. *Molecular ecology resources*, **18**, 1223–1233.
- Carlson M (2022) *GO.db: A set of annotation maps describing the entire Gene Ontology*.
- Carrión Y, Ntinou M, Badal E (2010) *Olea europaea* L. in the North Mediterranean Basin during the Pleniglacial and the Early-Middle Holocene. *Quaternary science reviews*, **29**, 952–968.
- Caye K, Jumentier B, Lepeule J, François O (2019) LFMM 2: Fast and Accurate Inference of Gene-Environment Associations in Genome-Wide Studies. *Molecular biology and evolution*, **36**, 852–860.
- Chang CC, Chow CC, Tellier LC *et al.* (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, **4**, 7.

- Chenault N, Arnaud-Haond S, Juteau M *et al.* (2011) SSR-based analysis of clonality, spatial genetic structure and introgression from the Lombardy poplar into a natural population of *Populus nigra* L. along the Loire River. *Tree genetics & genomes*, **7**, 1249–1262.
- Cingolani P, Platts A, Wang LL *et al.* (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly*, **6**, 80–92.
- Clark PU, Dyke AS, Shakun JD *et al.* (2009) The Last Glacial Maximum. *Science*, **325**, 710–714.
- Copernicus Land Monitoring Service (2020) CORINE Land Cover. *Land Copernicus Eu*.
- Cornille A, Antolín F, Garcia E *et al.* (2019) A Multifaceted Overview of Apple Tree Domestication. *Trends in plant science*, **24**, 770–782.
- Cruz F, Julca I, Gómez-Garrido J *et al.* (2016) Genome sequence of the olive tree, *Olea europaea*. *GigaScience*, **5**.
- D’Agostino N, Taranto F, Camposeo S *et al.* (2018) GBS-derived SNP catalogue unveiled wide genetic variability and geographical relationships of Italian olive cultivars. *Scientific reports*, **8**, 15877.
- Danecek P, Auton A, Abecasis G *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.
- Danielson JJ, Gesch DB (2011) Global multi-resolution terrain elevation data 2010 (GMTED2010).
- Davis MB, Shaw RG (2001) Range Shifts and Adaptive Responses to Quaternary Climate Change. *Science*, **292**, 673–679.
- De Mita S, Thuillet AC, Gay L *et al.* (2013) Detecting selection along environmental gradients: Analysis of eight methods and their effectiveness for outbreeding and selfing populations. *Molecular ecology*, **22**, 1383–1399.
- Diez CM, Trujillo I, Martínez-Urdiroz N *et al.* (2015) Olive domestication and diversification in the Mediterranean Basin. *The New phytologist*, **206**, 436–447.
- Dingle H, Drake VA (2007) What is migration? *Bioscience*, **57**, 113–121.
- Duan N, Bai Y, Sun H *et al.* (2017) Genome re-sequencing reveals the history of apple and supports a two-stage model for fruit enlargement. *Nature communications*, **8**.
- Ellstrand NC, Prentice HC, Hancock JF (1999) Gene Flow and Introgression from Domesticated Plants into their Wild Relatives. *Annual review of ecology and systematics*, **30**, 539–563.
- Erdei B, Hably L, Kázmér M, Utescher T, Bruch AA (2007) Neogene flora and vegetation development of the Pannonian domain in relation to palaeoclimate and palaeogeography. *Palaeogeography, palaeoclimatology, palaeoecology*, **253**, 115–140.
- Excoffier L, Marchi N, Marques DA *et al.* (2021) fastsimcoal2: demographic inference under complex evolutionary scenarios. *Bioinformatics*, **37**, 4882–4885.
- Feurtey A, Cornille A, Shykoff JA, Snirc A, Giraud T (2017) Crop-to-wild gene flow and its fitness consequences for a wild fruit tree: Towards a comprehensive

REFERENCES

- conservation strategy of the wild apple in Europe. *Evolutionary applications*, **10**, 180–188.
- Fick SE, Hijmans RJ (2017) WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, **37**, 4302–4315.
- Fiorino P, Mancuso S (2000) Differential thermal analysis, supercooling and cell viability in organs of *Olea europaea* at subzero temperatures. *Advances in horticultural science*, **14**, 23–27.
- Fournier-Level A, Korte A, Cooper MD *et al.* (2011) A map of local adaptation in *Arabidopsis thaliana*. *Science*, **334**, 86–89.
- Frichot E, François O (2015) LEA: An R package for landscape and ecological association studies. *Methods in ecology and evolution / British Ecological Society*, **6**, 925–929.
- Gain C, Rhoné B, Cubry P *et al.* (2023) A Quantitative Theory for Genomic Offset Statistics. *Molecular biology and evolution*, **40**.
- Gering E, Incorvaia D, Henriksen R *et al.* (2019) Getting Back to Nature: Feralization in Animals and Plants. *Trends in ecology & evolution*, **34**, 1137–1151.
- Goudet J, Jombart T (2022) Hierfstat: Estimation and tests of hierarchical F-statistics.
- Green PS (2002) A Revision of *Olea* L. (Oleaceae). *Kew bulletin / Royal Botanic Gardens*, **57**, 91–140.
- Green PS, Wickens GE (1989) *The Olea europaea Complex* (EU Press, Ed.), Edinburg, UK.
- Grivet D, Sebastiani F, Alía R *et al.* (2011) Molecular footprints of local adaptation in two Mediterranean conifers. *Molecular biology and evolution*, **28**, 101–116.
- Gros-Balthazard M, Besnard G, Sarah G *et al.* (2019) Evolutionary transcriptomics reveals the origins of olives and the genomic changes associated with their domestication. *The Plant journal: for cell and molecular biology*, **100**, 143–157.
- Haberman A, Bakhshian O, Cerezo-Medina S *et al.* (2017) A possible role for flowering locus T-encoding genes in interpreting environmental and internal cues affecting olive (*Olea europaea* L.) flower induction. *Plant, cell & environment*, **40**, 1263–1280.
- Hampe A, Petit RJ (2005) Conserving biodiversity under climate change: the rear edge matters. *Ecology letters*, **8**, 461–467.
- Harris I, Osborn TJ, Jones P, Lister D (2020) Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Scientific Data*, **7**, 1–18.
- Hauser TP, Shim SI (2007) Survival and flowering of hybrids between cultivated and wild carrots (*Daucus carota*) in Danish grasslands. *Environmental biosafety research*, **6**, 237–247.
- Hoban S, Kelley JL, Lotterhos KE *et al.* (2016) Finding the Genomic Basis of Local Adaptation: Pitfalls, Practical Solutions, and Future Directions. *The American naturalist*, **188**, 379–397.
- Hoffmann A, Griffin P, Dillon S *et al.* (2015) A framework for incorporating evolutionary genomics into biodiversity conservation and management. *Climate Change Responses*, **2**.

- Howden MS, Soussana J-F, Tubiello FN *et al.* (2007) Adapting agriculture to climate change. *Proceedings of the National Academy of Sciences*, **104**, 19691–19696.
- Jeger M, Caffier D, Candresse T *et al.* (2018) Updated pest categorisation of *Xylella fastidiosa*. *EFSA journal*, **16**, e05357.
- Jelić M, Patenković A, Skorić M *et al.* (2015) Indigenous forests of European black poplar along the Danube River: genetic structure and reliable detection of introgression. *Tree genetics & genomes*, **11**, 89.
- Johnsson M, Gering E, Willis P *et al.* (2016) Feralisation targets different genomic loci to domestication in the chicken. *Nature communications*, **7**, 12950.
- Jordan R, Hoffmann AA, Dillon SK, Prober SM (2017) Evidence of genomic adaptation to climate in *Eucalyptus microcarpa*: Implications for adaptive potential to projected climate change. *Molecular ecology*, **26**, 6002–6020.
- Jouganous J, Long W, Ragsdale AP, Gravel S (2017) Inferring the joint demographic history of multiple populations: Beyond the diffusion approximation. *Genetics*, **206**, 1549–1567.
- Julca I, Marcet-Houben M, Cruz F *et al.* (2020) Genomic evidence for recurrent genetic admixture during the domestication of Mediterranean olive trees (*Olea europaea* L.). *BMC biology*, **18**, 1–25.
- Kaniewski D, Van Campo E, Boiy T *et al.* (2012) Primary domestication and early uses of the emblematic olive tree: Palaeobotanical, historical and molecular evidence from the Middle East. *Biological reviews of the Cambridge Philosophical Society*, **87**, 885–899.
- Karger DN, Conrad O, Böhner J *et al.* (2017) Climatologies at high resolution for the earth's land surface areas. *Scientific Data*, **4**, 1–20.
- Kassambara A (2023) *rstatix: Pipe-Friendly Framework for Basic Statistical Tests*.
- Kawecki TJ, Ebert D (2004) Conceptual issues in local adaptation. *Ecology letters*, **7**, 1225–1241.
- Keller M, Kollmann J, Edwards PJ (2000) Genetic introgression from distant provenances reduces fitness in local weed populations. *The Journal of applied ecology*, **37**, 647–659.
- Kelly M (2019) Adaptation to climate change through genetic accommodation and assimilation of plastic phenotypes. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **374**.
- Khadari B, El Bakkali A (2018) Primary Selection and Secondary Diversification: Two Key Processes in the History of Olive Domestication. *International Journal of Agronomy*, **2018**, 1–9.
- Khadari B, El Bakkali A, Essalouh L *et al.* (2019) Cultivated Olive Diversification at Local and Regional Scales: Evidence From the Genetic Characterization of French Genetic Resources. *Frontiers in plant science*, **10**, 1593.
- Köppen W (1900) Versuch einer Klassifikation der Klimate, vorzugsweise nach ihren Beziehungen zur Pflanzenwelt. *Geographische Zeitschrift*, **6**, 593–611.
- Kostopoulou Z, Therios I, Roumeliotis E, Kanellis AK, Molassiotis A (2015) Melatonin combined with ascorbic acid provides salt adaptation in *Citrus aurantium* L.

REFERENCES

- seedlings. *Plant physiology and biochemistry: PPB / Societe francaise de physiologie vegetale*, **86**, 155–165.
- Kremer A, Ronce O, Robledo-Arnuncio JJ *et al.* (2012) Long-distance gene flow and adaptation of forest trees to rapid climate change. *Ecology letters*, **15**, 378–392.
- Langgut D, Cheddadi R, Carrión JS *et al.* (2019) The origin and spread of olive cultivation in the Mediterranean Basin: The fossil pollen evidence. *Holocene*, **29**, 902–922.
- Lascoux M, Glémin S, Savolainen O (2016) Local adaptation in plants. *Elsevier Oceanography Series*, 1–7.
- Lavee S, Zohary D (2011) The potential of genetic diversity and the effect of geographically isolated resources in olive breeding. *Israel journal of plant sciences*, **59**, 3–13.
- Le Corre V, Siol M, Vigouroux Y, Tenaillon MI, Délye C (2020) Adaptive introgression from maize has facilitated the establishment of teosinte as a noxious weed in Europe. *Proceedings of the National Academy of Sciences of the United States of America*, **117**, 25618–25627.
- Leimu R, Fischer M (2008) A meta-analysis of local adaptation in plants. *PloS one*, **3**, e4010.
- Lenormand T (2002) Gene flow and the limits to natural selection. *Trends in ecology & evolution*, **17**, 183–189.
- Leroy T, Louvet JM, Lalanne C *et al.* (2020) Adaptive introgression as a driver of local adaptation to climate in European white oaks. *The New phytologist*, **226**, 1171–1182.
- Leveau P, Heinz C, Laval H, Marinval P, Medus J (1991) Les origines de l'oléiculture en Gaule du Sud [Données historiques, archéologiques et botaniques]. *Revue d'archéométrie*, **15**, 83–94.
- Li L-F, Cushman SA, He Y-X, Li Y (2020) Genome sequencing and population genomics modeling provide insights into the local adaptation of weeping forsythia. *Horticulture Research*, **7**, 1–12.
- Li H, Ralph P (2018) Local PCA Shows How the Effect of Population Structure Differs Along the Genome. *Genetics*, **211**, 289–304.
- Lionello P, Abrantes F, Congedi L *et al.* (2012) Introduction: Mediterranean Climate—Background Information. In: *The Climate of the Mediterranean Region* (ed Lionello P), pp. xxxv–xc. Elsevier, Oxford.
- Lionello P, Malanotte-Rizzoli P, Boscolo R *et al.* (2006) The Mediterranean climate: An overview of the main characteristics and issues. In: *Developments in Earth and Environmental Sciences* (eds Lionello P, Malanotte-Rizzoli P, Boscolo R), pp. 1–26. Elsevier.
- Lionello P, Scarascia L (2018) The relation between climate change in the Mediterranean region and global warming. *Regional Environmental Change*, **18**, 1481–1493.
- Loiseau F (Ed.) (2023) *Le prix de l'huile d'olive flambe à cause de la sécheresse*. Reporterre.

- Lotterhos KE, Whitlock MC (2015) The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Molecular ecology*, **24**, 1031–1046.
- Lumaret R, Ouazzani N (2001) Ancient wild olives in Mediterranean forests. *Nature*, **413**, 700–700.
- Lumaret R, Ouazzani N, Michaud H *et al.* (2004) Allozyme variation of oleaster populations (wild olive tree) (*Olea europaea* L.) in the Mediterranean Basin. *Heredity*, **92**, 343–351.
- Lumaret R, Ouazzani N, Michaud H, Villemur P (1997) Cultivated olive and oleaster: Two closely connected partners of the same species (*Olea europaea* L.): Evidence from allozyme polymorphism. *Bocconea*, 39–42.
- Lütz C, Engel L (2007) Changes in chloroplast ultrastructure in some high-alpine plants: adaptation to metabolic demands and climate? *Protoplasma*, **231**, 183–192.
- Martin SH, Amos W (2020) Signatures of Introgression across the Allele Frequency Spectrum. *Molecular biology and evolution*, **38**, 716–726.
- Mary N, Iannuccelli N, Petit G *et al.* (2022) Genome-wide analysis of hybridization in wild boar populations reveals adaptive introgression from domestic pig. *Evolutionary applications*, **15**, 1115–1128.
- Medail F, Quezel P (1997) Hot-Spots Analysis for Conservation of Plant Biodiversity in the Mediterranean Basin. *Annals of the Missouri Botanical Garden. Missouri Botanical Garden*, **84**, 112–127.
- Miedes E, Vanholme R, Boerjan W, Molina A (2014) The role of the secondary cell wall in plant resistance to pathogens. *Frontiers in plant science*, **5**, 358.
- Mitchell-Olds T, Willis JH, Goldstein DB (2007) Which evolutionary processes influence natural genetic variation for phenotypic traits? *Nature reviews. Genetics*, **8**, 845–856.
- Neale DB, Kremer A (2011) Forest tree genomics: growing resources and applications. *Nature reviews. Genetics*, **12**, 111–122.
- Newton C, Lorre C, Sauvage C, Ivorra S, Terral JF (2014) On the origins and spread of *Olea europaea* L. (olive) domestication: Evidence for shape variation of olive stones at Ugarit, Late Bronze Age, Syria—a window on the Mediterranean Basin and on the westward diffusion of olive varieties. *Vegetation history and archaeobotany*, **23**, 567–575.
- Nielsen R (2005) Molecular signatures of natural selection. *Annual review of genetics*, **39**, 197–218.
- Nogués-Bravo D, Rodríguez-Sánchez F, Orsini L *et al.* (2018) Cracking the Code of Biodiversity Responses to Past Climate Change. *Trends in ecology & evolution*, **33**, 765–776.
- Oksanen J, Simpson GL, Blanchet FG *et al.* (2022) *vegan: Community Ecology Package*.
- Ozenda P (1975) Sur les étages de végétation dans les montagnes du bassin méditerranéen. *Documents de cartographie écologique*, **XVI**, 1–32.

REFERENCES

- Pagès H, Carlson M (2023) *AnnotationDbi: Manipulation of SQLite-based annotations in Bioconductor*.
- Pérez-López D, Ribas F, Moriana A, Rapoport HF, De Juan A (2008) Influence of temperature on the growth and development of olive (*Olea europaea* L.) trees. *The journal of horticultural science & biotechnology*, **83**, 171–176.
- Petit RJ, Bodénès C, Ducouso A, Roussel G, Kremer A (2004) Hybridization as a mechanism of invasion in oaks. *The New phytologist*, **161**, 151–164.
- Petit RJ, Hampe A (2006) Some evolutionary consequences of being a tree. *Annual review of ecology, evolution, and systematics*, **37**, 187–214.
- Petrucelli R, Bartolini G, Ganino T *et al.* (2022) Cold Stress, Freezing Adaptation, Varietal Susceptibility of *Olea europaea* L.: A Review. *Plants*, **11**.
- Planchais N (1984) Palynologie lagunaire: l'exemple du Languedoc-Roussillon. *Annales de Géographie*, **93**, 268–275.
- Presotto A, Ureta MS, Cantamutto M, Poverene M (2012) Effects of gene flow from IMI resistant sunflower crop to wild *Helianthus annuus* populations. *Agriculture, ecosystems & environment*, **146**, 153–161.
- Privé F, Luu K, Vilhjálmsson BJ, Blum MGB (2020) Performing Highly Efficient Genome Scans for Local Adaptation with R Package pcadapt Version 4. *Molecular biology and evolution*, **37**, 2153–2154.
- Purcell S, Chang C (2015) *PLINK*.
- Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.
- R Core Team (2022) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rallo L, Martin GC (1991) The Role of Chilling in Releasing Olive Floral Buds from Dormancy. *Journal of the American Society for Horticultural Science*, **116**, 1058–1062.
- Ramírez-Valiente JA, Santos del Blanco L, Alía R, Robledo-Arnuncio JJ, Climent J (2022) Adaptation of Mediterranean forest species to climate: Lessons from common garden experiments. *The Journal of ecology*, **110**, 1022–1042.
- Rellstab C, Fischer MC, Zoller S *et al.* (2017) Local adaptation (mostly) remains local: Reassessing environmental associations of climate-related candidate SNPs in *Arabidopsis halleri*. *Heredity*, **118**, 193–201.
- Rellstab C, Gugerli F, Eckert AJ, Hancock AM, Holderegger R (2015) A practical guide to environmental association analysis in landscape genomics. *Molecular ecology*, **24**, 4348–4370.
- Rhoné B, Defrance D, Berthouly-Salazar C *et al.* (2020) Pearl millet genomic vulnerability to climate change in West Africa highlights the need for regional collaboration. *Nature communications*, **11**, 1–9.
- Rivas Martínez S, Gandullo Gutiérrez JM, Allué Andrade JL, Montero de Burgos JL, González Rebollar JL (1987) *Memoria del Mapa de series de vegetación de España 1:400.000*.
- Rodríguez-Sánchez F, Guzmán B, Valido A, Vargas P, Arroyo J (2009) Late Neogene history of the laurel tree (*Laurus*L., Lauraceae) based on phylogeographical

