



**HAL**  
open science

# Adaptation locale des téosintes *Zea mays* ssp. *parviglumis* et *Zea mays* ssp. *mexicana* le long de gradients altitudinaux

Margaux-Alison Fustier

► **To cite this version:**

Margaux-Alison Fustier. Adaptation locale des téosintes *Zea mays* ssp. *parviglumis* et *Zea mays* ssp. *mexicana* le long de gradients altitudinaux. Biodiversité et Ecologie. Université Paris Saclay (COmUE), 2016. Français. NNT: 2016SACLS147. tel-01451110

**HAL Id: tel-01451110**

**<https://theses.hal.science/tel-01451110v1>**

Submitted on 31 Jan 2017

**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.

2016SACLS147

THESE DE DOCTORAT  
DE L'UNIVERSITE PARIS-SACLAY,  
préparée à l'Université Paris-Sud

ÉCOLE DOCTORALE N°567  
Sciences du Végétal: du gène à l'écosystème

Spécialité de doctorat de Biologie

Par

**Madame Margaux-Alison Fustier**

Adaptation locale des téosintes *Zea mays* ssp. *parviglumis* et ssp. *mexicana*  
le long de gradients altitudinaux

**Thèse présentée et soutenue à l'unité Génétique Quantitative et Evolution - Le Moulon,  
le 16 juin 2016 :**

**Composition du Jury :**

Monsieur Castric, Chargé de recherche CNRS, EEP, Lille	Rapporteur
Monsieur Roux, Chargé de recherche CNRS, LIPM, Toulouse	Rapporteur
Madame Dillmann, Professeure Univ. Paris-Sud, GQE-Le Moulon, Gif-sur-Yvette	Présidente du Jury
Madame Giraud, Directrice de recherche CNRS, ESE, Orsay	Examinatrice
Monsieur Nabholz, Maître de conférence Univ. Montpellier II, ISEM, Montpellier	Examineur
Madame Tenaillon, Directrice de recherche CNRS, GQE-Le Moulon, Gif sur Yvette	Directrice de thèse
Madame Manicacci, Maître de conférence Univ. Paris-Sud, GQE-Le Moulon, Gif sur-Yvette	Invitée



Ces dernières années ont été riches autant d'un point de vue professionnel que personnel. Je tiens à adresser mes plus vifs remerciements à Maud Tenaillon qui m'a donné l'opportunité de réaliser mon doctorat. Son énergie débordante tire les personnes vers le haut et m'a permis de donner le meilleur de moi-même. Maud m'a bien sûr encadrée scientifiquement, mais je tiens également à la remercier pour ses qualités humaines et pour son écoute sincère. Je sens que tu t'es toujours arrangée pour me donner les meilleures chances pour la suite de ma carrière. Je tiens aussi à remercier Domenica pour son encadrement précieux, ses explications très claires et pour ses nombreuses corrections au moment de la rédaction. Votre disponibilité à toutes les deux a été incroyable ces derniers mois, et m'a permis d'aborder la période de rédaction sereinement. Un grand Merci. Je remercie également Jean-Tristan Brandeburg pour toute l'aide bio-informatique et les conseils scientifiques qu'il m'a prodigués au quotidien.

Je remercie Vincent Castric et Fabrice Roux d'avoir accepté d'être les rapporteurs de ma thèse et Christine Dillmann, Tatiana Giraud et Benoit Nabholz d'avoir accepté d'être les examinateurs. Je veux aussi remercier Benoit Nabholz ainsi que Sylvain Glémin pour le stage de master 2 que j'ai réalisé sous leur direction, et pour m'avoir encouragé à faire mon doctorat avec Maud. Merci à Yves Vigouroux pour m'avoir mis en relation avec Maud pendant mon master. Je remercie Antoine Branca, Christine Dillmann, Simon Boitard, Renaud Vitalis d'avoir été dans mon comité de suivi de doctorat. Je remercie la direction de l'école doctorale Sciences du Végétal pour leur soutien, ainsi que tous les membres de l'association de doctorants Doc'en Herbe qui organisent chaque année des rencontres entre doctorants très réussies.

Je suis convaincue que la recherche est loin d'être un travail solitaire. Merci particulièrement à Simon Boitard et Yves Vigouroux pour leur implication dans ce projet. Je remercie Jessica Mélique et Jason Lapeyronie pour leur contribution à ce travail. Mon travail de doctorat a été rendu possible grâce à l'aide technique que j'ai reçue des techniciens du Moulon. Merci à Hélène Corti, Agnès Rousselet et Fabrice Dumas pour les innombrables extractions et PCR que vous avez réalisées pour mon projet. Merci Hélène, Fabrice et Anthony Venon pour avoir effectué les lectures de microsattellites. Anthony, c'était un sacré bizutage pour ton arrivée au Moulon, et tu t'es impliqué à 200%. Chapeau. Merci à Matthieu Falque pour nous avoir aidé à concevoir le dispositif de génotypage. Je remercie également Carine Remoué, Agnès Rousselet, Martine Le Guilloux, Valérie Combes et Charlotte Urien pour leur aide à mon arrivée quand j'ai commencé les manips au laboratoire, et Mathilde

Latron avec qui j'ai couplé mes manipulations, c'était plus sympa de travailler à deux. Merci aussi à Yannick de Oliviera, Pierre Montalent, Héloïse Giraud, Christophe Lecarpentier et Fabien Laporte pour leur aide en python et R quand Jean-Tristan n'était pas disponible.

Mi trabajo de doctorado contiene experimentos y viajes realizados en México. Estoy ansiosa por expresar mi gratitud a Luis Eguiarte Fruns, Valeria Souza Saldivar y Erika Aguirre Planter por sus considerables ayudas a organizar mis estancias en México y acogerme en el Laboratorio de Evolución Molecular y Experimental. Gracias a Jonas Aguirre por sus discusiones acerca de los teocintles y su generosa hospitalidad. Gracias a Salvador Montes Hernandez y María Guadalupe Herrera Hernández por la gestión de los experimentos, y por su recibimiento en Celaya. Progreso mucho con el idioma gracias a sus enseñanzas y su paciencia. Gracias a Salvador y su familia por hacerme descubrir Celaya, y las maravillosas ciudades de Queretaro y San Miguel de Allende. Gracias también a Manuel por su agradable compañía y su implicación en el proyecto. Gracias a Angélica Cibrián Jaramillo y su equipo por alojarme en su laboratorio, y muchas gracias a Flor Zamudio por su bienvenida y su ayuda técnica.

Je remercie Clémentine Vitte et Karine Alix avec qui j'ai travaillé en collaboration avec Maud sur un article de revue, et mon frère Sylvain Fustier pour la séance photo des fruits et céréales divers.

Je tiens à remercier toute l'équipe administrative. Merci à Valérie Lespinas pour sa disponibilité. Grâce à Valérie les formalités administratives paraissent presque simples. Merci également à Cyril Martin, Khalid Ouskou, Stéphanie Avodagbe et Frédéric Huynh avec lesquels les réponses aux questions administratives vont généralement de paire avec moult gourmandises. Ce bureau est magique.

Le Moulon est un lieu de travail agréable, grâce à l'ensemble des personnes qui y travaillent. Je tiens à remercier tous les membres du laboratoire, et ainsi que ceux avec qui j'ai partagé des activités : merci aux membres de l'équipe de Basket et à ses formidables organisateurs Yannick et Pierre, merci à Cyril Bauland pour les très intéressantes séances d'œnologie, merci aux gymastes d'avoir partagé les courbatures (et pardon à la nouvelle équipe d'avoir déserté les séances en faveur du basket) et merci aux jardiniers d'avoir partagé un bout de parcelle, les salades étaient délicieuses. Merci à Rozenn Le Guyader et Valérie Lespinas pour la gestion des commandes de légumes. Egalement un grand merci à Valérie Combes : c'est toujours un délice de découvrir tes gâteaux au petit matin.

Maintenant, place aux jeunes ! Je vous remercie pour votre accueil à mon arrivée au Moulon. Certains sont partis très vite après mon arrivée, mais j'ai pu lier de véritables relations d'amitié avec d'autres. Héloïse et Christophe, ce fut un plaisir de faire partie de la génération 2012 avec vous. Héloïse, on t'en a fait peut-être un peu baver avec Christophe et les autres, mais c'est qu'au fond, on t'aime bien. Christophe, mon « Michel » ego. Je prends un peu d'avance (ou plutôt moins de retard...) dans la fin de l'épisode doctorat. Je te souhaite beaucoup de courage pour finir ta thèse, tu verras, tu vas clore la cuvée 2012 en beauté! Charlotte, merci pour ton humour, je t'assure, moi, je te trouve vraiment drôle. Tu possèdes un flair incroyable pour déceler quand le moral n'est pas au top, merci de m'avoir soutenu pendant ces moments. Jean-Tristan, je suis très heureuse d'avoir partagé notre algeco/bureau pendant ces années. Ta bonne humeur quotidienne (...), ton humour mordant et la musique en boucle ont contribué à faire de notre bureau un agréable lieu de travail. Mais je veux surtout te remercier pour m'avoir soutenu et d'avoir été là dans les moments difficiles. Natalia, je suis ravie que tu nous ais rejoints. J'espère que tu tiens le coup d'être tombée dans notre bureau et bon début de doctorat! Zeineb, j'ai été ravie d'assister à ton mariage. L'accueil de ta famille et de tes amies a été incroyable et m'a beaucoup touchée. Je te souhaite une bonne continuation de doctorat. Courage! Pierre, merci pour les séjours à La Palmyre, je suis d'accord, ce sont vraiment les meilleurs bulots du monde. Fabien, vivement mon rendu de thèse, on a une sacrée liste de séries et de films à voir, merci d'être aussi froussard que moi. Je remercie Yannick, Stéphanie, Paulina, Bubar, Anthony, Xavier et Pierre Rivière pour les moments passés ensemble.

Christophe, Fabien, Héloïse, Pierre et Zeineb : je vous remercie pour les nombreux transports en voiture, vous m'avez ôté plus d'une fois une épine du pied.