- analyses of Mediterranean and Macaronesian populations. *Journal of biogeography*, **36**, 1270–1281.
- Rubio de Casas R, Besnard G, Schönswetter P, Balaguer L, Vargas P (2006) Extensive gene flow blurs phylogeographic but not phylogenetic signal in *Olea europaea* L. *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik*, **113**, 575–583.
- Saponari M, Boscia D, Nigro F, Martelli GP, Others (2013) Identification of DNA sequences related to *Xylella fastidiosa* in oleander, almond and olive trees exhibiting leaf scorch symptoms in Apulia (Southern Italy). *Journal of plant pathology: an international journal of the Italian Phytopathological Society*, **95**.
- Saumitou-Laprade P, Vernet P, Vekemans X *et al.* (2017) Elucidation of the genetic architecture of self-incompatibility in olive: Evolutionary consequences and perspectives for orchard management. *Evolutionary applications*, **10**, 867–880.
- Savolainen O, Lascoux M, Merilä J (2013) Ecological genomics of local adaptation. *Nature reviews. Genetics*, **14**, 807–820.
- Scott M, Lindsey R (2020) What's the hottest Earth's ever been? *Climate.gov*.
- Serre J-L (2006) *Génétique des populations*. Dunod.
- Slimp M, Williams LD, Hale H, Johnson MG (2021) On the potential of Angiosperms353 for population genomic studies. *Applications in plant sciences*, **9**.
- Sork VL (2017) Genomic Studies of Local Adaptation in Natural Plant Populations. *The Journal of heredity*, **109**, 3–15.
- Sork VL, Aitken SN, Dyer RJ *et al.* (2013) Putting the landscape into the genomics of trees: Approaches for understanding local adaptation and population responses to changing climate. *Tree genetics & genomes*, **9**, 901–911.
- Spennemann DHR, Richard Allen L (2000) The avian dispersal of olives *Olea europaea*: Implications for Australia. *The Emu*, **100**, 264–273.
- Steane DA, Potts BM, McLean E *et al.* (2014) Genome-wide scans detect adaptation to aridity in a widespread forest tree species. *Molecular ecology*, **23**, 2500–2513.
- Storey, Jd B, Aj D, A, Robinson, D (2022) *qvalue: Q-value estimation for false discovery rate control*.
- Suc J-P (1984) Origin and evolution of the Mediterranean vegetation and climate in Europe. *Nature*, **307**, 429–432.
- Sun X, Han G, Meng Z, Lin L, Sui N (2019) Roles of malic enzymes in plant development and stress responses. *Plant signaling & behavior*, **14**, e1644596.
- Svenning J-C (2003) Deterministic Plio-Pleistocene extinctions in the European cool-temperate tree flora. *Ecology letters*, **6**, 646–653.
- Terral, Alonso N, Buxó I Capdevila R *et al.* (2004) Historical biogeography of olive domestication (*Olea europaea* L.) as revealed by geometrical morphometry applied to biological and archaeological material. *Journal of biogeography*, **31**, 63–77.
- Terral J-F, Bonhomme V, Pagnoux C *et al.* (2021) The Shape Diversity of Olive Stones Resulting from Domestication and Diversification Unveils Traits of the Oldest

REFERENCES

- Known 6500-Years-Old Table Olives from Hishuley Carmel Site (Israel). *Agronomy*, **11**, 2187.
- Terral J-F, Durand A, Newton C, Ivorra S (2009) Archéo-biologie de la domestication de l'olivier en Méditerranée occidentale : de la remise en cause d'une histoire dogmatique à la révélation de son irrigation médiévale. *Etudes*, 1–19.
- Tiffin P, Ross-Ibarra J (2014) Advances and limits of using population genetics to understand local adaptation. *Trends in ecology & evolution*, **29**, 673–680.
- Trkulja V, Tomić A, Iličić R, Nožinić M, Milovanović TP (2022) Xylella fastidiosa in Europe: From the Introduction to the Current Status. *The plant pathology journal / the Korean Society of Plant Pathology*, **38**, 551–571.
- Vaahtera L, Schulz J, Hamann T (2019) Cell wall integrity maintenance during plant development and interaction with the environment. *Nature plants*, **5**, 924–932.
- Van Andel T, Veltman MA, Bertin A *et al.* (2019) Hidden Rice Diversity in the Guianas. *Frontiers in plant science*, **10**, 1161.
- Vanwallendael A, Soltani A, Emery NC *et al.* (2019) A Molecular View of Plant Local Adaptation: Incorporating Stress-Response Networks. *Annual review of plant biology*, **70**, 559–583.
- Venzhik YV, Shchyogolev SY, Dykman LA (2019) Ultrastructural Reorganization of Chloroplasts during Plant Adaptation to Abiotic Stress Factors. *Russian journal of plant physiology: a comprehensive Russian journal on modern phytophysiology*, **66**, 850–863.
- de Villemereuil P, Frichot É, Bazin É, François O, Gaggiotti OE (2014) Genome scan methods against more complex models: when and how much should we trust them? *Molecular ecology*, **23**, 2006–2019.
- Whitlock MC, Lotterhos KE (2015) Reliable detection of loci responsible for local adaptation: Inference of a null model through trimming the distribution of FST. *The American naturalist*, **186**, S24–S36.
- Wright SI, Bi IV, Schroeder SG *et al.* (2005) The effects of artificial selection on the maize genome. *Science*, **308**, 1310–1314.
- Wu T, Hu E, Xu S *et al.* (2021) clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Cambridge (Mass.))*, **2**, 100141.
- Yerushalmi S, Yakir E, Green RM (2011) Circadian clocks and adaptation in Arabidopsis. *Molecular ecology*, **20**, 1155–1165.
- Yu G (2023) *enrichplot: Visualization of Functional Enrichment Result*.
- Zhang C, Dong S-S, Xu J-Y, He W-M, Yang T-L (2018) PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics*, **35**, 1786–1788.
- Zhu S, Niu E, Shi A, Mou B (2019) Genetic Diversity Analysis of Olive Germplasm (*Olea europaea* L.) With Genotyping-by-Sequencing Technology. *Frontiers in genetics*, **10**, 755.
- Zohary D, Spiegel-roy P (1975) Beginnings of Fruit Growing in the Old World. *American Association for the Advancement of Science*, **187**, 319–327.

REFERENCES

- Zunino L, Cubry P, Sarah G *et al.* (2023a) Genomic evidence of genuine wild versus admixed olive populations evolving in the same natural environments in western Mediterranean Basin. *bioRxiv*, 2023.04.18.537296.
- Zunino L, Sarah G, Cubry P, Khadari B (2023b) Selection footprints in *Olea europaea* subsp. *europaea*: Detection in Wild, Admixed and Cultivated olive trees from the Western Mediterranean Basin - Clues to their evolutionary history ?

Supplementary Chapter 1

Supplemental Information for:

**Genomic evidence of genuine wild versus admixed olive populations evolving
in the same natural environments in western Mediterranean Basin**

Lison Zunino^{1,2}, Philippe Cubry², Gautier Sarah¹, Pierre Mournet^{1,3}, Ahmed El Bakkali⁴, Laila Aqbouch¹, Evelyne Costes¹ and Bouchaib Khadari^{1,5}

Affiliations:

1 UMR AGAP Institut, Univ Montpellier, CIRAD, INRAE, Institut Agro, Montpellier F-34398, France

2 DIADE, Univ Montpellier, CIRAD, IRD, Montpellier F-34830, France

3 CIRAD, UMR AGAP Institut, F-34398 Montpellier, France

4 INRA, UR Amélioration des Plantes et Conservation des Ressources Phytogénétiques, Meknes 50102, Morocco

5 Conservatoire Botanique National Méditerranéen (CBNMed), UMR AGAP Institut, F-34398 Montpellier, France

2 | Methods

DNA extraction:

DNA extraction was performed on leaves dried after a minimum of 10 days in silica gel. 25 mg of dried samples were sorted in a Deepwell 96 well plate (Corning 1.1mL Deepwell ThermoFisher Scientific) with beads. The plate was grinded using GenoGrinder (SPEX SamplePrep). The DNA was extracted using a MATAB buffer-based extraction method. Genomic DNA was quantified with Fluoroskan (Thermo fisher Scientific Inc). The quality was checked using agarose gel electrophoresis and Nanoquant plate (Tecan)

Library preparation:

The input DNA was fragmented for 10 minutes with fragmentase enzymes. The fragmentation size was checked with Tapestation with a mean average size of 160 bp. The adaptors were ligated then the samples were purified with AMPure beads (Beckman Coulter, Inc) at 0.8X to do a low and high size selection of 200 bp. Unique Dual Index (NEBNext Multiplex oligos for Illumina) were added to the adaptor-ligated DNA before doing an enrichment PCR as required for NovaSeq experiments. Libraries were amplified for 11 cycles and cleaned with AMPure beads at 0.9X. The library quality was checked with Tapestation. Each library was quantified by qPCR on a LightCycler (Roche Molecular Systems, Inc) before being mixed together on a 48 libraries equimolarity pool. Hybridization capture for targeted NGS protocol is applied on each pool. The baits are added in a pool before hybridizing with beads for at least 16 hours at 60°C. The beads are then washed and the Bait-target hybrids are amplified by PCR for 12 cycles. The reaction is purified with AMPure beads. The captured library quality was verified by Tapestation.

Each capture-pool was quantified by qPCR on LightCycler before being grouped in one final pool of 384 libraries based on equimolarity.

Cluster assignment to create subgroup inside cultivated:

Using sNMF from LEA packages V3.11.3 (Frichot & François 2015), we did structure analyses of the subgroup of cultivated (cultivated from west, including varieties from Spain, Morocco and France). Five runs were performed with K from 1 to 10. According to the cross-entropy, only K of 4 was represented in the results. Using this, we did an assignment of each variety to a cluster, if the variety was assigned to a cluster more than 70%. If the individual is not assigned to any cluster with these parameters, then it is assigned to an admixed cluster, C0. The C2 group was composed only with 3 individuals, which are not sufficient to a TreeMix analysis, we decided to put them in the C0 group and remove C2.

References:

Frichot, Eric, and Olivier François. 2015. "LEA: An R Package for Landscape and Ecological Association Studies." *Methods in Ecology and Evolution* / British Ecological Society 6 (8): 925–29.

Supplementary figures:

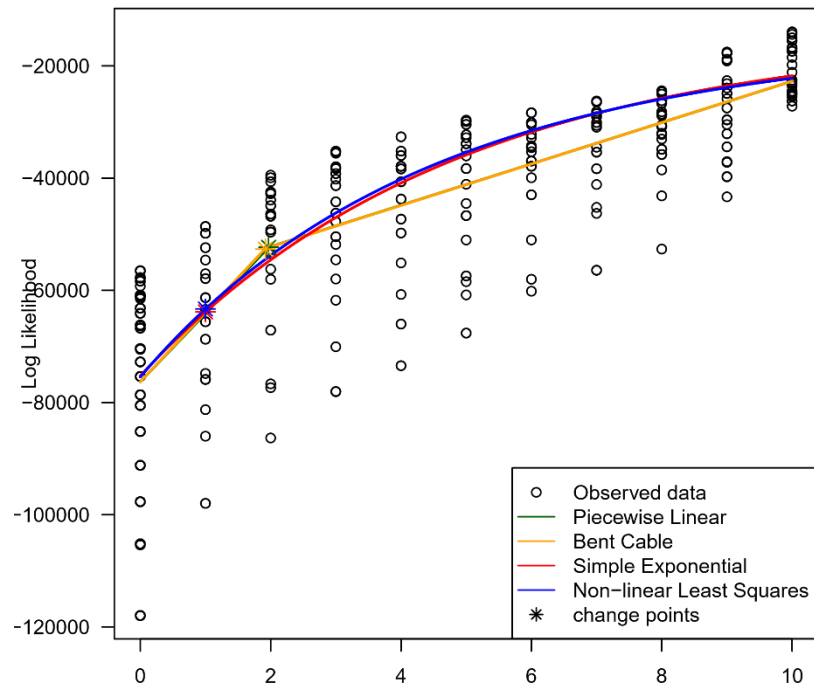


Figure S1. Multiple linear models of possible migration events between 27 naturally occurring populations, eastern wild Mediterranean population and cultivated accessions of olives using TreeMix. One hundred TreeMix runs were done with a random SNPs block size between 100 and 1000, from 1 to 10 migrations each with M29 as outgroup. The analysis was performed with *OptM* package.

m (migration edges)

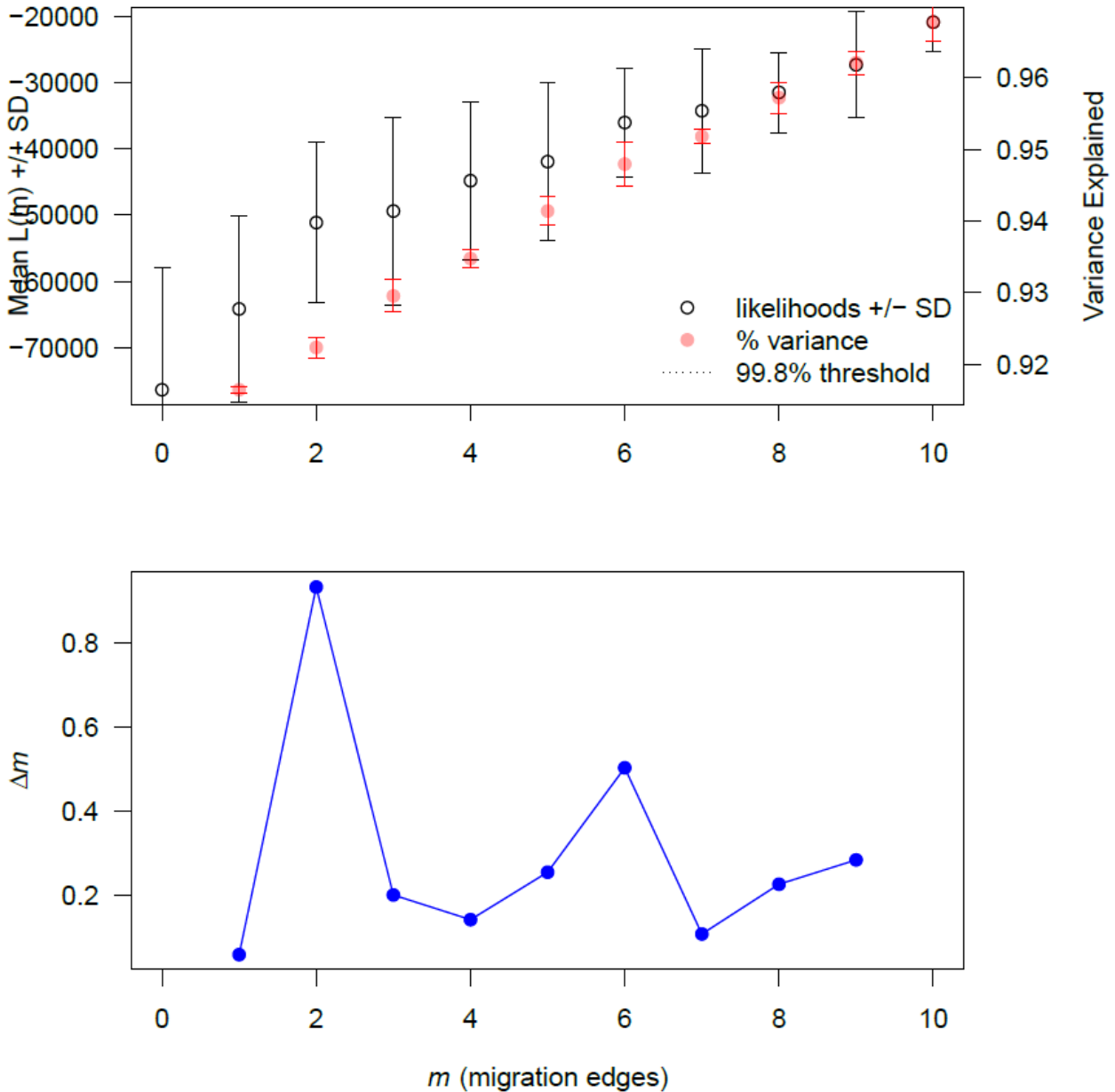


Figure S2. Evanno test of possible migration events between 27 naturally occurring populations, eastern wild Mediterranean population and cultivated accessions of olives using TreeMix. One hundred TreeMix runs were done with a random SNPs block size between 100 and 1000, from 1 to 10 migrations each with M29 as outgroup. The analysis was performed with *OptM* package.

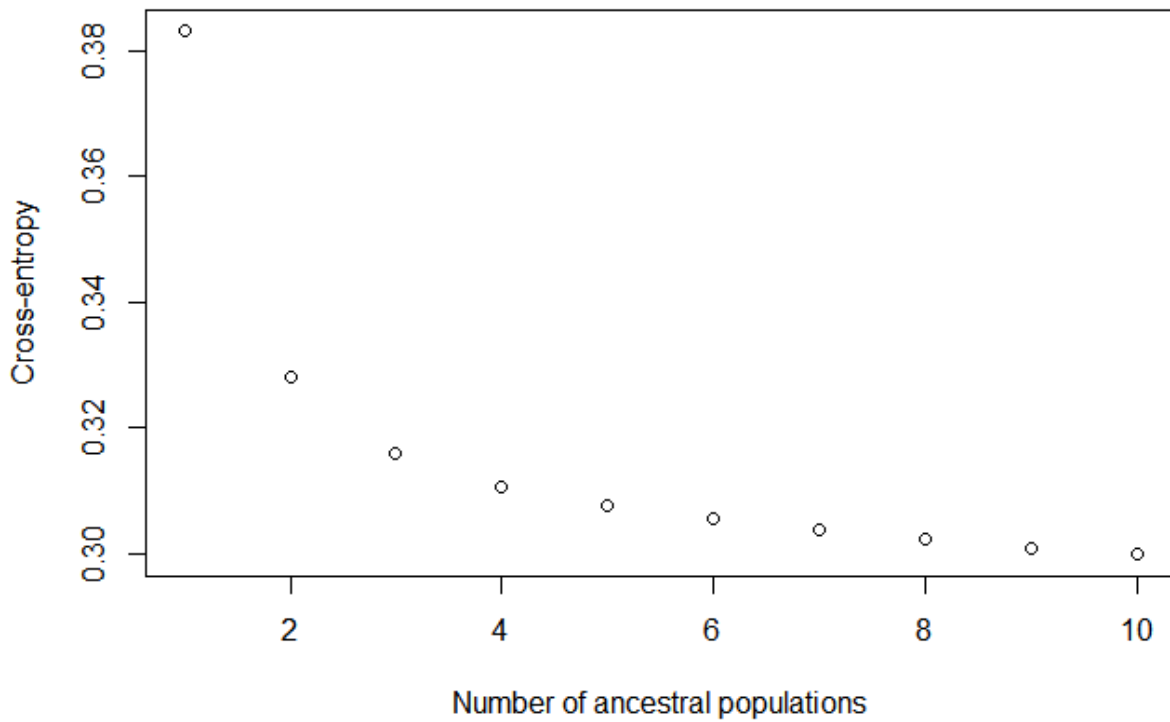


Figure S3. Cross-entropy criterion inferred by sNMF analyze performed on the genome-wide SNPs diversity of natural populations of *O. europaea L.* collected in France (143), Spain (123), Morocco (96) and Turkey (13) and cultivated *O. europaea L.* from the western Mediterranean Basin (145), using 142,060 SNPs.

Table S1. List of the baits used in the study. The baits were designed based on annotated gene of *Olea europaea* var. *europaea* (cv. Farga) Oe9 genome assembly (Julca et al. 2020). This table is in the supplementary xls file.