Mi estancia en el DF ha sido como un regalo. La gente allá es formidable. Gracias a todos los jóvenes del Laboratorio de Evolución Molecular y Experimental. En particular, gracias a Cecilia, Mirna y Feli por su bienvenida en mi primera estancia en el DF. Yocelyn, gracias por hacerme descubrir los Pumas, soy completamente fan! Gracias a Doña Silvia por su bienvenida y darnos los boletos para el estadio. Gracias Yocelyn, Mirna, Helena, Gaby, Ale, Feli, Paty, Roberto, Jaime, Enrique, Gabriel y todos los otros por hacerme descubrir la fabulosa comida mexicana : pozole, barbacoa, birria, camote (no, gracias), las cantinas, las cervezas, el mezcal.... Me gusto mucho su país!

Merci Eodiah, Louise-Roxane, Charlotte et Manon pour avoir gardé mon tigre miniature lors de mes séjours au Mexique. Je vous ai laissé Ortie les yeux fermés. A vrai dire, j'étais plus inquiète pour vous et vos plantes, mais apparemment, vous ne nous en voulez pas trop. Merci et désolée...

Merci Cyril pour ta présence. J'ai pu déconnecter du doctorat lorsque l'on se voyait, même dans les périodes les plus stressantes. Merci d'avoir accepté sans trop râler mon retrait des derniers temps, et d'avoir géré l'intendance. Je retire ce que j'ai dit, les patates, c'est bon.

Mes derniers remerciements sont adressés à ma famille. Merci à ma sœur Louise-Roxane et mon frère Sylvain. Votre présence sur Paris a largement contribué à ma montée « à la capitale ». Merci d'être là (même si loulou est partie depuis...), et de me laisser parler « science ». Enfin, je remercie mes parents. Vous m'avez continuellement soutenu et encouragé dans mes études. Vous m'avez laissé une liberté totale dans mes choix, et m'avez toujours aidé dans les déménagements conséquents. Je suis arrivée ici grâce à vous, et je vous en remercie du fond du cœur.







<b>I. INTRODUCTION.....</b>	<b>5</b>
I.1 ETUDIER L'ADAPTATION LOCALE .....	7
I.1.1 Adaptation locale .....	7
I.1.1.a Définition et mécanisme de mise en place de l'adaptation locale.....	7
I.1.1.b Mise en évidence de l'adaptation locale.....	8
I.1.1.c Plasticité phénotypique.....	11
I.1.1.d Adaptation locale versus plasticité phénotypique .....	13
I.1.1.e Adaptation locale, un processus dynamique dans le temps et l'espace.....	15
I.1.1.f Deux modèles d'adaptation locale : neutralité conditionnelle et pléiotropie antagoniste.....	17
I.1.1.g Nature des forces de sélection et leur impact dans la mise en place de l'adaptation locale .....	19
I.1.2 Bases génétiques de l'adaptation.....	20
I.1.2.a Nombre de locus impliqués.....	20
I.1.2.b.Nature des polymorphismes impliqués .....	21
I.1.2.c Standing variation <i>versus</i> mutation <i>de novo</i> .....	23
I.1.2.d Convergence adaptative .....	26
I.1.3 Méthodes moléculaires pour détecter l'adaptation .....	28
I.1.3.a Génotypage haut-débit de positions prédéfinies.....	28
I.1.3.b (Re-)séquençage haut-débit de génomes complets.....	33
I.1.3.c Détection des polymorphismes par re-séquençage NGS.....	36
I.1.3.d Séquençage en mélange .....	37
I.1.3.e Méthodes statistiques de détection de la sélection .....	38
I.1.3.f Etudes d'association avec le phénotype.....	44
I.1.3.g Etudes <i>Qst/Fst</i> .....	49
I.2 <i>ZEA MAYS</i> COMME MODELE BIOLOGIQUE .....	49
I.2.1 Système <i>Zea mays</i> .....	49
I.2.2 Génome de <i>Zea mays</i> .....	54
I.2.2.a Génome du maïs.....	54
I.2.2.b Outils de génotypage disponibles.....	56
I.2.2.c Identification de gènes ou régions d'intérêt chez le maïs.....	57
I.2.3 Adaptation locale chez les téosintes.....	60
I.3 OBJECTIFS DE LA THESE ET PRESENTATION DU TRAVAIL EFFECTUE.....	62
REFERENCES BIBLIOGRAPHIQUES .....	66
<b>II. CHAPITRE 1 : LOCAL ADAPTATION OF TEOSINTES ALONG ALTITUDINAL GRADIENTS USING WHOLE GENOME SEQUENCING OF POOLED SAMPLES .....</b>	<b>75</b>
II.1 ABSTRACT.....	77
II.2 INTRODUCTION.....	78
II.3 MATERIAL AND METHODS.....	81
II.3.1 Sampling information.....	81
II.3.2 Sequencing procedure.....	81
II.3.3 SNP detection, coverage and depth.....	82
II.3.4 Allele frequency estimation.....	83

II.3.5 Site-based neutrality tests. ....	83
II.3.5.a Differentiation-based methods. ....	83
II.3.5.b Diversity-based method. ....	84
II.3.5.c Correlation-based method. ....	85
II.3.6 Haplotype-based neutrality tests. ....	85
II.4 RESULTS. ....	86
II.4.1 Sampling 6 populations based on environmental analyses. ....	86
II.4.2 Sequencing statistics. ....	88
II.4.3 Allele frequencies and differentiation between populations. ....	89
II.4.4 Combining genome-scan statistics to identify candidate loci. ....	92
II.4.5 Validating candidate regions with haplotype analyses. ....	104
II.5 DISCUSSION. ....	106
ACKNOWLEDGEMENTS. ....	110
REFERENCES. ....	111
AUTHOR CONTRIBUTIONS. ....	115
SUPPORTING INFORMATION. ....	115
<b>III. CHAPITRE 2 : VALIDATION DES SNP CANDIDATS PAR L'ETUDE DE CLINES DE FRÉQUENCES</b> .....	<b>135</b>
III.1 INTRODUCTION. ....	137
III.2 OBTENTION DES 270 SNP CANDIDATS. ....	139
III.2.1 Cohérence des signaux de détection de la sélection. ....	139
III.2.2 Fst-scan. ....	140
III.2.3 Synthèse Bibliographique .....	145
III.2.4 Génotypage par la méthode KASPar .....	148
III.3 CLINES DE FREQUENCES .....	160
III.4 DISCUSSION .....	169
REFERENCES BIBLIOGRAPHIQUES .....	173
<b>IV. CHAPITRE 3 : RELIER LE POLYMORPHISME DES SNP CANDIDATS À DES VARIATIONS PHÉNOTYPIQUES</b> .....	<b>179</b>
IV.1 INTRODUCTION. ....	181
IV.2 DISPOSITIF DE GENETIQUE D'ASSOCIATION. ....	183
IV.2.1 Choix des caractères phénotypiques. ....	183
IV.2.2 Choix des populations .....	186
IV.2.3 Dispositif expérimental. ....	187
IV.3 ANALYSES DES DONNEES PHENOTYPIQUES .....	189
IV.3.1 Nettoyages des données phénotypiques .....	189
IV.3.2 Modélisation des effets environnementaux intra-essai. ....	190
IV.3.3 Analyses des données phénotypiques .....	192
IV.3.4 Effets populations et plantes-mères .....	195
IV.4 ACQUISITION ET ANALYSE DES DONNEES GENETIQUES. ....	199

IV.4.1 Prélèvements et extractions d'ADN .....	199
IV.4.2 Description des marqueurs microsatellites .....	199
IV.4.3 Génotypage des microsatellites .....	201
IV.4.4 Modélisation de la structuration neutre .....	205
IV.5 CONSTRUCTION DES MODELES DE GENETIQUE D'ASSOCIATION .....	207
IV.6 DISCUSSION.....	209
REFERENCES BIBLIOGRAPHIQUES .....	213
<b>V. DISCUSSION GENERALE.....</b>	<b>217</b>
V.1 LES TEOSINTES COMME MODELE D'ADAPTATION LOCALE. ....	219
V.2 REFLEXIONS SUR LES METHODES EMPLOYEES.....	220
V.3 L'ADAPTATION A L'ALTITUDE CHEZ LA TEOSINTE .....	224
V.4 PERSPECTIVES .....	227
REFERENCES BIBLIOGRAPHIQUES .....	229
<b>ANNEXES .....</b>	<b>231</b>
ANNEXE I.1 THE BRIGHT SIDE OF TRANSPOSONS IN CROP EVOLUTION.....	232
ANNEXE I.2 DESCRIPTION DES 37 POPULATIONS.....	253
ANNEXE III.1 SYNTHESE BIBLIOGRAPHIQUE.....	258
ANNEXE III.2 : TABLE DE FREQUENCES ALLELIQUES.....	295
ANNEXE IV.1 : PLANTES-MERES .....	316



# **I. INTRODUCTION**



L'environnement dans lequel se trouve une espèce présente souvent d'importantes variations spatiales et temporelles. Pour faire face à cette hétérogénéité environnementale, les espèces peuvent avoir recours à plusieurs stratégies : la migration – avec des conséquences sur l'aire de répartition de l'espèce –, l'adaptation locale ou la plasticité phénotypique. Alors que la migration implique un déplacement de l'espèce, les deux dernières stratégies impliquent une modification du phénotype sans changement génétique (plasticité) ou avec changement génétique (adaptation). Adaptation et plasticité phénotypique sont parfois difficiles à distinguer. Nous nous concentrerons dans cette introduction sur la comparaison de ces deux stratégies, avec un effort particulier sur l'adaptation locale et sa détection qui est le mécanisme auquel je me suis plus particulièrement intéressée au cours de ma thèse.

## **I.1 ETUDIER L'ADAPTATION LOCALE**

### **I.1.1 Adaptation locale**

#### **I.1.1.a Définition et mécanisme de mise en place de l'adaptation locale**

Les populations présentent de la diversité : les individus sont différents les uns des autres. Cette diversité phénotypique est l'expression de différences génétiques entre individus et/ou de variations dues à l'environnement. C'est sur cette variabilité que s'exerce la sélection naturelle, mise en évidence par Darwin dans l'Origine des espèces (1859). Dans des conditions données, certains phénotypes – ou valeur de certains caractères – assurent une plus grande valeur sélective définie par une meilleure chance de survie et/ou de reproduction. Les individus porteurs de ce phénotype donné vont donc majoritairement contribuer à la prochaine génération. Si le phénotype est déterminé en partie génétiquement, il est héréditaire. Lorsque les individus avantagés par un phénotype se reproduisent, seule une partie de la valeur génétique est transmise à leurs descendants. C'est sur cette valeur (héritabilité) que se base la sélection naturelle. Les allèles composant le génotype vont augmenter en fréquence dans la population et contribuer au mécanisme d'adaptation de la population à un environnement donné. L'adaptation à l'échelle d'une population se définit donc comme un changement de fréquence(s) allélique(s), qui entraîne une modification de la distribution d'un phénotype. Les individus porteurs des nouveaux phénotypes mieux adaptés voient leur valeur sélective augmenter par rapport aux individus porteurs du phénotype ancestral.