Table S2. Summary information of the sequenced cultivated accessions of *O. europaea* L. used in the study

Sample ID	Code	Accession origine	Cultivar origine	Source
Unkown-VS2-545	MAR00545	Morocco	Morocco	World Olive Germplasm Bank of Marrakech
Acebuchera	MAR00215	Spain	Spain	World Olive Germplasm Bank of Marrakech
Aglandau	MAR00187	France	France	World Olive Germplasm Bank of Marrakech
Alameno Blanco	MAR00216	Spain	Spain	World Olive Germplasm Bank of Marrakech
Alameno de Montilla	MAR00218	Spain	Spain	World Olive Germplasm Bank of Marrakech
Amargoso	MAR00219	Spain	Spain	World Olive Germplasm Bank of Marrakech
Arbequina	MAR00220	Spain	Spain	World Olive Germplasm Bank of Marrakech
Azul	MAR00221	Spain	Spain	World Olive Germplasm Bank of Marrakech
Berri Meslal-397	MAR00397	Morocco	Morocco	World Olive Germplasm Bank of Marrakech
Berri Meslal-532	MAR00532	Morocco	Morocco	World Olive Germplasm Bank of Marrakech
Bical	MAR00333	Spain	Spain	World Olive Germplasm Bank of Marrakech
Blanqueta	MAR00222	Spain	Spain	World Olive Germplasm Bank of Marrakech
Bolvino	MAR00223	Spain	Spain	World Olive Germplasm Bank of Marrakech

SUPPLEMENTARY CHAPTER 1

Borriolenca	MAR00334	Spain	Spain	World Olive Germplasm Bank of Marrakech
Bouchouika	MAR00394	Morocco	Morocco	World Olive Germplasm Bank of Marrakech
Bouteillan	MAR00189	France	France	World Olive Germplasm Bank of Marrakech
Callosina	MAR00391	Morocco	Spain	World Olive Germplasm Bank of Marrakech
Canivano Negro	MAR00224	Spain	Spain	World Olive Germplasm Bank of Marrakech
Carrasqueno de Jumilla	MAR00226	Spain	Spain	World Olive Germplasm Bank of Marrakech
Carrasquillo	MAR00335	Spain	Spain	World Olive Germplasm Bank of Marrakech
Cayon	MAR00191	France	France	World Olive Germplasm Bank of Marrakech
Cerezuela	MAR00349	Spain	Spain	World Olive Germplasm Bank of Marrakech
Changlot Real	MAR00227	Spain	Spain	World Olive Germplasm Bank of Marrakech
Chorroo	MAR00229	Spain	Spain	World Olive Germplasm Bank of Marrakech
Cirujal	MAR0043	Italy	Spain	World Olive Germplasm Bank of Marrakech
Corbella	MAR00230	Spain	Spain	World Olive Germplasm Bank of Marrakech
Cornezuelo de Jaen	MAR00231	Spain	Spain	World Olive Germplasm Bank of Marrakech
Cornicabra	MAR00232	Spain	Spain	World Olive Germplasm Bank of Marrakech
Cucca	MAR0032	Spain	Spain	World Olive Germplasm Bank of Marrakech
Dolce di Rossano	MAR0011	Spain	Spain	World Olive Germplasm Bank of Marrakech
Dressi	MAR00286	Spain	Spain	World Olive Germplasm Bank of Marrakech

SUPPLEMENTARY CHAPTER 1

Dulzal	MAR00233	Spain	Spain	World Olive Germplasm Bank of Marrakech
El Lewa	MAR00494	Spain	Spain	World Olive Germplasm Bank of Marrakech
Empeltre	MAR00234	Spain	Spain	World Olive Germplasm Bank of Marrakech
Enagua de Arenas	MAR00336	Spain	Spain	World Olive Germplasm Bank of Marrakech
Escarabajuelo de Posadas	MAR00235	Spain	Spain	World Olive Germplasm Bank of Marrakech
Escarabajuelo de Úbeda	MAR00236	Spain	Spain	World Olive Germplasm Bank of Marrakech
Farga	MAR00338	Spain	Spain	World Olive Germplasm Bank of Marrakech
Frantoio	MAR0039	Spain	Spain	World Olive Germplasm Bank of Marrakech
Fulla de Salze	MAR00339	Spain	Spain	World Olive Germplasm Bank of Marrakech
Gentile di chieti	MAR0015	Spain	Spain	World Olive Germplasm Bank of Marrakech
Gordal Sevillana	MAR00108	Italy	Spain	World Olive Germplasm Bank of Marrakech
Gordal de Granada	MAR00238	Spain	Spain	World Olive Germplasm Bank of Marrakech
Grappolo	MAR0041	Spain	Spain	World Olive Germplasm Bank of Marrakech
Grossane-194	MAR00194	France	France	World Olive Germplasm Bank of Marrakech
Habichuelero de Grazalema	MAR00239	Spain	Spain	World Olive Germplasm Bank of Marrakech
Hojiblanca	MAR00240	Spain	Spain	World Olive Germplasm Bank of Marrakech
Idleb	MAR00591	Spain	Spain	World Olive Germplasm Bank of Marrakech
Jabaluna	MAR00241	Spain	Spain	World Olive Germplasm Bank of Marrakech

SUPPLEMENTARY CHAPTER 1

Jaropo	MAR00242	Spain	Spain	World Olive Germplasm Bank of Marrakech
Khashabi-631	MAR00631	Spain	Spain	World Olive Germplasm Bank of Marrakech
Lastrino	MAR0023	Spain	Spain	World Olive Germplasm Bank of Marrakech
Lazzero di prata	MAR00367	Spain	Spain	World Olive Germplasm Bank of Marrakech
Lechin de Sevilla	MAR00243	Spain	Spain	World Olive Germplasm Bank of Marrakech
Lechin de Granada	MAR00340	Spain	Spain	World Olive Germplasm Bank of Marrakech
Lentisca-244	MAR00244	Spain	Spain	World Olive Germplasm Bank of Marrakech
Limoncillo	MAR00341	Spain	Spain	World Olive Germplasm Bank of Marrakech
Lloron de Atarfe	MAR00245	Spain	Spain	World Olive Germplasm Bank of Marrakech
Llumeta	MAR00343	Spain	Spain	World Olive Germplasm Bank of Marrakech
Loaime	MAR00344	Spain	Spain	World Olive Germplasm Bank of Marrakech
Lucques	MAR00195	France	France	World Olive Germplasm Bank of Marrakech
Machorron	MAR00247	Spain	Spain	World Olive Germplasm Bank of Marrakech
Manzanilla Cacerena	MAR00248	Spain	Spain	World Olive Germplasm Bank of Marrakech
Manzanilla de Sevilla	MAR00251	Spain	Spain	World Olive Germplasm Bank of Marrakech
Manzanilla de Agua	MAR00345	Spain	Spain	World Olive Germplasm Bank of Marrakech
Manzanilla de Hellin	MAR00346	Spain	Spain	World Olive Germplasm Bank of Marrakech
Manzanilla de Montefrio	MAR00250	Spain	Spain	World Olive Germplasm Bank of Marrakech

SUPPLEMENTARY CHAPTER 1

Mesyaf-641	MAR00641	Spain	Spain	World Olive Germplasm Bank of Marrakech
Mignolo Cerretano	MAR0046	Spain	Spain	World Olive Germplasm Bank of Marrakech
Minekiri	MAR00634	Spain	Spain	World Olive Germplasm Bank of Marrakech
Mollar de Cieza	MAR00348	Spain	Spain	World Olive Germplasm Bank of Marrakech
Morchione	MAR0067	France	France	World Olive Germplasm Bank of Marrakech
Morisca	MAR00254	Spain	Morocco	World Olive Germplasm Bank of Marrakech
Morona	MAR00246	Spain	Spain	World Olive Germplasm Bank of Marrakech
Morrut	MAR00350	Spain	Spain	World Olive Germplasm Bank of Marrakech
Negral de Sabinan-255	MAR00255	Spain	Spain	World Olive Germplasm Bank of Marrakech
Negrillo Redondo	MAR00257	Spain	Spain	World Olive Germplasm Bank of Marrakech
Negrillo de Arjona	MAR00256	Spain	Spain	World Olive Germplasm Bank of Marrakech
Negrillo de Estepa	MAR00351	Spain	Spain	World Olive Germplasm Bank of Marrakech
Negrillo de Iznalloz	MAR00352	Spain	Spain	World Olive Germplasm Bank of Marrakech
Nerba	MAR00123	Spain	Spain	World Olive Germplasm Bank of Marrakech
Nevado Azul	MAR00354	Spain	Spain	World Olive Germplasm Bank of Marrakech
Nevado Basto	MAR00225	Spain	Spain	World Olive Germplasm Bank of Marrakech
Nevado Rizado	MAR00355	Spain	Spain	World Olive Germplasm Bank of Marrakech
Ocal	MAR00258	Spain	Spain	World Olive Germplasm Bank of Marrakech

SUPPLEMENTARY CHAPTER 1

Ojo de Liebre	MAR00259	Spain	Spain	World Olive Germplasm Bank of Marrakech
Olivo de Mancha Real	MAR00260	Spain	Spain	World Olive Germplasm Bank of Marrakech
Olivo di Mandanici	MAR00136	Syria	Spain	World Olive Germplasm Bank of Marrakech
Palomar	MAR00262	Spain	Spain	World Olive Germplasm Bank of Marrakech
Patronet	MAR00263	Spain	Spain	World Olive Germplasm Bank of Marrakech
Picholine	MAR00196	France	France	World Olive Germplasm Bank of Marrakech
Picholine Marocaine	MAR00540	Morocco	Morocco	World Olive Germplasm Bank of Marrakech
Pico Limon de Grazalema	MAR00265	Spain	Spain	World Olive Germplasm Bank of Marrakech
Picual	MAR00267	Spain	Spain	World Olive Germplasm Bank of Marrakech
Picudo	MAR00356	Spain	Spain	World Olive Germplasm Bank of Marrakech
Plementa Bjelica	MAR00402	Spain	Spain	World Olive Germplasm Bank of Marrakech
Puntoza	MAR00499	Spain	Spain	World Olive Germplasm Bank of Marrakech
Racimal	MAR00268	Spain	Spain	World Olive Germplasm Bank of Marrakech
Rapasayo	MAR00357	Spain	Spain	World Olive Germplasm Bank of Marrakech
Razzaio	MAR0063	France	France	World Olive Germplasm Bank of Marrakech
Rechino	MAR00269	Spain	Spain	World Olive Germplasm Bank of Marrakech
Ronde de la Menara	MAR00543	Morocco	Morocco	World Olive Germplasm Bank of Marrakech
Rossellino	MAR0060	Spain	Spain	World Olive Germplasm Bank of Marrakech

SUPPLEMENTARY CHAPTER 1

Royal de Cazorla	MAR00270	Spain	Spain	World Olive Germplasm Bank of Marrakech
Sabatera	MAR00271	Spain	Spain	World Olive Germplasm Bank of Marrakech
Salonenque	MAR00197	France	France	World Olive Germplasm Bank of Marrakech
Santa Martinenga	MAR00145	Syria	Spain	World Olive Germplasm Bank of Marrakech
Sayali	MAR00287	Spain	Spain	World Olive Germplasm Bank of Marrakech
Sevillano de Jumilla	MAR00272	Spain	Spain	World Olive Germplasm Bank of Marrakech
Sevillenca	MAR00358	Spain	Spain	World Olive Germplasm Bank of Marrakech
Sinopolese	MAR0018	Spain	Spain	World Olive Germplasm Bank of Marrakech
Storta	MAR00406	Spain	Spain	World Olive Germplasm Bank of Marrakech
Tabelout	MAR00437	Spain	Spain	World Olive Germplasm Bank of Marrakech
Tebabs	MAR00661	Spain	Spain	World Olive Germplasm Bank of Marrakech
Teffah	MAR00427	Spain	Spain	World Olive Germplasm Bank of Marrakech
Tempranillo de Yeste-274	MAR00274	Spain	Spain	World Olive Germplasm Bank of Marrakech
Tonda Iblea	MAR0012	Spain	Spain	World Olive Germplasm Bank of Marrakech
Unkown-OT2-537	MAR00537	Morocco	Morocco	World Olive Germplasm Bank of Marrakech
Unkown-VS1-544	MAR00544	Spain	Spain	World Olive Germplasm Bank of Marrakech
Unkown-VS2-545	MAR00546	Morocco	Morocco	World Olive Germplasm Bank of Marrakech
Unkown-VS2-545	MAR00545	Morocco	Morocco	World Olive Germplasm Bank of Marrakech

SUPPLEMENTARY CHAPTER 1

Unkown-VS5-547	MAR00547	Morocco	Morocco	World Olive Germplasm Bank of Marrakech
Uovo di Piccione	MAR00141	Syria	Spain	World Olive Germplasm Bank of Marrakech
Varudo	MAR00273	Spain	Spain	World Olive Germplasm Bank of Marrakech
Varudo-275	MAR00275	Spain	Spain	World Olive Germplasm Bank of Marrakech
Vera	MAR00276	Spain	Spain	World Olive Germplasm Bank of Marrakech
Verdala	MAR00278	Spain	Spain	World Olive Germplasm Bank of Marrakech
Verdale	MAR00199	France	France	World Olive Germplasm Bank of Marrakech
Verdial de Badajoz	MAR00342	Spain	Spain	World Olive Germplasm Bank of Marrakech
Verdial de Huevar	MAR00213	Portugal	Spain	World Olive Germplasm Bank of Marrakech
Verdiell	MAR00279	Spain	Spain	World Olive Germplasm Bank of Marrakech
Villalonga	MAR00201	Portugal	Spain	World Olive Germplasm Bank of Marrakech
Zalmati-299	MAR00299	Spain	Spain	World Olive Germplasm Bank of Marrakech
Zarza	MAR00280	Spain	Spain	World Olive Germplasm Bank of Marrakech
Zeletni	MAR00421	Spain	Spain	World Olive Germplasm Bank of Marrakech
Zinzala	CTO_E8_N4	France	France	Technical Centrer of Olive
Petit Ribier	CTO_P10_F06	France	France	Technical Centrer of Olive
Sabine	CTO_P10_N27	France	France	Technical Centrer of Olive
Cayet Roux	POR_P39_33_10	France	France	Porquerolles collection
Salonenque	POR_P39_33_16	France	France	Porquerolles collection
Brun	POR_P39_37_21	France	France	Porquerolles collection
Tanche	POR_P43_02_05	France	France	Porquerolles collection
Cayon	POR_P43_03_12	France	France	Porquerolles collection

SUPPLEMENTARY CHAPTER 1

Reymet	POR_P43_06_01	France	France	Porquerolles collection
Grossane	POR_P43_17_13	France	France	Porquerolles collection

SUPPLEMENTARY CHAPTER 1

Table S3. List of filters and remaining SNPs for each on genomic data of target sequencing of 561 genotypes of *O. europaea* L.

	Total SNPs remaining
Raw data	27275679
Remove Indel	24547413
Quality > 200	22340559
SNP Biallelic only	21036126
Min mean depth 8	958648
Max mean depth 400	957569
SNP cluster (3,10)	383904
Min depth 8	383904
Missing data Site > 0.15	155063
heterozygoty > 0.85	154675
Minor allele count 1	142060

Table S4. List of individuals of *O. europaea* L. removed from genomic data set because of missing data >0.2

Sample ID	Type
removed	
Manzanilla_de_Abla	Cultivated
OES_E13_03	Wild
OES_E15_08	Wild
OES_E15_12	Wild
OES_E16_05	Wild
OES_E16_10	Wild
OES_E16_13	Wild
OES_E17_08	Wild
OES_E18_02	Wild
OES_E18_04	Wild
OES_E18_11	Wild
OES_E19_12	Wild
OES_E20_02	Wild
OES_F02_01	Wild
OES_F02_15	Wild
OES_F07_08	Wild
OES_F07_10	Wild
OES_F10_05	Wild
OES_F10_13	Wild
OES_F11_04	Wild
OES_M21_03	Wild
OES_M21_13	Wild
OES_M23_03	Wild
OES_M23_05	Wild
OES_M23_07	Wild
OES_M23_09	Wild
OES_M23_10	Wild
OES_M23_11	Wild
OES_M23_13	Wild
OES_M23_14	Wild

SUPPLEMENTARY CHAPTER 1

OES_M23_15	Wild
OES_M28_10	Wild
OES_M29_04	Wild
OST072	Wild
OST082	Wild

Table S5. Results of bootstrapping over loci of pairwise F_{ST} performed on the genome-wide SNPs diversity of natural populations of *O. europaea* L. collected in France (143), Spain (123), Morocco (96) and Turkey (13) using 142,060 SNPs. The upper limit ci is the upper part and the lower limit ci in the bottom part of the matrix. This table is in the supplementary xls file.

Table S6. List of cultivated accessions of *O. europaea* L. and their assignment to genetic cluster (C0, C1, C3 or C4) according to sNMF assignation at $K = 4$. The detailed method is developed above (Supplementary information)

Accession names	Cluster assignation
Unkown-VS2-545	C0
Acebuchera	C4
Aglandau	C1
Alameno Blanco	C4
Alameno de Montilla	C0
Amargoso	C1
Arbequina	C3
Azul	C0
Berri Meslal-397	C0
Berri Meslal-532	C0
Bical	C4
Blanqueta	C3
Bolvino	C0
Borriolenca	C0
Bouchouika	C4
Bouteillan	C1
Zinzala	C3
Petit Ribier	C3

Sabine	C0
Callosina	C4
Canivano Negro	C4
Carrasqueno de Jumilla	C4
Carrasquillo	C4
Cayon	C1
Cerezuela	C4
Changlot Real	C0
Chorruo	C4
Cirujal	C0
Corbella	C0
Cornezuelo de Jaen	C0
Cornicabra	C4
Cucca	C0
Dolce di Rossano	C0
Dressi	C0
Dulzal	C4
El Lewa	C1
Empeltre	C0
Enagua de Arenas	C4
Escarabajuelo	de
Posadas	C4
Escarabajuelo	de
Úbeda	C0
Farga	C0
Frantoio	C3
Fulla de Salze	C4
Gentile di chieti	C3
Gordal Sevillana	C1
Gordal de Granada	C4
Grappolo	C0
Grossane-194	C1
Habichuelero	de
Grazalema	C4
Hojiblanca	C4

SUPPLEMENTARY CHAPTER 1

Idleb	C1
Jabaluna	C0
Jaropo	C4
Khashabi-631	C1
Lastrino	C0
Lazzero di prata	C3
Lechin de Sevilla	C0
Lechin de Granada	C0
Lentisca-244	C0
Limoncillo	C4
Lloron de Atarfe	C4
Llumeta	C0
Loaime	C4
Lucques	C1
Machorron	C4
Manzanilla Cacerena	C4
Manzanilla de Sevilla	C4
Manzanilla de Agua	C4
Manzanilla de Hellin	C4
Manzanilla de Montefrio	C4
Mesyaf-641	C1
Mignolo Cerretano	C3
Minekiri	C1
Mollar de Cieza	C4
Morchione	C0
Morisca	C4
Morona	C4
Morrut	C0
Negral de Sabinan-255	C1
Negrillo Redondo	C4
Negrillo de Arjona	C4
Negrillo de Estepa	C4
Negrillo de Iznalloz	C4
Nerba	C1
Nevado Azul	C4

Nevado Basto	C4
Nevado Rizado	C4
Ocal	C4
Ojo de Liebre	C4
Olivo de Mancha Real	C4
Olivo di Mandanici	C3
Cayet Roux	C0
Salonenque	C1
Brun	C1
Tanche	C1
Cayon	C0
Reymet	C3
Grossane	C1
Palomar	C0
Patronet	C0
Picholine	C0
Picholine Marocaine	C4
Pico Limon de	
Grazalema	C4
Picual	C4
Picudo	C4
Plementa Bjelica	C0
Puntoza	C0
Racimal	C4
Rapasayo	C0
Razzaio	C3
Rechino	C4
Ronde de la Menara	C4
Rossellino	C3
Royal de Cazorla	C4
Sabatera	C0
Salonenque	C1
Santa Martinenga	C0
Sayali	C0
Sevillano de Jumilla	C4

SUPPLEMENTARY CHAPTER 1

Sevillenca	C0
Sinopolese	C0
Storta	C0
Tabelout	C3
Tebabs	C1
Teffah	C1
Tempranillo de Yeste- 274	C4
Tonda Iblea	C1
Unkown-OT2-537	C3
Unkown-VS1-544	C0
Unkown-VS2-545	C0
Unkown-VS2-545	C0
Unkown-VS5-547	C0
Uovo di Piccione	C1
Varudo	C4
Varudo-275	C4
Vera	C3
Verdala	C4
Verdale	C1
Verdial de Badajoz	C4
Verdial de Huevar	C4
Verdiell	C0
Villalonga	C0
Zalmati-299	C3
Zarza	C0
Zeletni	C0

Table S7. Descriptive data of depth and enrichment rate of *O. europaea* L. using target sequencing method versus whole genome sequencing method.