La notion d'environnement est importante lorsque l'on étudie l'adaptation. Les contraintes sélectives qui s'appliquent à une espèce sont en effet hétérogènes dans l'espace ou



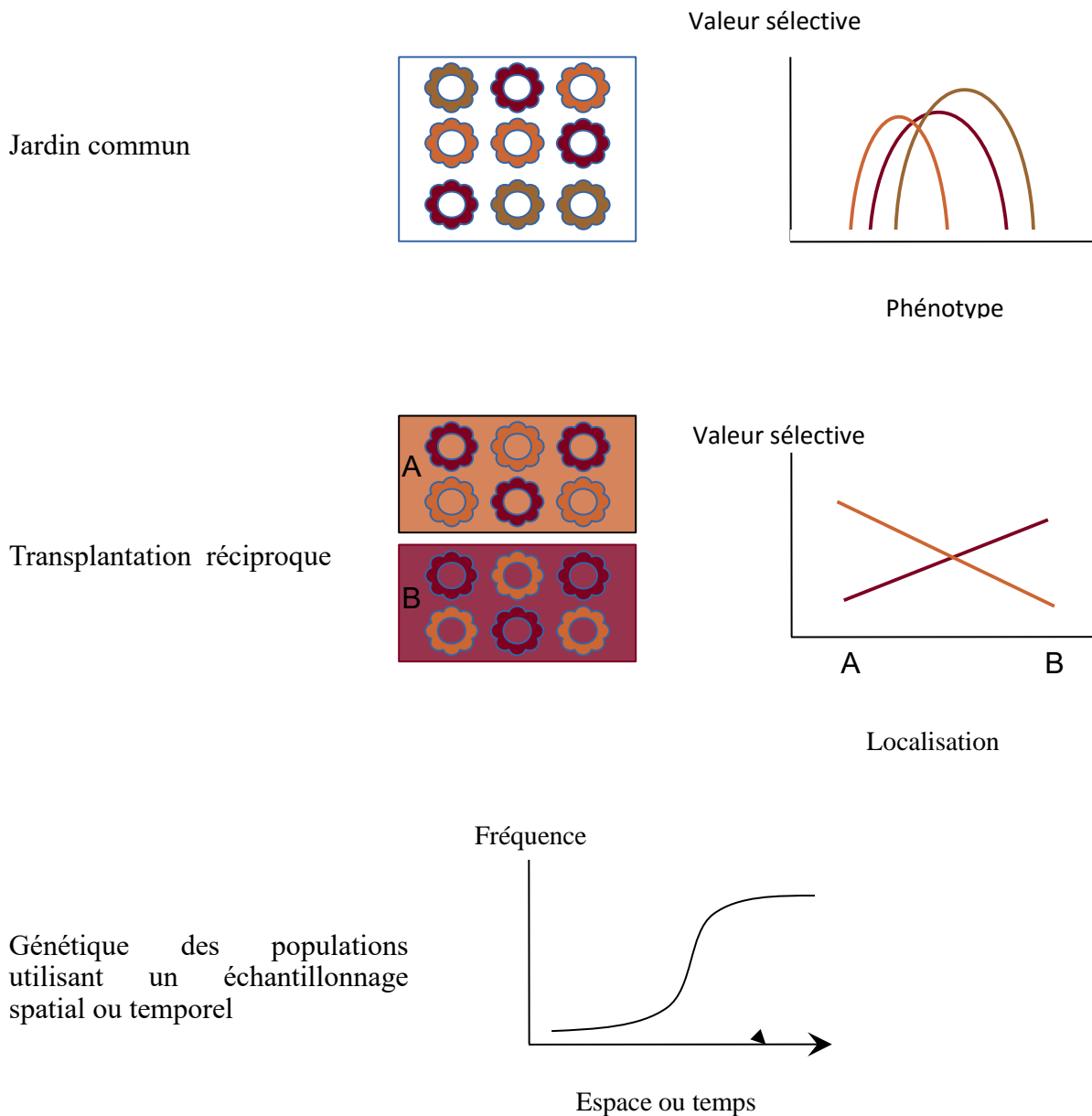
dans le temps. La réponse d'une population particulière aux contraintes qu'elle rencontre est dénommée *adaptation locale*. Bien que principalement mise en avant dans le rôle de l'adaptation, la sélection naturelle n'est pas la seule force évolutive influençant l'adaptation locale. La mutation par exemple est à l'origine de la création de nouveaux variants génétiques. La recombinaison, lors de la méiose, mélange les allèles provenant des deux chromosomes homologues. C'est sur le phénotype que va produire ces combinaisons d'allèles, potentiellement avantageuses, neutres ou délétères, que vont s'exercer conjointement la sélection naturelle et la dérive génétique. La dérive génétique est une force évolutive qui peut conduire à la perte ou à la fixation d'allèles, qu'ils soient adaptatifs ou mal-adaptatifs, dans les populations suite à des processus stochastiques. La migration peut aussi contribuer à l'adaptation locale en apportant de nouveaux variants avantageux dans la population receveuse. Cependant, elle peut également empêcher sa mise en place en apportant un flux constant de nouveaux variants mal-adaptés qui empêche la fixation d'allèles avantageux.

L'adaptation locale est donc la résultante de forces évolutives conduisant à un phénotype performant dans certaines conditions environnementales, dites « natives », de la population. Ainsi, les individus d'une population présenteront un meilleur phénotype dans leur environnement d'origine relativement aux individus d'autres populations non-natives : ils posséderont localement une meilleure valeur sélective.

### **I.1.1.b Mise en évidence de l'adaptation locale**

Une différenciation génétique et phénotypique entre populations échantillonnées dans des environnements différents ou à des temps différents peut indiquer l'existence d'adaptation locale. Les études spatiales permettent d'observer un grand nombre de populations dans des environnements différents à un temps donné, alors que les études temporelles permettent d'accéder directement aux réponses évolutives (Figure 1 ; Franks *et al.* 2014). Les différenciations génétique et phénotypique entre populations peuvent être mises en évidence par des études de génétique des populations (voir section I.1.3.e). Cependant, pour vraiment démontrer l'existence d'adaptation locale il faut avoir recours à des stratégies expérimentales. Une première stratégie peut être l'étude de jardins communs où plusieurs populations prélevées dans des environnements différents sont cultivées dans un (des) environnement(s) commun(s) (Figure 1). Leur valeur sélective est alors mesurée et corrélée avec des variables environnementales d'origine. Dans le cas des jardins communs, il est parfois possible de

contrôler pour les effets maternels causés par l'environnement de la mère (Wolf & Wade 2009). Ces effets peuvent venir perturber les estimations des effets génétiques en créant de fausses associations entre le phénotype et le génotype des individus.



**Figure 1 : Dispositifs pour étudier l'adaptation locale.**

Une autre stratégie pour étudier l'adaptation locale est la mise en œuvre de transplantations réciproques entre populations (Merilä & Hendry 2014). Dans ce cas, la valeur sélective de la population native dans son environnement est comparée directement avec celle

de la population introduite, et réciproquement (Figure 1). Une dernière stratégie est d'effectuer des expériences d'évolution expérimentale au cours desquelles des populations évoluent sous une pression de sélection contrôlée – température/concentration de CO<sub>2</sub>/disponibilité en eau. Les valeurs sélectives des populations sont comparées entre elles et aux populations ancestrales dont elles dérivent, ce qui permet de tester les potentialités d'adaptation locale. Dans cette dernière démarche, ce n'est pas l'adaptation locale réalisée mais plutôt le potentiel adaptatif de la population qui est étudié.

Ces trois stratégies passent par des mesures phénotypiques comme proxy de la valeur sélective puis par des comparaisons des performances des populations ou des individus évaluées dans ces dispositifs. Dans les expériences en jardins communs, la valeur sélective des populations est étudiée dans un même lieu afin de minimiser les effets de l'environnement. Ainsi, chez *Arabidopsis thaliana* plusieurs écotypes ont été plantés dans quatre sites couvrant l'aire de répartition naturelle de l'espèce. Cette approche a permis d'identifier des locus liés à des variations de valeur sélective entre écotypes. Les allèles liés à de meilleures performances étaient surreprésentés dans les écotypes provenant de populations naturelles poussant dans des environnements similaires à ceux des sites de jardins communs où ils confèrent un avantage (Fournier-Level *et al.* 2011). Ces dispositifs sont plus difficiles à mener sur les modèles animaux pour des raisons logistiques. On peut néanmoins suivre les populations dans leur environnement naturel. Le rongeur *Peromyscus polionotus* présente par exemple une gamme de coloris de robe qui est corrélée avec la coloration des sols. Mullen & Hoekstra (2008) ont mesuré les réflectances du pelage des individus par rapport à la coloration du substrat d'origine, qu'ils ont ensuite utilisées comme proxy de la valeur sélective dans un contexte de prédation par des oiseaux. Ils ont ainsi pu démontrer que la couleur du pelage résultait d'un processus d'adaptation locale.