Sample ID	Mean depth whole genome	Mean depth target sequencing	Enrichment rate
OES_E13_09	1.21	40.40	33.27
OES_F10_03	1.27	34.72	27.31
Picholine	0.77	18.20	23.59
Picholine_Marocaine	1.27	67.20	52.88
Mean_all	1.13	40.13	34.26

Supplementary Chapter 2

Selection footprints in *Olea europaea* subsp. *europaea*: Detection in wild, admixed and cultivated olive trees from the Western Mediterranean Basin - Clues to their evolutionary history?

In preparation

Lison Zunino et al

2 | Methods

2.1 | Sampling and sequencing

Table S1. List of the baits on chromosome and scaffolds used in this study. The baits were designed based on annotated gene of *Olea europaea* var. *europaea* (cv. Farga) Oe9 genome assembly (Julca *et al.* 2020). Table available here: <https://zenodo.org/records/10058461>

2.2 | SNP filtering

Table S2. Assignment of each individual to a genetic group (wild, admixed, cultivated)

Summary information of the assignment of each sequenced individual of *Olea europaea* to a specific genetic group: wild, admixed and cultivated. Table available here: <https://zenodo.org/records/10058461>

WOGBM: World Olive Germplasm Bank of Marrakech; CTO: Centre Technique de l'Olivier; POR: Porquerolles collection.

3 | Results

3.1 | Target sequencing characterization

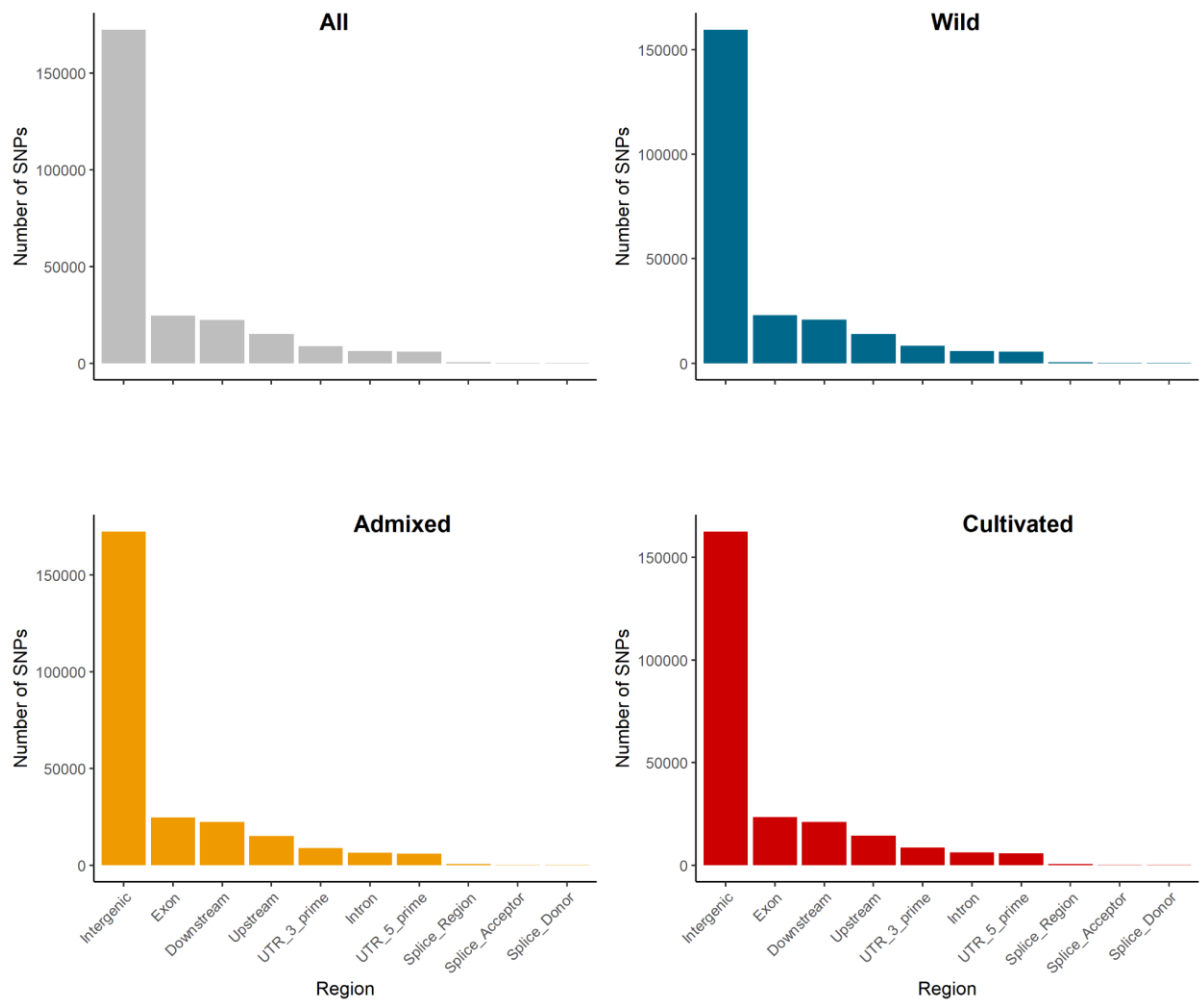


Figure S1. Number of sequenced SNPs per type of region in the genome of *O. europaea* cv. Farga for each genetic group: All individuals in grey, wild individuals in blue, admixed individuals in yellow and cultivated individuals in red.

3.2 | Genomic diversity, LD decay and genomic structure

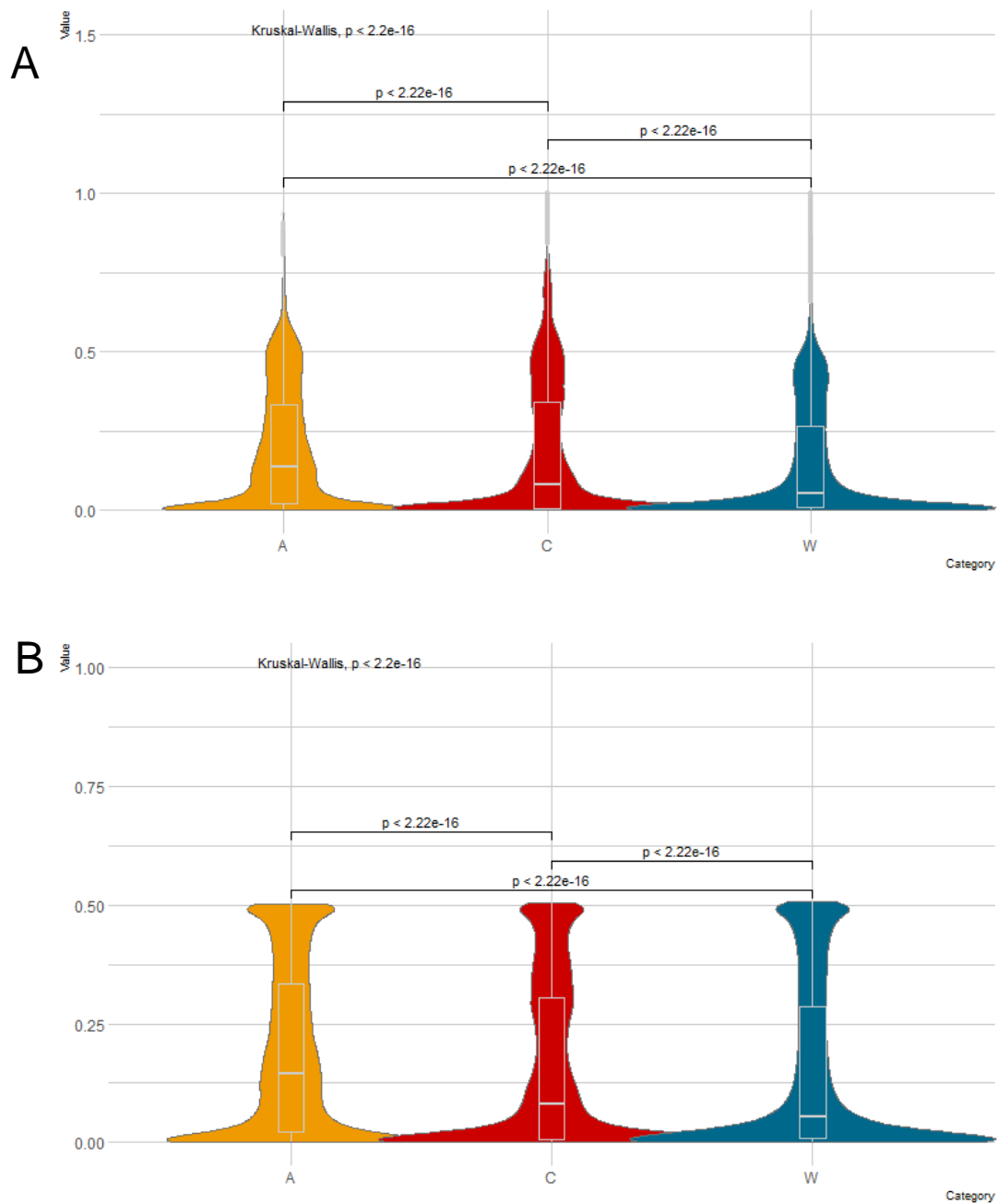


Figure S2. Measure of genetic diversity per loci of *O.europaea* for admixed (A; in yellow), cultivated C; (in red) and wild (W; in blue), with (A) measure of observed heterozygosity H_o and (B) measure of expected heterozygosity H_s . Differences in distribution between pairs of genetic groups were tested using Mann-Whitney test (p -value < 0.05). The difference between distribution of all groups was tested using kruskall-Wallis (p -value < 0.05).

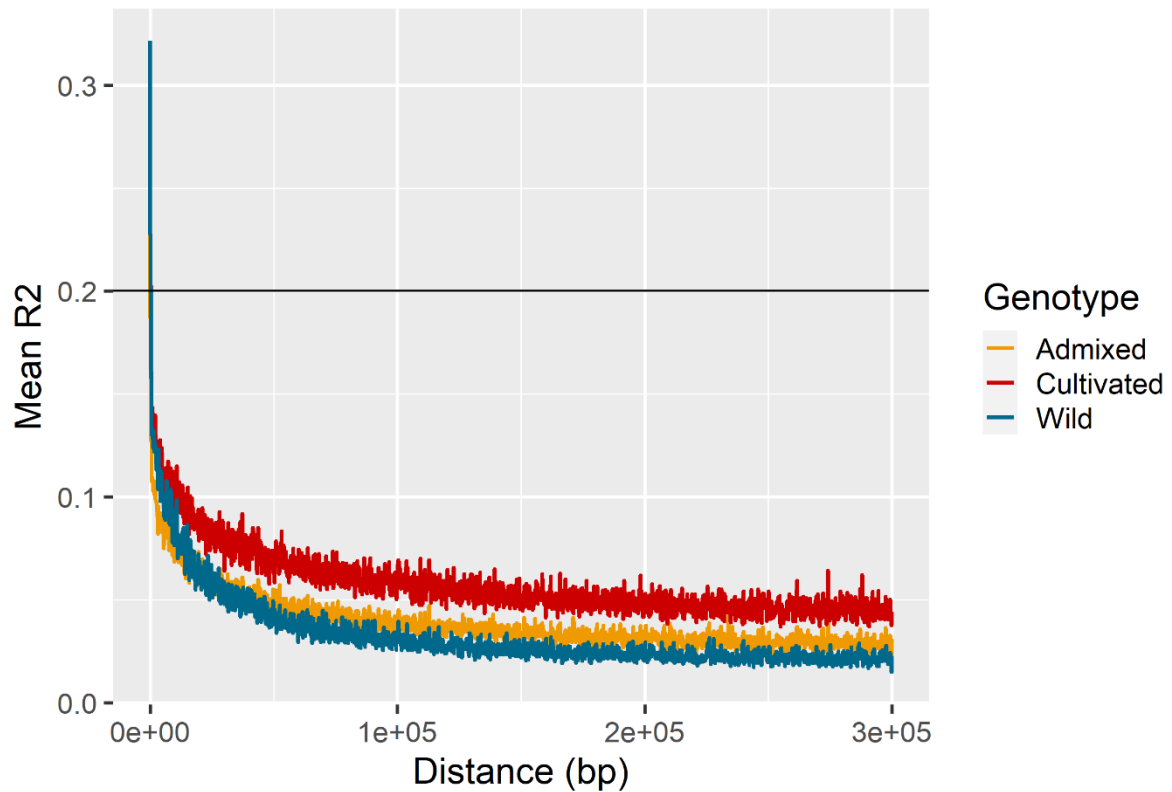


Figure S3. Linkage disequilibrium decay in different groups of *O. europaea* based on SNPs. It corresponds to values of mean r^2 according to distance in base pairs between SNPs for three genotypic groups: wild (in blue), admixed (in yellow) and cultivated (in red).

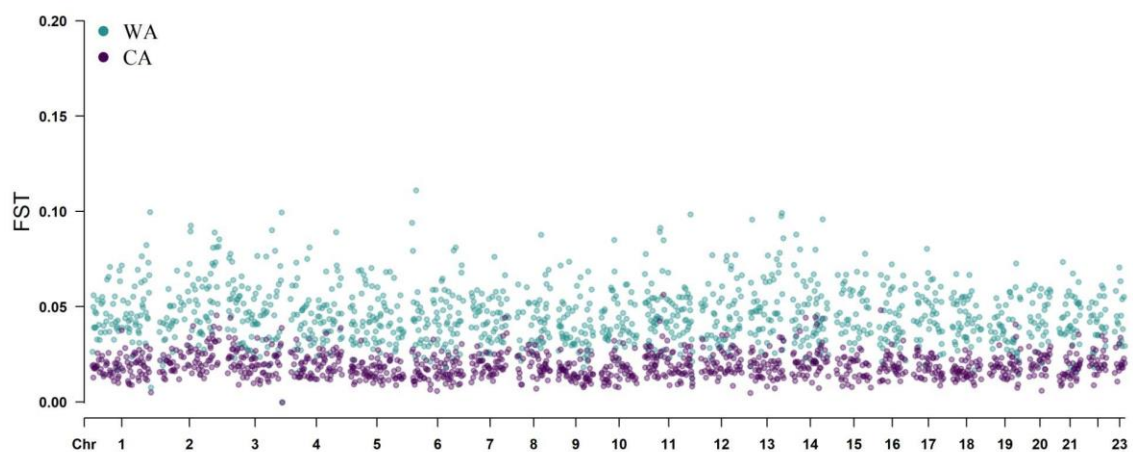


Figure S4. Mean pairwise F_{ST} calculated on 100 SNPs windows between genetic groups of *Olea europaea*. Two mean pairwise calculations are represented here, with in blue wild-admixed mean pairwise F_{ST} (WA) and in purple cultivated-admixed mean pairwise F_{ST} (CA).

3.3 | Biological significance of outlier SNPS

Table S2. Summary of gene information for candidate SNPs for selection for wild, cultivated and admixed olives *O. europaea* L. Table available here: <https://zenodo.org/records/10058461>

Table S3. Number of effects by impact (high, low, moderate and modifier) for candidate SNPs for selection, for wild, admixed and cultivated genetic groups of *Olea europaea* using SnpEff (Cingolani *et al.* 2012).

Predict impact	Number of SNPs		
	Wild	Admixed	Cultivated
High	1	0	0
Low	72	2	53
Moderate	91	1	57
Modifier	1353	63	1054

High (e.g. stop gained), moderate (e.g. missense-variant), modifier (e.g. downstream gene variant) or low (e.g. synonym variant).

References

- Cingolani P, Platts A, Wang LL, et al.** 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* 6: 80–92.
- Julca I, Marcet-Houben M, Cruz F, et al.** 2020. Genomic evidence for recurrent genetic admixture during the Cingolani e domestication of Mediterranean olive trees (*Olea europaea* L.). *BMC biology* 18: 1–25.

Supplementary Chapter 3

Does crop-to-wild gene flow lead to maladaptation of perennial species to their natural environment? Insights from the study of local adaptation of wild olive trees to their environment.

Lison Zunino *et al*

In preparation

1 | Methods

2.1 | Sampling and sequencing

Table S1. Summary information of the sequenced of *O. europaea* L. used in the study.

Table available here: <https://zenodo.org/doi/10.5281/zenodo.10044433>

WOGBM: World Olive Germplast Bank of Marrakech; CTO: Technical Centre of Olive;

POR: Porquerolles collection.

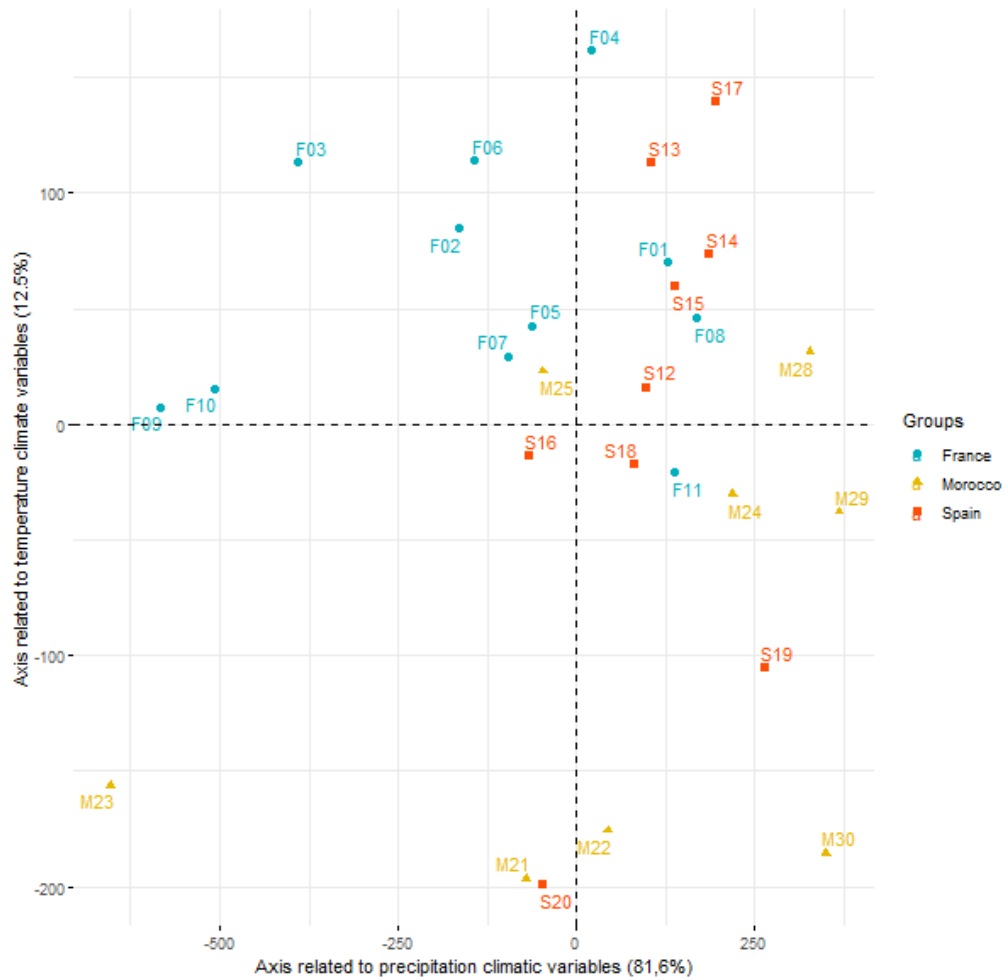


Figure S1. PCA led on climatic data from native populations of olives tree *O. europaea* in France (blue circle), in Spain (red square) and in Morocco (yellow triangle). The first dimension is correlated to the precipitation-related variables and the second dimension to the temperature-related variables from 1981 to 2010 in CHELSA V2 (Karger *et al.* 2017).

2.3 | Selection of environmental variables

Table S2. Summary information of all environmental data of 26 naturally occurring populations of *O. europaea* used in this study.

Site	Bio1	Bio2	Bio3	Bio4	Bio5	Bio6	Bio7	Bio8	Bio9	Bio10	Bio11	Bio12	Bio13	Bio14	Bio15	Bio16	Bio17	Bio18	Bio19	gmted
F01	15.75	7.7	31.5	584.6	29.15	4.65	24.5	12.45	23.55	23.55	8.55	609.2	89.8	14.9	40.2	220.1	70.6	70.6	188.6	2.52
F02	14.25	8.3	33.3	584.2	28.05	03.05	25	10.85	22.15	22.15	7.15	866.5	133.8	22.6	40.1	331.7	106.8	106.8	247.5	142.85
F03	13.25	9.8	36.2	607.2	28.35	1.25	27.1	13.55	21.45	21.45	5.75	1068.5	165.9	28.6	42.3	416.1	131.5	131.5	290.6	317.28
F04	14.95	9.9	34	659.4	30.95	1.75	29.2	15.15	23.85	23.85	6.85	705.1	104.1	27.1	40.7	284	101.9	101.9	153.5	81.42
F06	15.55	9.2	33.6	624	30.05	2.65	27.4	11.95	23.95	23.95	7.85	846.4	121.3	17.9	43.1	336.3	102.3	102.3	232	101.79
F07	15.75	5.4	26.6	537.6	25.95	5.55	20.4	13.15	22.75	22.75	9.15	798.1	122.9	16.2	50	343.2	79.5	79.5	224.1	71.99
F08	16.65	5.3	25.2	545.4	27.65	6.85	20.8	14.25	23.85	23.85	10.15	574.8	83.9	9.6	42.7	225	65.2	65.2	153.6	33.63
F09	16.55	4.4	23	522.3	26.65	7.45	19.2	14.55	23.35	23.45	10.35	1215	197.2	13.3	51.6	518.4	95.7	148	376.8	49.70
F10	16.15	6.2	29	524.5	27.75	6.45	21.3	13.95	22.95	23.05	9.95	1157.6	178.2	15.9	48.4	476.2	90.8	137.7	354.2	63.19
F11	16.75	2.7	16.6	471.1	25.65	9.35	16.3	15.35	22.85	23.25	11.35	600.7	88.1	7.3	50.5	249.7	40.2	69.1	162.3	45.58
S12	15.55	3.7	21.8	478.9	24.75	7.75	17	16.95	21.85	22.05	9.95	644.3	103.4	18.2	38.3	225	80.9	107	173.9	36.83
S13	14.85	7.4	32.9	544.8	26.35	3.75	22.6	16.15	21.95	22.05	8.25	640.1	86.4	25.2	32.2	231.4	123.3	165.7	129.4	241.22
S14	16.55	6.3	29.8	534.7	27.65	6.45	21.2	17.95	23.55	23.65	10.05	561.1	87.7	14.8	43.2	220.8	78.7	128.1	116.7	40.93
S15	17.75	8.1	34.6	550.7	29.45	6.15	23.3	18.75	24.95	24.95	11.05	604	91.3	12.5	41	233.2	72.8	72.8	147.7	9.99
S16	17.85	2.9	18.4	463	26.55	10.65	15.9	19.75	23.75	24.25	12.65	776.6	117.5	6.7	50.7	321.1	53.6	138.8	218.2	48.26
S17	16.75	10.8	34.3	718.6	33.85	2.35	31.5	12.25	26.45	26.45	08.05	534.3	84.7	4.9	51.2	210.6	38.4	38.4	183.3	606.55
S18	18.65	11.1	39.3	570.6	34.15	5.95	28.2	15.35	26.15	26.15	11.55	612	115.7	2	68.8	292.6	19.1	19.1	248.6	37.31
S19	17.75	3.8	23.6	427.5	26.45	10.25	16.2	13.65	23.35	23.65	12.75	464	85.3	1.2	68.1	221.3	14.6	27.5	189.5	200.24
S20	17.95	5.3	31.7	368.3	27.15	10.45	16.7	14.45	22.55	22.95	13.55	710.3	149.9	1.9	76.5	355.1	13.5	28	326.8	40.58
M21	17.45	5	30.9	366.2	26.35	10.35	16	14.05	22.05	22.45	13.15	733.9	140.9	2.4	75.6	370.8	13	30.4	323.6	72.11
M22	17.55	5.9	33.5	381.4	27.45	9.75	17.7	14.05	22.45	22.85	13.15	639.7	125.9	1.6	74.2	315.5	11.7	27.3	288.9	157.56
M25	16.35	12.7	40.1	631.5	33.25	1.65	31.6	8.75	24.75	24.85	8.75	719.1	121.7	7.6	62.7	319.3	37.5	41.4	319.3	1098.03
M28	17.15	12.4	42.9	544.6	32.55	3.65	28.9	13.35	24.45	24.45	10.55	431.1	66.2	8.5	43.7	162	43.3	43.3	130.4	983.07
M29	18.25	12.4	41.9	539.1	35.05	5.35	29.7	11.55	25.35	25.55	11.55	360	76.4	1.7	82.8	182.4	9.3	11.7	182.4	905.89
M30	19.05	8.1	42.6	349.5	29.15	10.25	18.9	15.55	23.55	23.95	14.75	379.2	76	1.3	85.3	199.4	5.7	7.3	195.2	234.78

Climate data extracted from CHELSA (Karger *et al.* 2017) and topographic data extracted from GMTED2010 (Danielson & Gesch 2011).

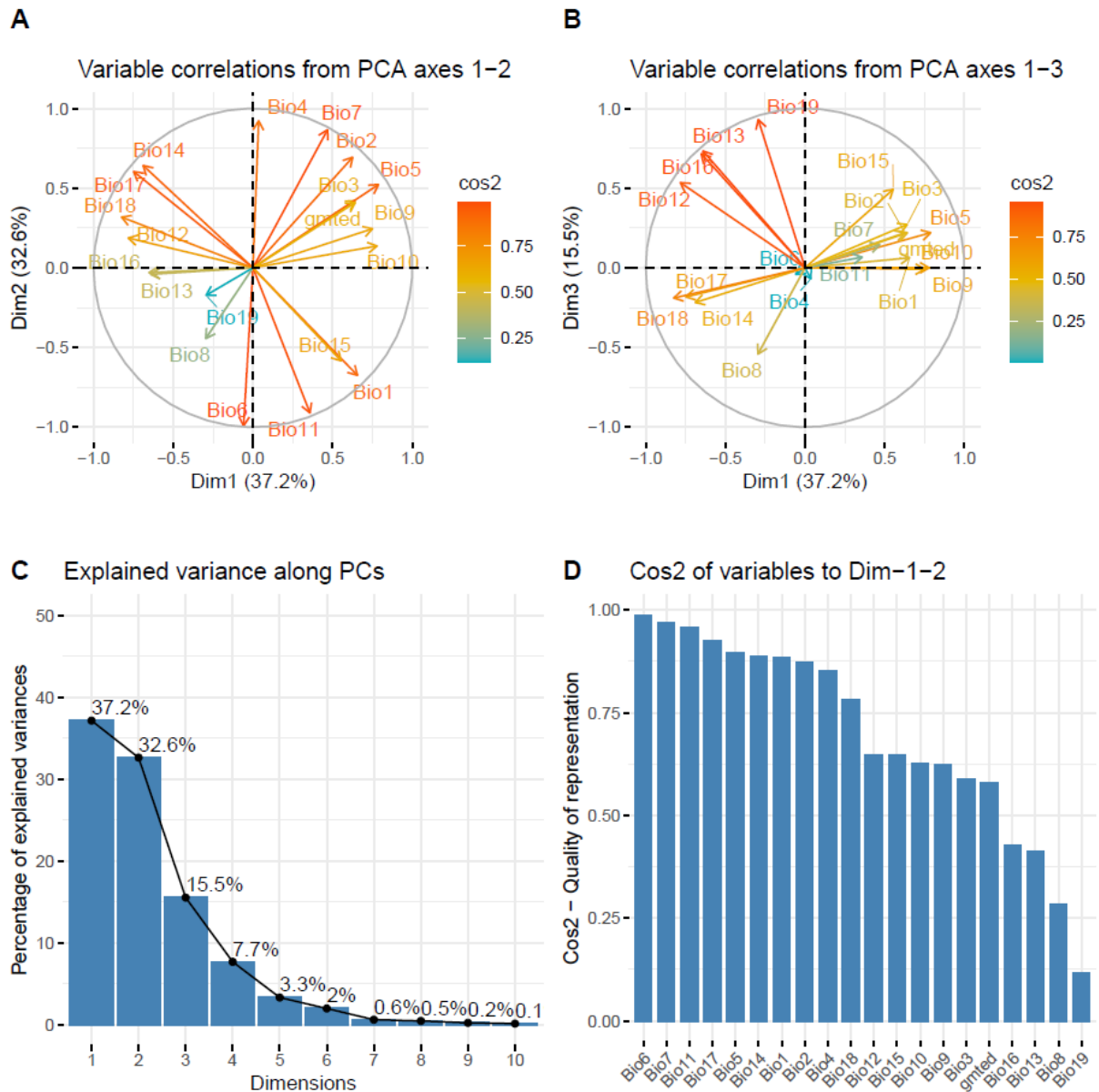


Figure S2. Principal Component analysis of all environmental data for the 26 sites sampled with (A) the variable correlations axes 1-2 (B) the variable correlations axes 1-3 (C) the explained variance for each dimension (D) the Cos2 of variables for Dim-1-2. Analysis based on Benjamin Dauphin script available here: <https://github.com/bcm-uga/SSMPG2022>

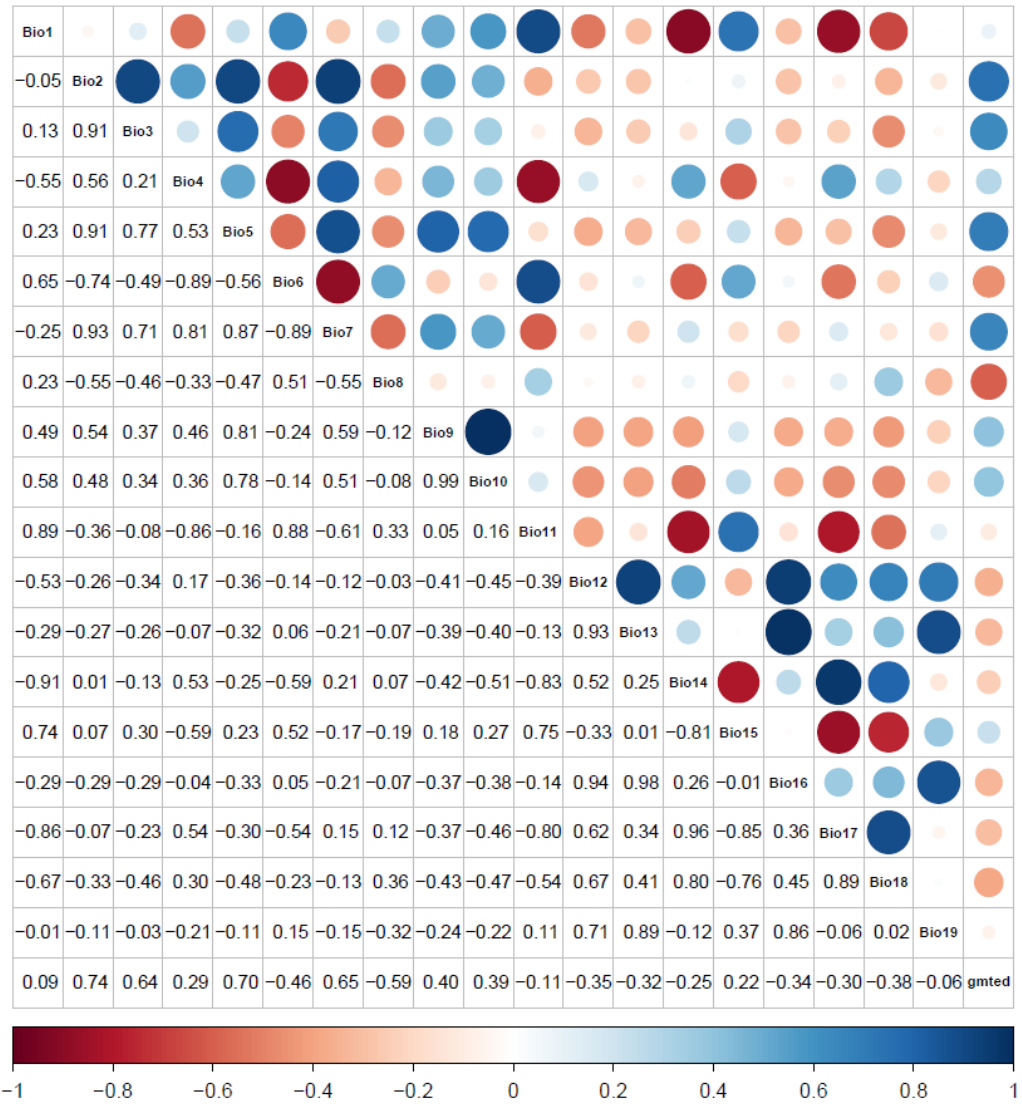


Figure S3. Pearson correlation matrix of all environmental data per sites.

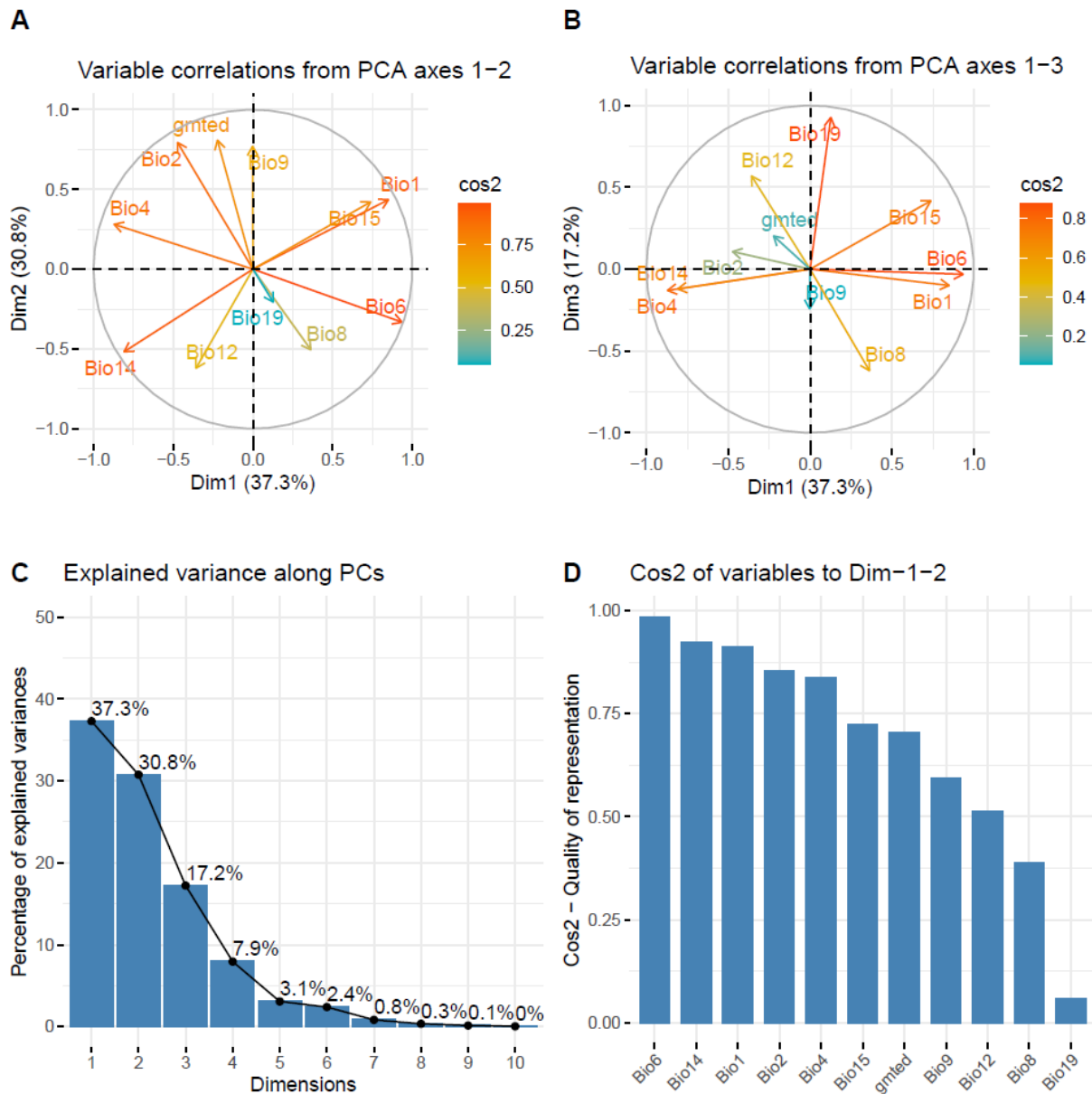


Figure S4. Principal Component analysis of reduce correlations environmental data (<75% Pearson Correlation) for the 26 sites sampled with (A) the variables correlation axes 1-2 (B) the variable correlations axes 1-3 (C) the explained variance for each dimension (D) the Cos2 of variables for Dim-1-2.