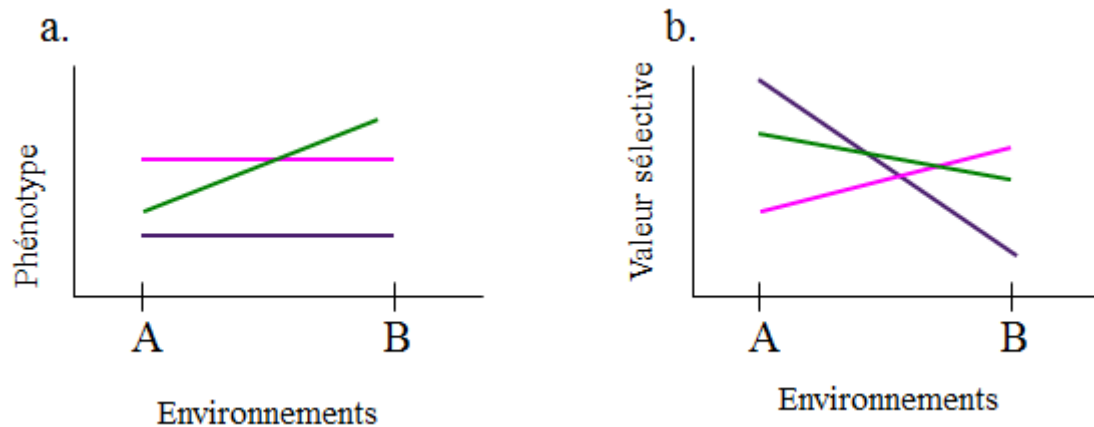
Dans les expériences de transplantation réciproque, l'adaptation locale est avérée lorsque les performances en terme de valeur sélective d'une population dans son environnement d'origine sont meilleures que celles de n'importe quelle autre population transplantée dans ce même environnement (Figure 1) (Kawecki & Ebert 2004). Par exemple, Ågren & Schemske (2012) ont réalisé ce type de dispositif sur des populations d'*Arabidopsis thaliana* provenant d'Italie et de Suède sur cinq années à différents stades de développement de la plante. Les plantes locales présentent de meilleurs taux de survie et/ou fécondité (production de graines) que les plantes transplantées, indiquant une forte adaptation locale liée notamment à l'adaptation aux sols froids pour les populations suédoises. Anderson *et al.*

(2013) ont créé des lignées de l'espèce autogame *Boechnera stricta* en croisant deux individus provenant de deux régions différentes des Etats-Unis, le Colorado et le Montana. Ils ont fait pousser dans ces deux environnements les lignées natives et les descendants hybrides. Ils ont ensuite comparé les probabilités de floraison en fonction des allèles « natifs » et « extérieurs » sur 164 marqueurs. Sur ces marqueurs, près de 11% étaient associés à une meilleure probabilité de floraison dans au moins un site lorsque les allèles provenant de la population locale étaient portés par les individus étudiés, reflétant l'adaptation locale.

Les expériences en évolution contrôlée sont un bon moyen de tester l'effet d'une variable sélective sur un organisme. Leurs contraintes logistiques les limitent généralement aux organismes à cycle de vie court, et ne permettent pas souvent de se placer dans des conditions naturelles véritables (Merilä & Hendry 2014). Elles n'en restent pas moins très intéressantes. Lohbeck *et al.* (2012) ont permis de mettre en évidence le potentiel adaptatif d'une espèce de phytoplancton en faisant évoluer sur 500 générations des populations dans les concentrations de CO<sub>2</sub> prévues pour les prochaines années. D'autres études ont porté sur des organismes à cycle de vie plus long, comme l'expérience de sélection divergente récurrente pour la précocité de floraison conduite sur des lignées de maïs dans des conditions agronomiques par Durand *et al.* (2015). Les 10 individus les plus tardifs et les 10 individus les plus précoces ont été prélevés pour constituer la génération suivante. Après 16 générations, il y a encore un potentiel adaptatif pour une floraison précoce, tandis que le potentiel pour une floraison tardive s'est épuisé pour certaines populations au bout de sept générations.

### **I.1.1.c Plasticité phénotypique**

La plasticité phénotypique est la gamme de phénotypes que peut donner un génotype dans des environnements distincts (Figure 2). Elle est liée à l'expression de différences développementales dans les environnements rencontrés. Par exemple, chez la plante modèle *Arabidopsis thaliana*, la floraison dépend de la répression du gène *FLC*. Chez les écotypes sensibles à la vernalisation, l'exposition au froid dans les stades précoces de développement entraîne la répression du gène *FLC* au niveau transcriptionnel. Pour un même génotype, cette répression est d'autant plus forte que l'intensité et la durée du froid sont importantes (Sheldon *et al.* 2000). Il s'agit donc d'un exemple de réponse plastique liée à l'environnement.



**Figure 2 : Adaptation locale (génotype rose et violet) versus plasticité phénotypique (génotype vert).** (a) Phénotype et (b) valeur sélective des différents génotypes dans les environnements A et B. Le génotype violet est adapté à l'environnement A tandis que le génotype rose est adapté à l'environnement B. Ces deux génotypes ne sont pas plastiques, c'est-à-dire qu'ils présentent le même phénotype quel que soit l'environnement, contrairement au génotype vert qui est plastique, c'est-à-dire qu'il présente un phénotype variable. De plus sa plasticité est adaptative, c'est à dire que son phénotype est proche du meilleur phénotype dans chaque environnement.

Longtemps définie comme un « bruit » autour du génotype, la plasticité est aujourd'hui considérée comme pouvant être héritable et d'une importance non négligeable pour l'évolution des espèces (Lande 2009 ; Pigliucci 2005). Pendant longtemps les gènes de la plasticité phénotypique ont été recherchés. Deux écoles se sont affrontées : celle des gènes « sensitifs » et celle des gènes « régulateurs ». Dans le cas de gènes « sensitifs », l'idée est que l'expression des gènes codant directement le phénotype est modulée par l'environnement. Dans le cas des gènes « régulateurs », c'est l'expression de gènes modulant l'expression d'autres gènes impliqués dans les différences phénotypiques qui est sensible aux stimuli environnementaux (revue Via *et al.* 1995). Les deux catégories de gènes ont été identifiées par des études sur les gènes de la plasticité, chaque réponse plastique pouvant être déterminée par des mécanismes plus ou moins complexes, comme c'est le cas pour n'importe quel caractère phénotypique (Pigliucci 2005).