3 | Results

3.1 | Identification of candidate SNPs in wild and admixed populations

Table S3. Q-value threshold retained per variable tested during genome-environment association analysis for wild genotypes and admixed genotypes.

Environmental variable	q-value	
	Wild	Admixed
<i>all</i>	1.13E-03	9.74E-04
<i>temperature</i>	6.88E-04	6.78E-05
<i>precipitation</i>	2.52E-04	1.57E-04
<i>elevation</i>	3.82E-05	NA
<i>Bio1</i>	NA	NA
<i>Bio2</i>	1.17E-05	NA
<i>Bio4</i>	NA	NA
<i>Bio6</i>	1.11E-05	NA
<i>Bio8</i>	5.25E-05	5.59E-07
<i>Bio9</i>	NA	NA
<i>Bio12</i>	NA	1.40E-07
<i>Bio14</i>	1.29E-08	NA
<i>Bio15</i>	5.11E-05	NA
<i>Bio19</i>	NA	NA

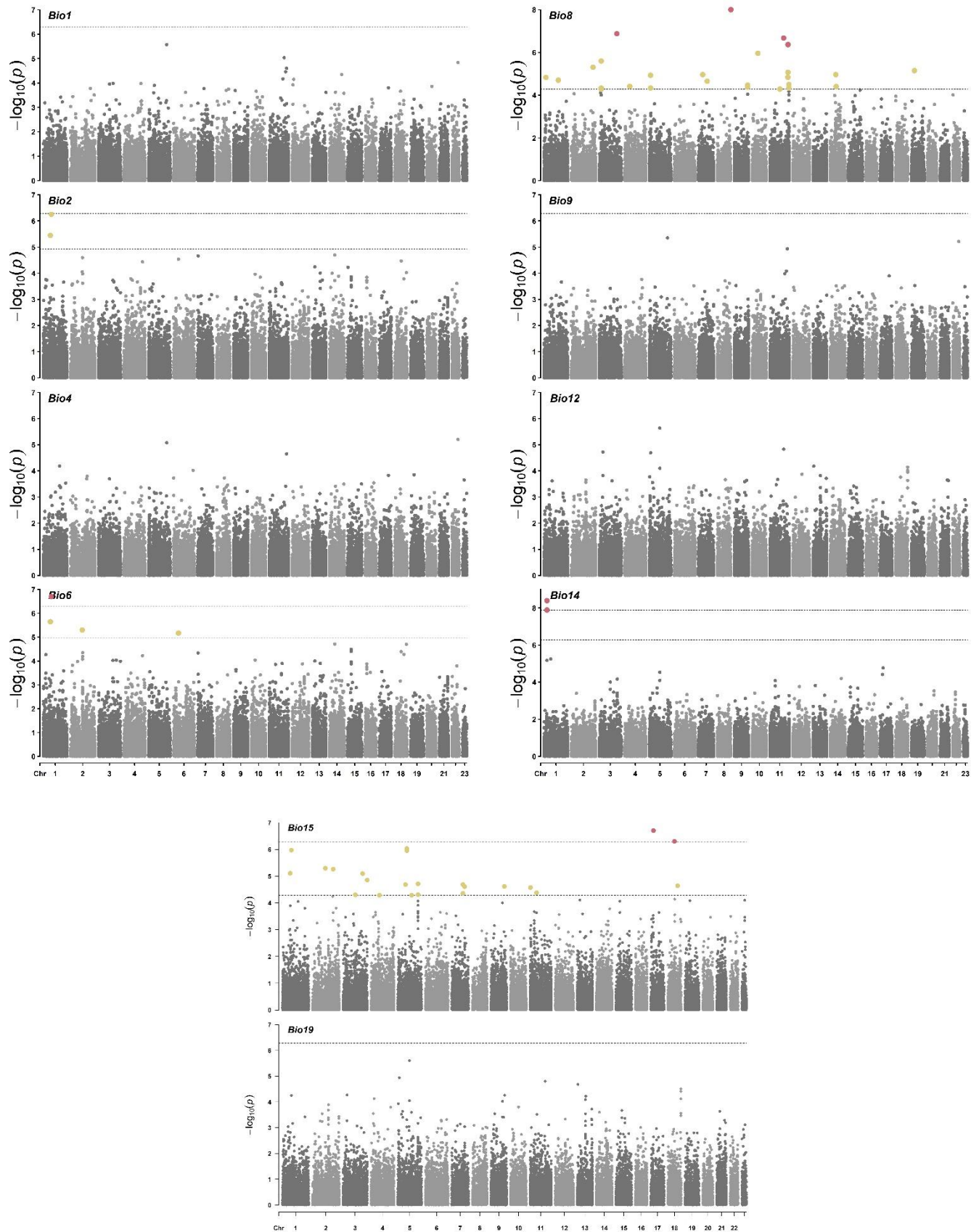


Figure S5. Manhattan plots of different genomic-environment associations (GEA) for western Mediterranean Basin of wild olive trees. Each point represents a SNP. The

yellow points represent minor outlier SNPs, with association score q -value < 0.1 . The pink points represent major outlier SNPs, with association score above p -value < 0.05 after Bonferroni correction with (A) Wild genotype associations to environmental variables in order “All”, “Temperature”, “Precipitations” and “Topography”. (B) Admixed genotype associations to environmental variables in order “All”, “Temperature”, “Precipitations” and “Topography”.

SUPPLEMENTARY CHAPTER 3

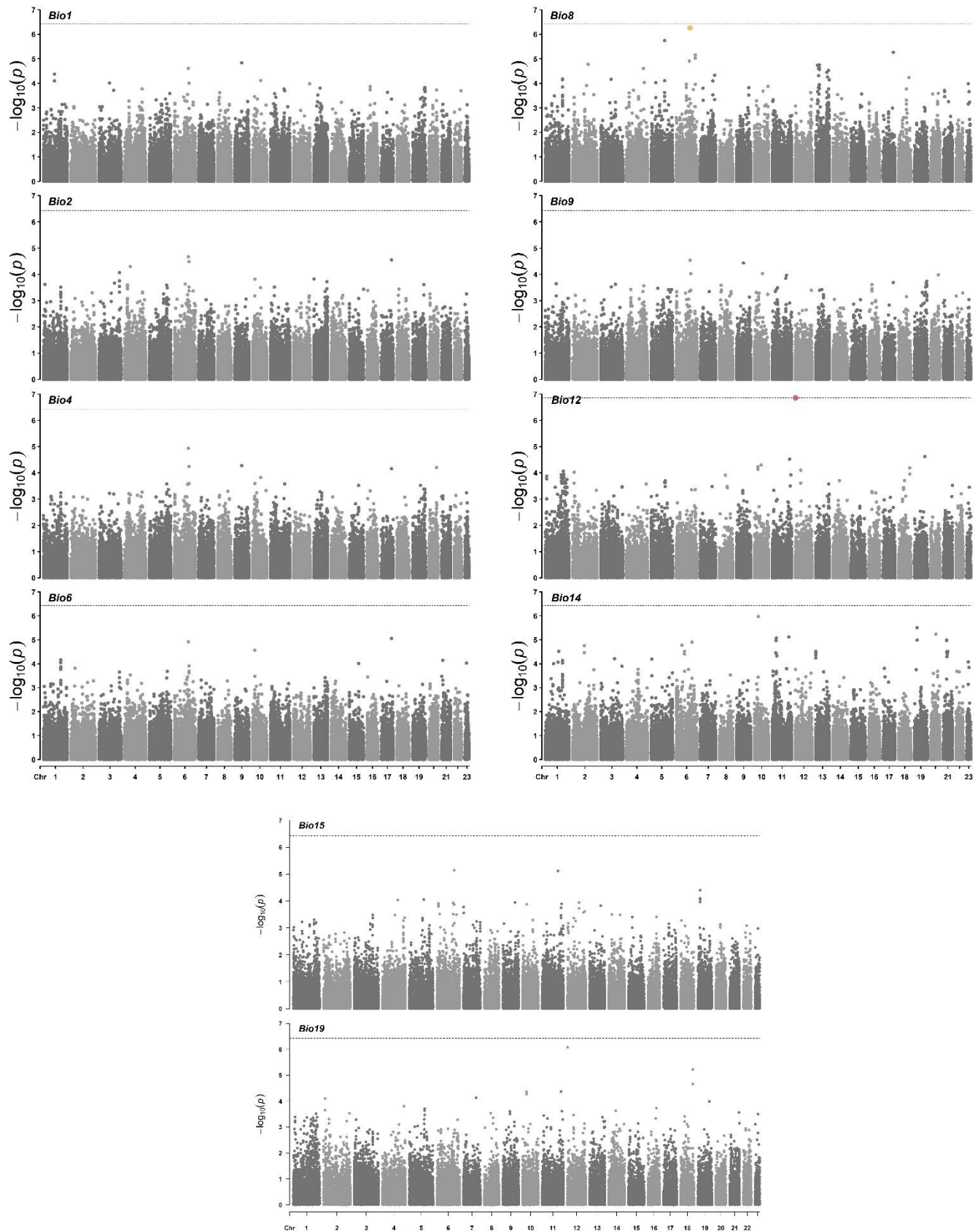


Figure S6. Manhattan plots of different genomic-environment associations (GEA) for western Mediterranean Basin of admixed olive trees. Each point represents a SNP.

The yellow points represent minor outlier SNPs, with association score q -value < 0.1 . The pink points represent major outlier SNPs, with association score p -value < 0.05 after Bonferroni correction with (A) Wild genotype associations to environmental variables in order “All”, “Temperature”, “Precipitations” and “Topography”. (B) Admixed genotype associations to environmental variables in order “All”, “Temperature”, “Precipitations” and “Topography”.

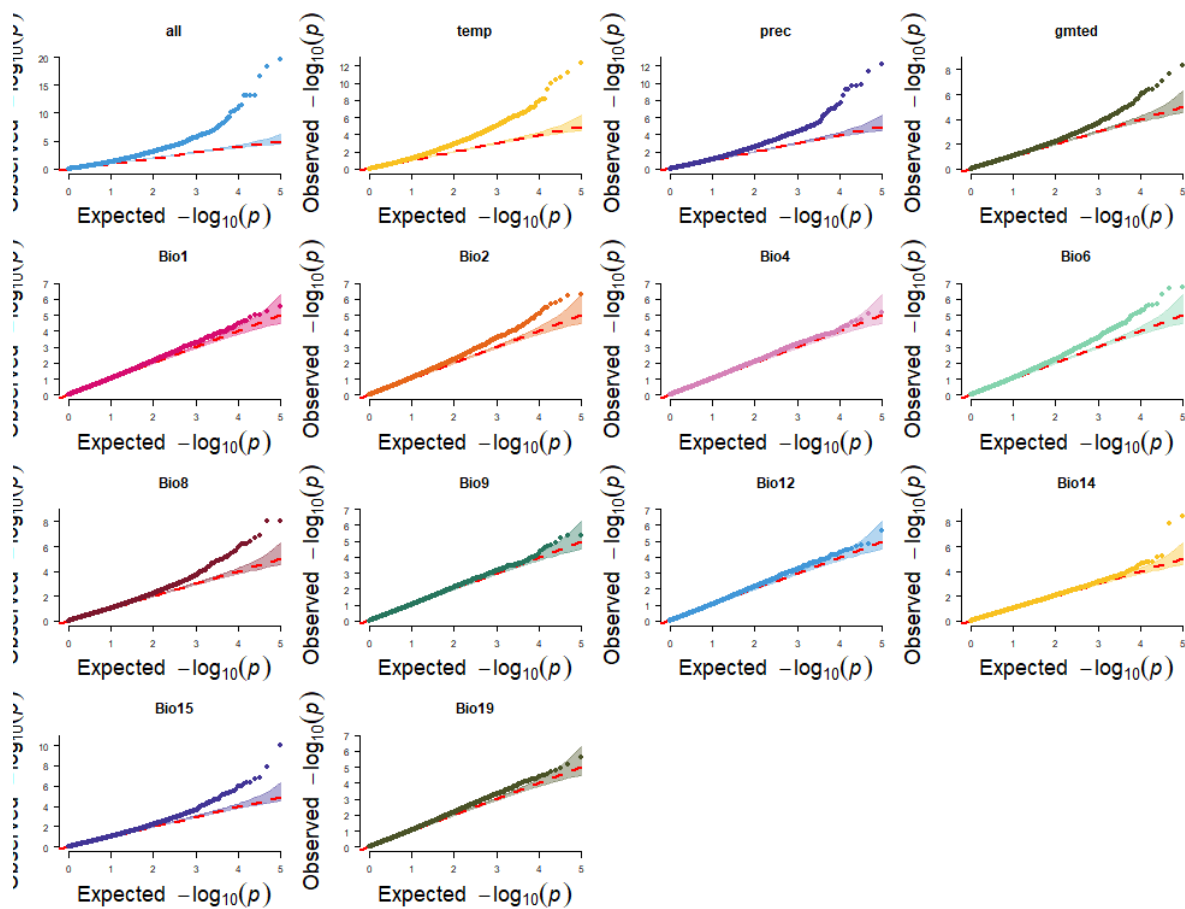


Figure S7. QQplot of each environmental variable tested for genomic environment association analysis with wild genetic group.

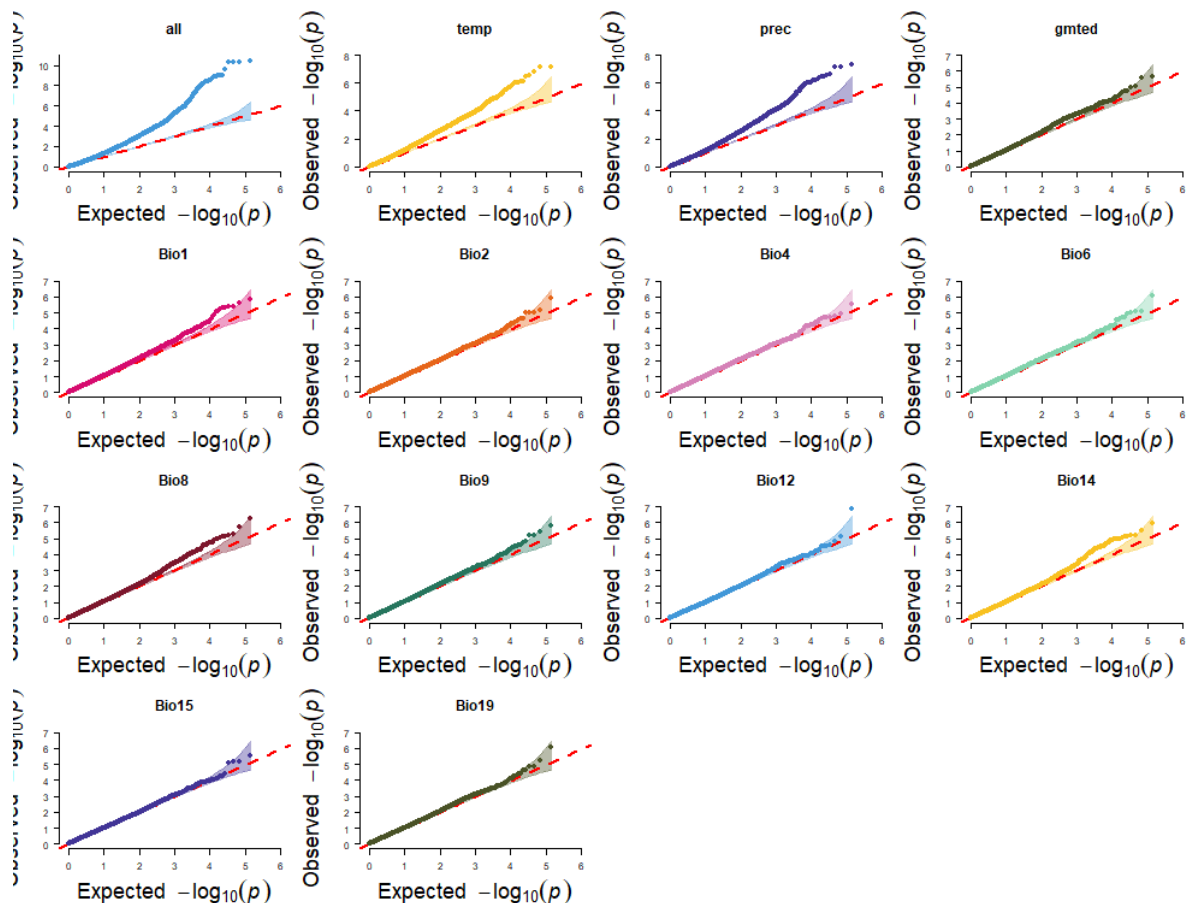


Figure S8. QQplot of each environmental variable tested for genomic environment association analysis with admixed genetic group.

SUPPLEMENTARY CHAPTER 3

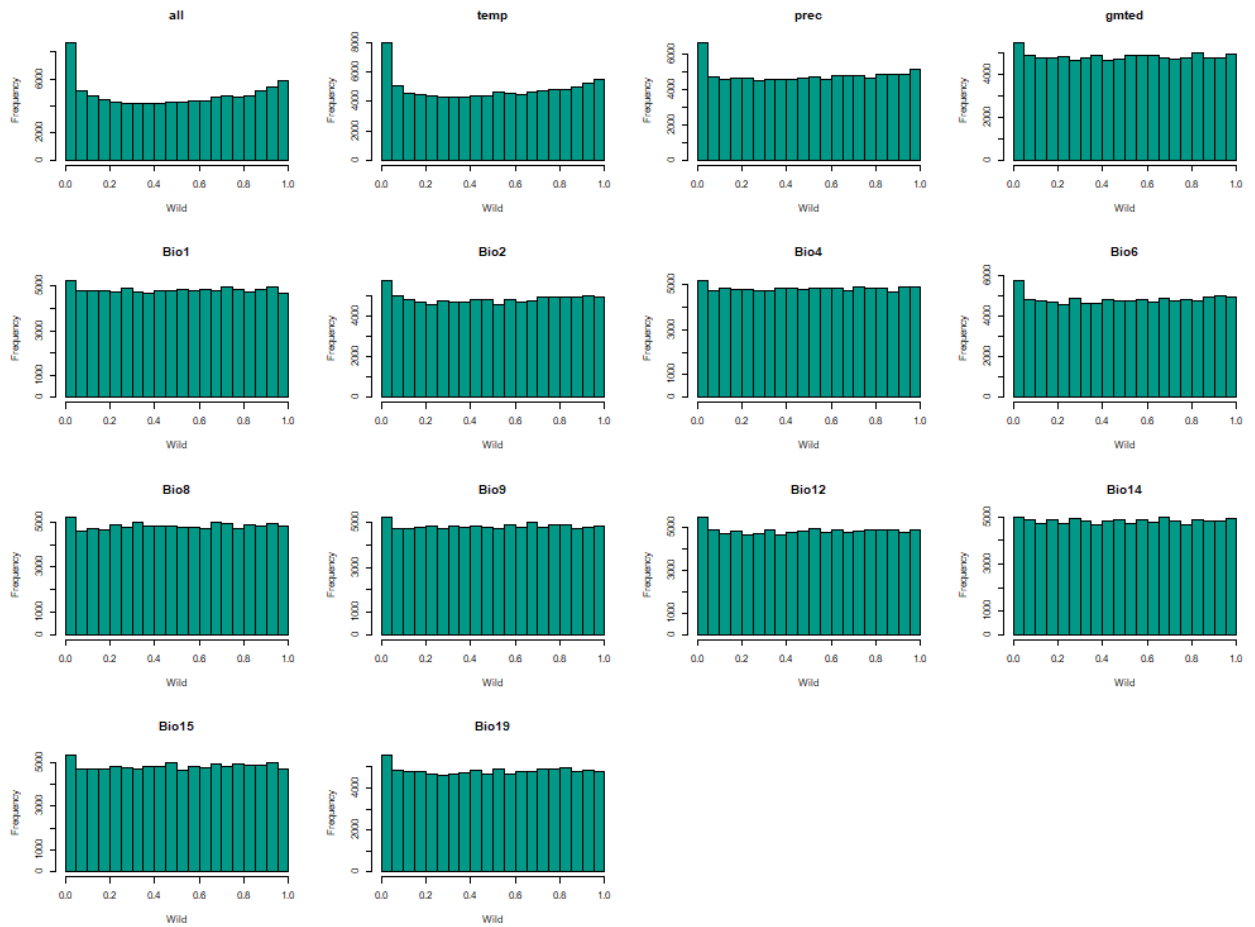


Figure S9. Histogram of *p-value* frequencies for each environmental variable tested for genomic environment association analysis with wild genetic group.

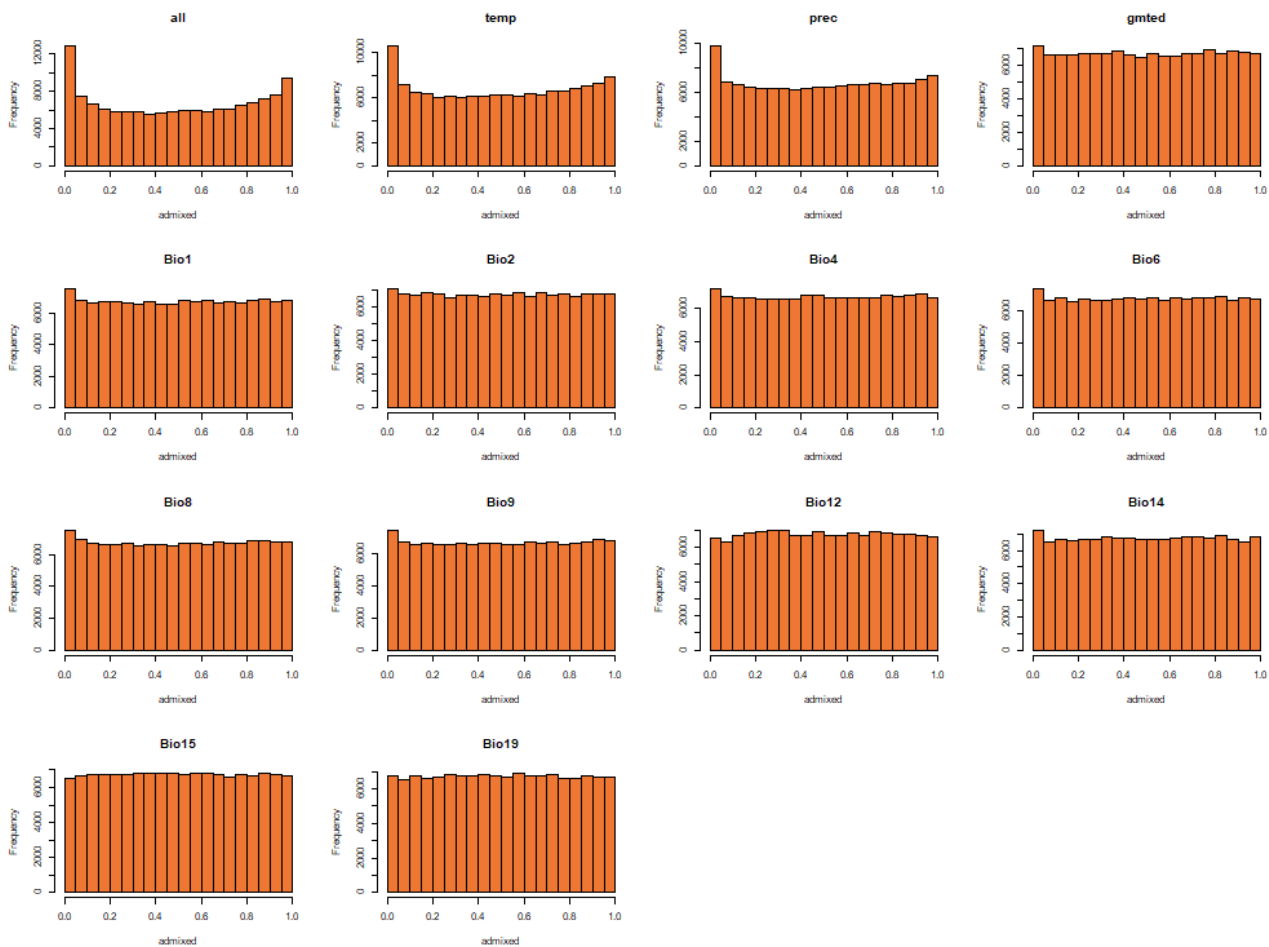
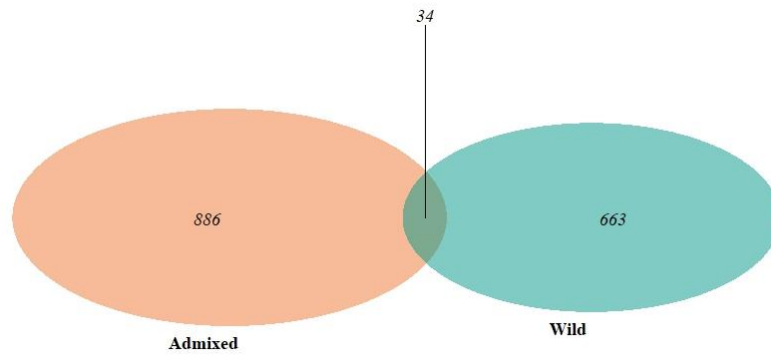
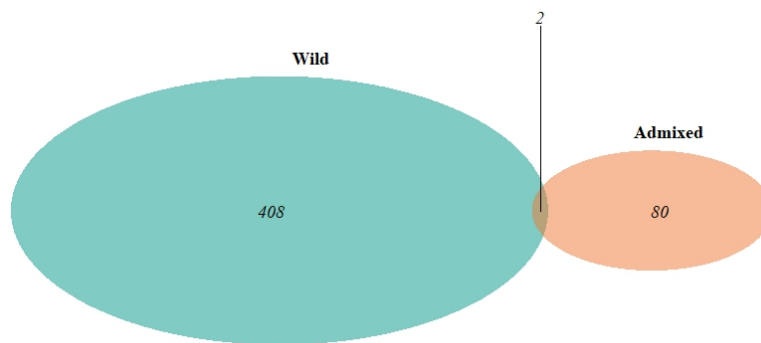


Figure S10. Histogram of *p-value* frequencies for each environmental variable tested for genomic environment association analysis with admixed genotypes.

A



B



C

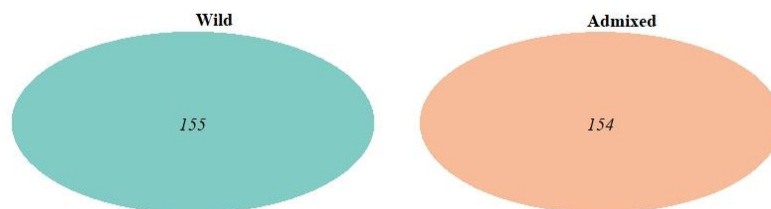


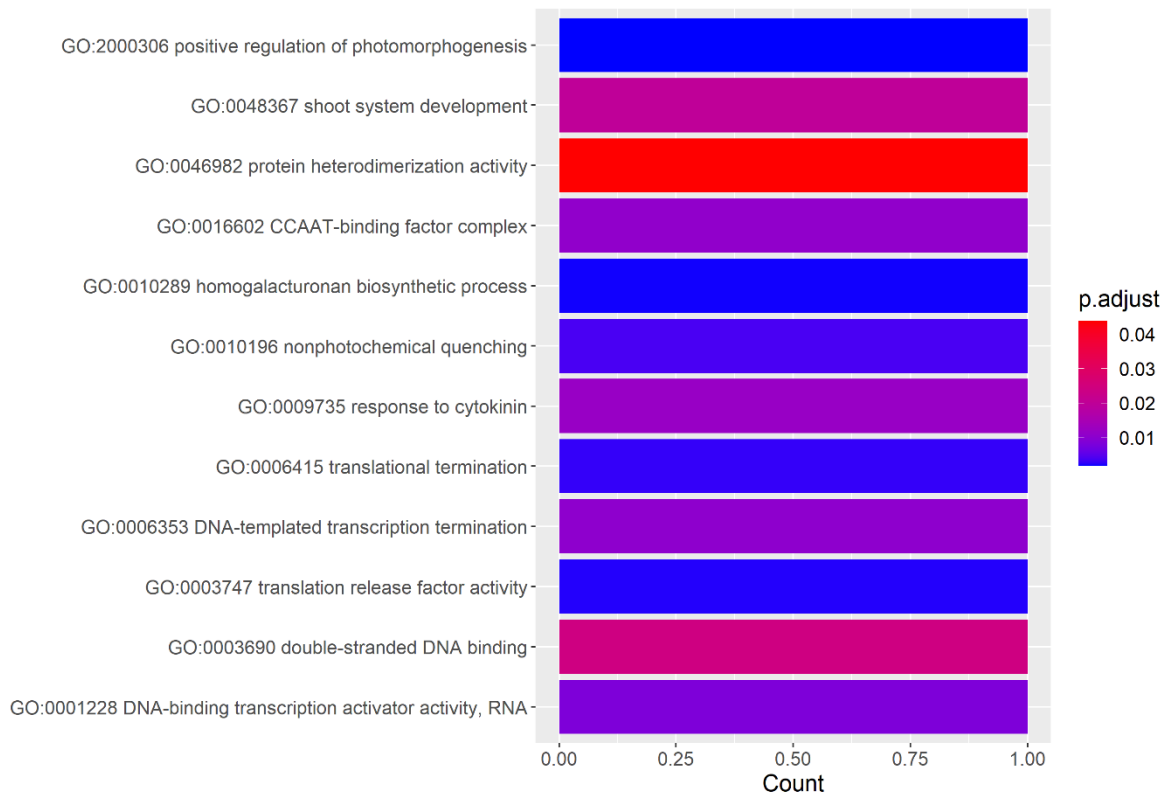
Figure S11. Venn Diagram of common genes between wild and admixed genotypes associated to environmental variables in genomic-environment analysis for (A) "All" variables, (B) "Temperature-related" variables and (C) "Precipitation-related" variables.

3.2 | Description of associated SNPs and genes in wild and admixed populations

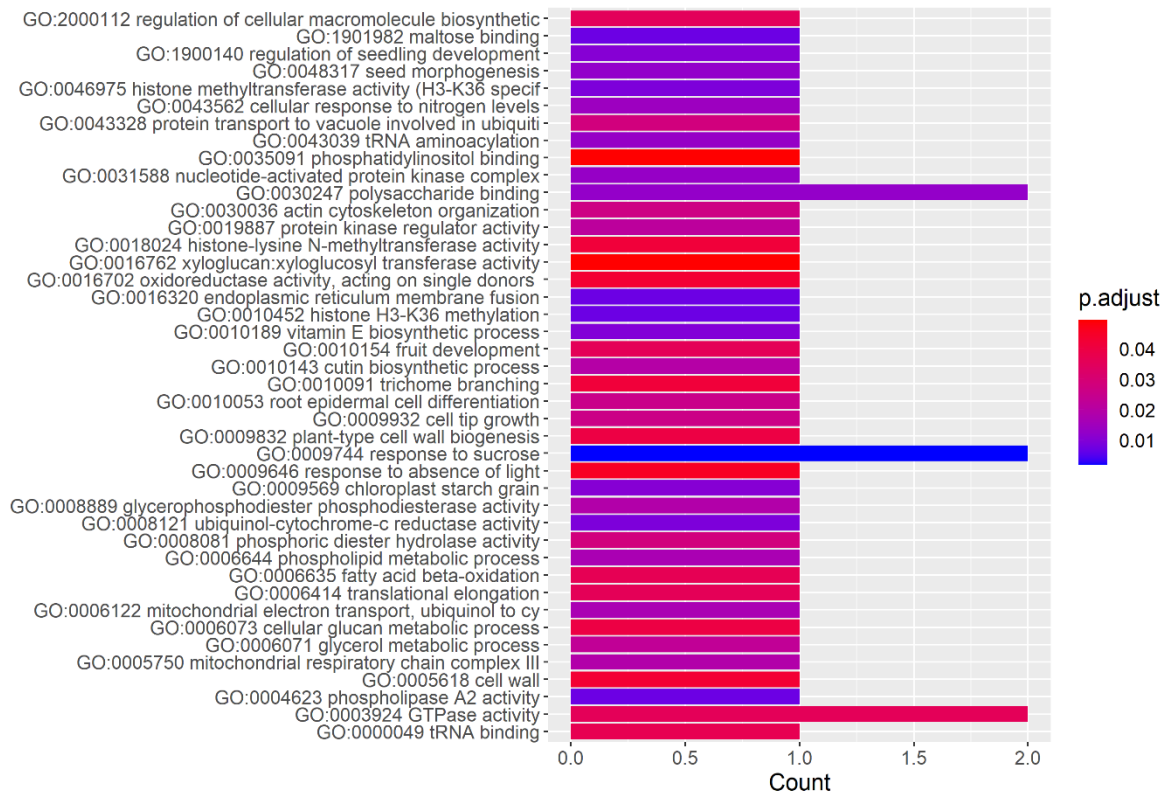
Table S4. Summary of gene information for SNPs detected as associated to climate for wild and admixed olives *O. europaea* L. Table available here: <https://zenodo.org/doi/10.5281/zenodo.10044433>

SUPPLEMENTARY CHAPTER 3

A



B



C

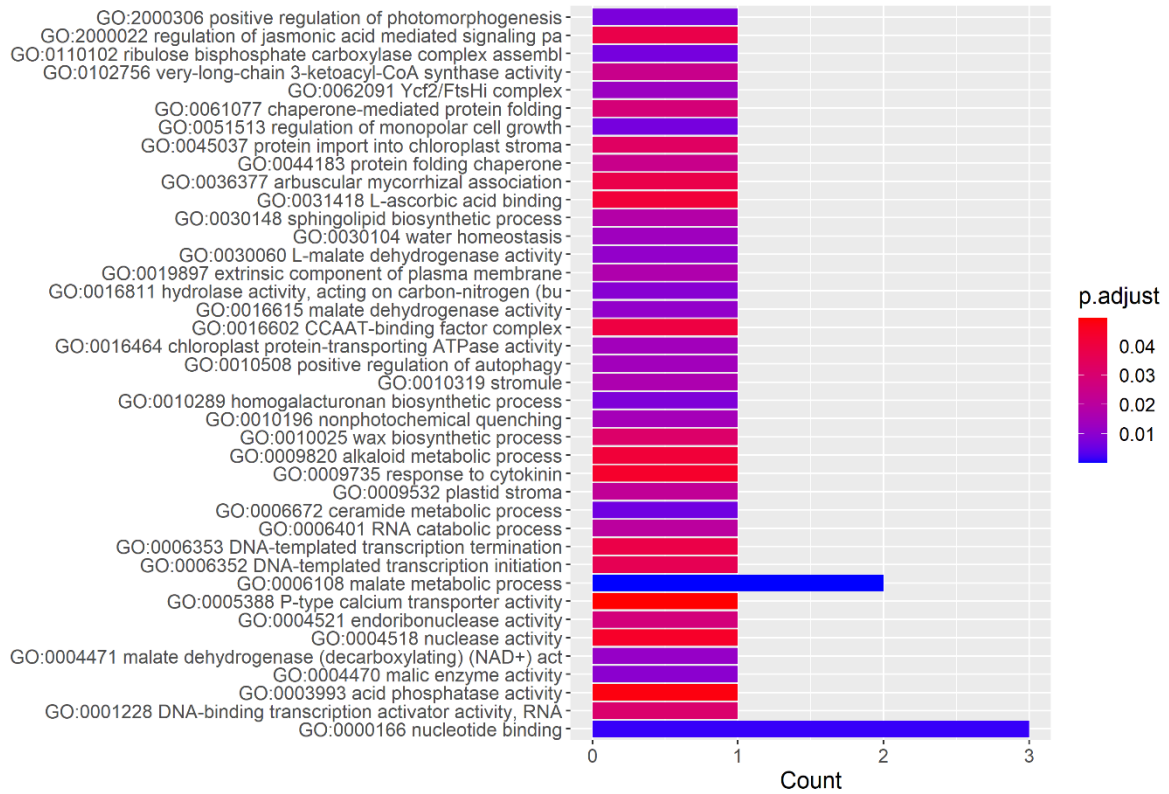


Figure S12. Gene ontology (GO) term enrichment analysis based on SNPs associated with environmental variables for wild genetic groups. (A) SNPs associated with “Bio6” for wild, (B) SNPs associated with “Bio8” for wild and (C) SNPs associated with “Bio15” for wild.

3.3 | SNP characterisation and localisation along the geographic gradient according to wild and admixed groups

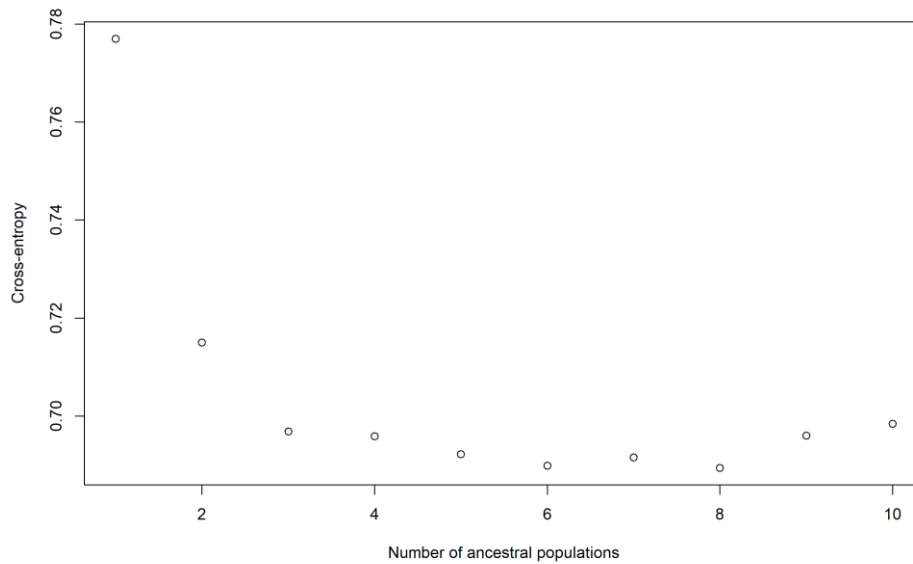


Figure S13. sNMF cross-entropy with wild SNPs tested as associated to climate (n = 1400).

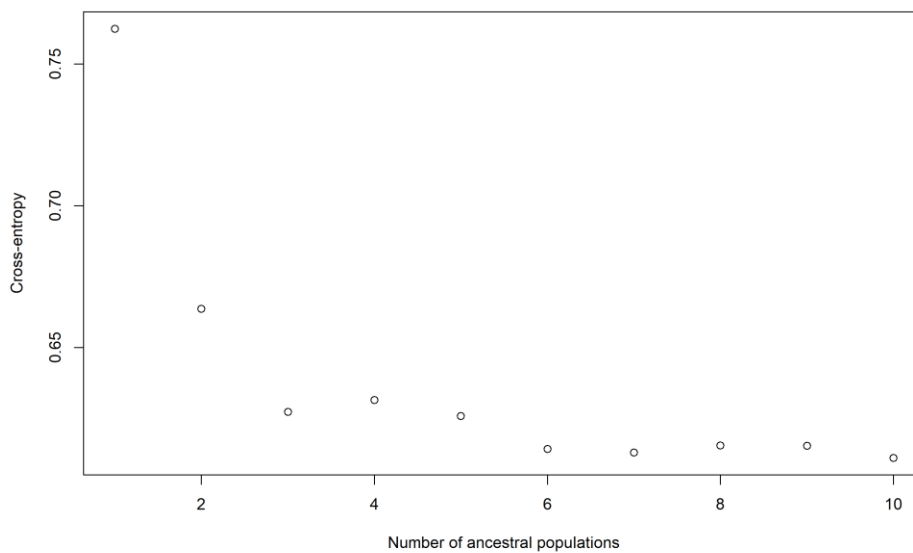


Figure S14. sNMF cross-entropy with major wild SNPs tested as associated to climate (n = 86).

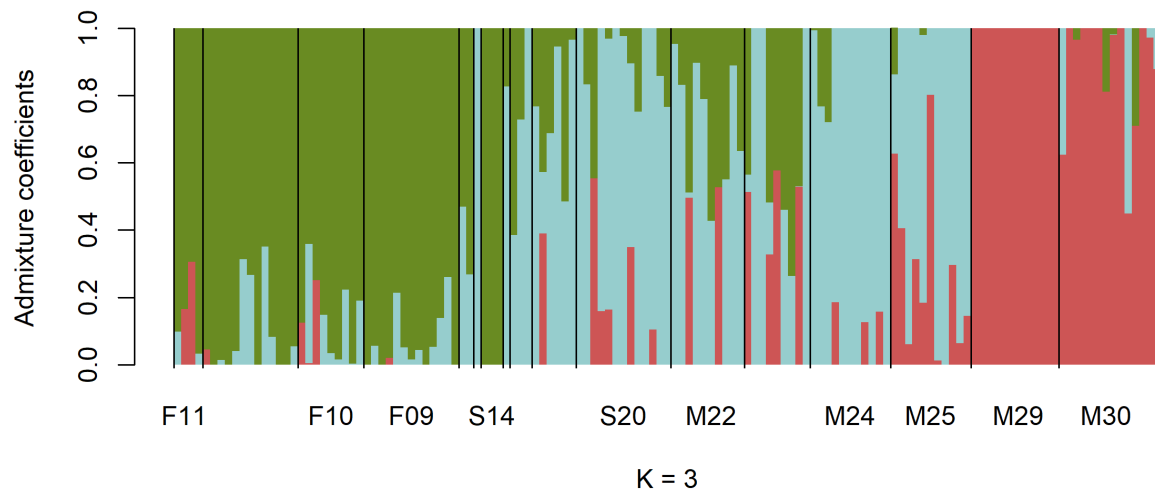


Figure S15. Genetic structure using major wild SNPs tested as associated to climate ($n = 86$) for 135 olive trees.

SUPPLEMENTARY CHAPTER 3

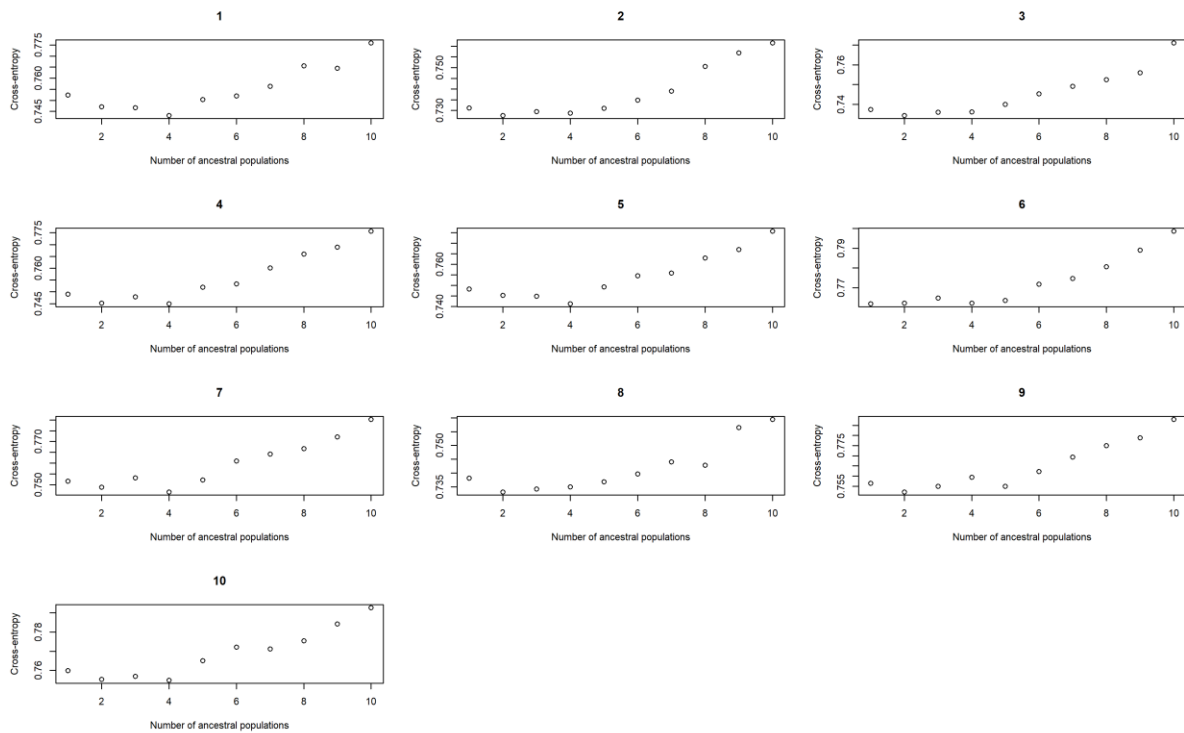


Figure S16. Cross-entropy for ten random drawn of SNPs within wild dataset ($n = 1400$).

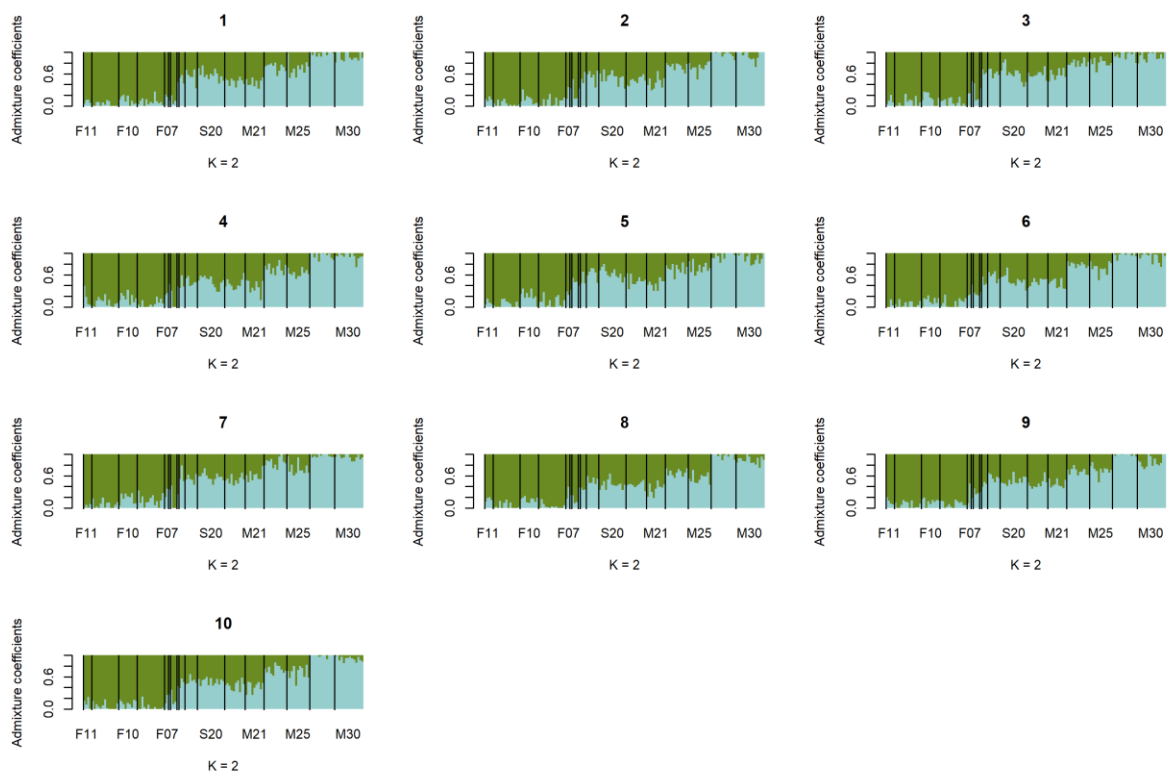


Figure S17. Genetic structure obtained with ten random drawn of SNPs within wild dataset at $K = 2$ ($n = 1400$).

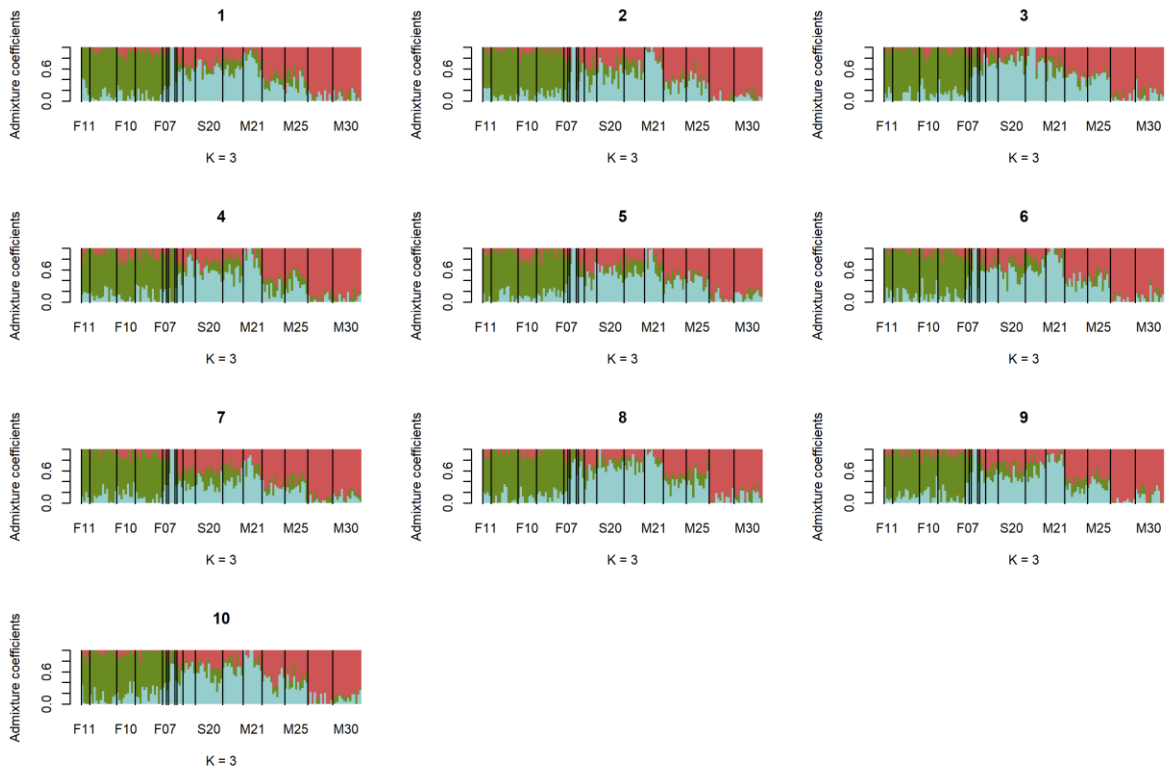


Figure S18. Genetic structure obtained with ten random drawn of SNPs within wild dataset at $K = 3$ ($n = 1400$).

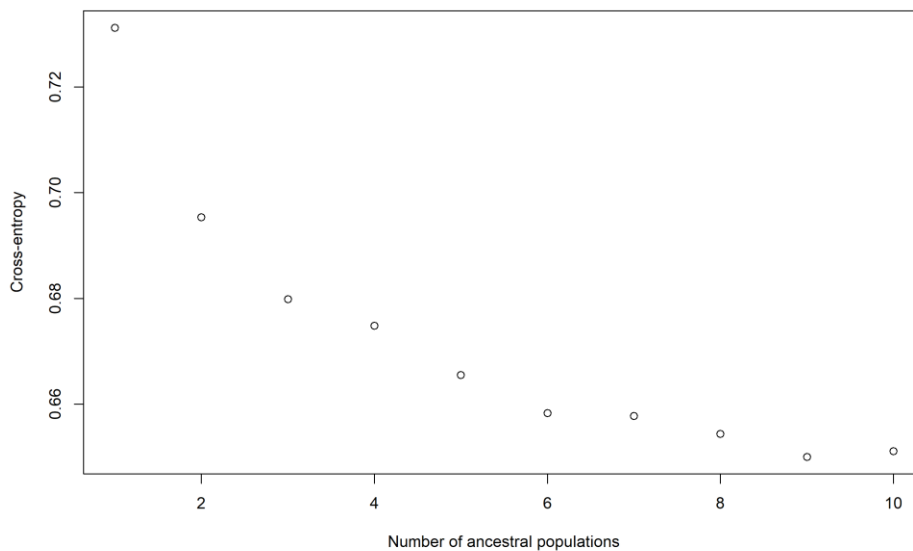


Figure S19. sNMF cross-entropy using corresponding SNPs tested as associated to climate within the wild dataset, found in the admixed dataset ($n = 1125$).

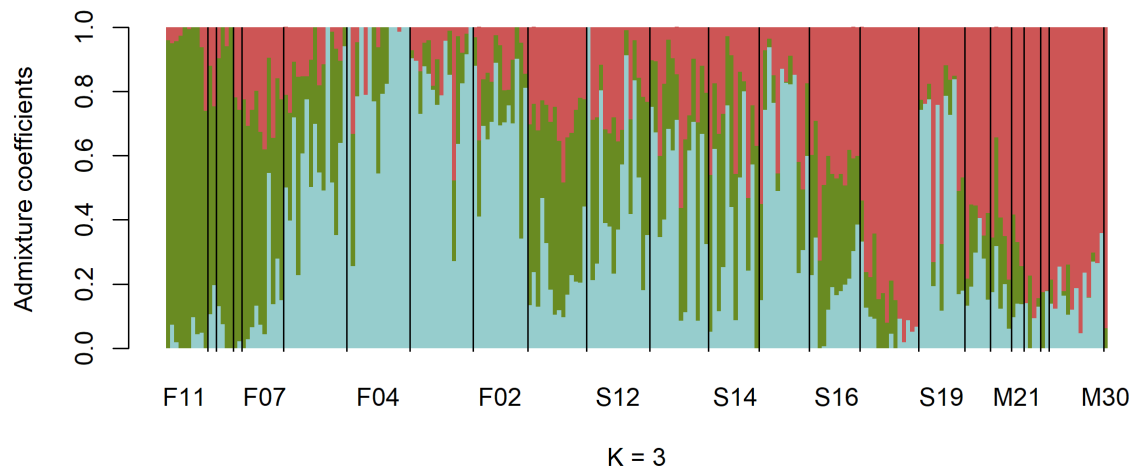


Figure S20. Genetic structure using corresponding SNPs tested as associated to climate within the wild dataset, found in the admixed dataset at $K = 3$ ($n = 1125$).

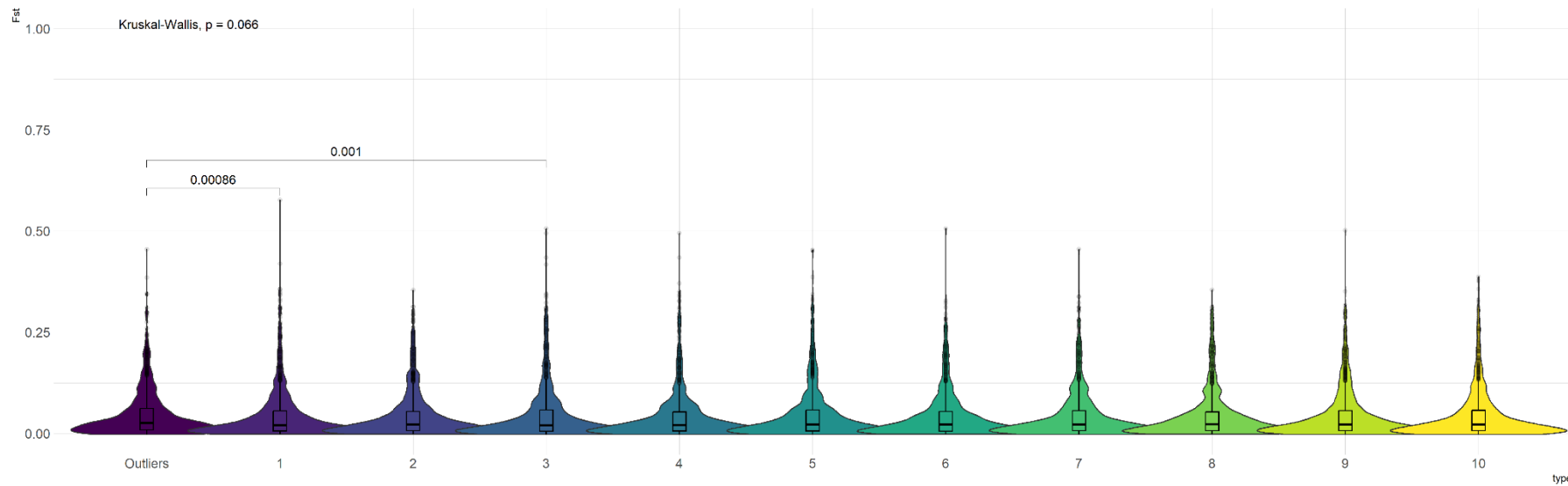


Figure S21. Violin plot of F_{ST} between wild and admixed group for SNPs associated to climate within wild GEA results and 10 random drawn of SNPs from wild dataset. The difference was statically tested with Wilcoxon test and Krsukal-Wallis, with Bonferroni correction.