La plasticité phénotypique est fondamentalement différente de l'adaptation locale. Dans le cadre de l'adaptation locale, la population présente une meilleure valeur sélective dans son environnement d'origine. Dans le cadre de la plasticité phénotypique, un phénotype différent est produit selon les environnements pour optimiser la valeur sélective. Notons cependant que ce phénotype plastique confère souvent un avantage sélectif moins important que celui résultant de l'adaptation locale, car le phénotype peut être contraint et/ou plus coûteux (DeWitt *et al.* 1998). Un exemple remarquable de ce phénomène peut être trouvé

chez la grenouille *Rana temporaria*, qui peut accélérer son développement en fonction des risques de sécheresse. Les populations de grenouilles les plus plastiques pour ce caractère présentent des tailles à la métamorphose plus petites lorsque les conditions de sécheresse ne sont pas réunies (Merilä *et al.* 2004). Cet exemple illustre bien le compromis évolutif entre plasticité et adaptation. Mais adaptation et plasticité ne sont pas mutuellement exclusives : certains caractères peuvent répondre à une évolution adaptative, d'autres présenter de la plasticité, et d'autres encore une combinaison des deux. Anderson *et al.* (2012) ont mené une expérience sur des populations de *Boechera stricta* qu'ils ont suivies pendant 38 ans dans un contexte de changement climatique. La précocité de floraison était mesurée ainsi que des variables environnementales. Les auteurs ont montré que la précocité de floraison était liée aux variables environnementales (plasticité phénotypique) mais aussi qu'il y avait de la sélection pour des génotypes précoces (adaptation). Enfin, les deux peuvent interagir (Franks *et al.* 2013). Ainsi, la plasticité peut ralentir l'adaptation ou au contraire la promouvoir en permettant par exemple la subsistance des populations dans des conditions environnementales fluctuantes. Par exemple, dans le cas d'une population île recevant un flux continu d'allèles mal-adaptés d'un continent, l'augmentation de la plasticité phénotypique dans la population île est sélectionnée. Cela permet d'obtenir une gamme phénotypique plus importante mais dont la moyenne se rapproche de l'optimal phénotypique (Chevin & Lande 2011). Lind *et al.* (2011) ont mesuré la plasticité phénotypique de différentes populations de grenouilles *Rana temporaria* dans différents environnements. Ils ont mis en évidence une corrélation positive entre la plasticité phénotypique et le taux de migration lorsque les conditions entre les populations diffèrent.

#### **I.1.1.d Adaptation locale versus plasticité phénotypique**

Distinguer adaptation locale et plasticité phénotypique n'est pas chose aisée. Alors même que le mécanisme génétique de l'adaptation locale est découvert, il est difficile d'éliminer totalement une participation de la plasticité. Inversement, peu de dispositifs permettent de trouver le ou les gène(s) responsable(s) de la variation phénotypique (Franks *et al.* 2014 ; Merilä & Hendry 2014). Les expérimentations en jardins communs peuvent trancher pour l'un ou l'autre des phénomènes si le pedigree est connu ou les individus apparentés, car dans ce cas la variation phénotypique observée peut être liée à un génotype particulier (Figure 1, Figure 2). La réplication des expériences dans différents environnements permet de dissocier effets génétiques et effets environnementaux. Les études temporelles permettent également d'évaluer la part de la plasticité et celle de l'adaptation (Figure 1). Par exemple, dans leur

suivi des populations de *Boechnera stricta* sur 38 ans, Anderson *et al.* (2012) ont attribué la part de variation phénotypique due à la réponse de l'adaptation d'une année sur l'autre, et la part de variation due à la plasticité phénotypique due à la variation environnementale (température, date de la première chute de neige) indépendamment des années. Dans le cas de transplantations réciproques, Wang *et al.* (2010) ont proposé une méthode pour distinguer les effets environnementaux de l'adaptation. Ils ont appliqué cette méthode sur un dispositif comportant 140 populations de pins tordus *Pinus contorta* sur 62 sites différents. Les auteurs ont conclu là aussi à l'action conjointe de l'adaptation locale et de la plasticité phénotypique. De tels dispositifs sont cependant lourds et parfois impossibles à mettre en place techniquement.

La mise en place de l'une ou de l'autre stratégie dépend de la variabilité de l'environnement et des forces sélectives et démographiques auxquelles sont soumises les populations. Les conditions dans lesquelles s'établissent l'une ou l'autre des stratégies sont assez discutées (Franks *et al.* 2014). Une variabilité temporelle fine de l'environnement, c'est à dire intra-générationnelle, favorise la plasticité phénotypique. La plasticité phénotypique est aussi favorisée par une variabilité spatiale fine, surtout si la population ne peut se soustraire aux conditions environnementales comme dans le cas des plantes (Sultan & Spencer 2002 ; Van Kleunen & Fischer 2005). En cas de migration importante entre populations provenant d'environnements hétérogènes, la plasticité phénotypique est également favorisée par rapport à l'adaptation locale (Sultan & Spencer 2002). En effet, les flux de gènes entretenus par la migration s'opposent à l'adaptation locale en introduisant des allèles provenant de populations adaptées à d'autres conditions environnementales (Chevin & Lande 2011 ; Slatkin 1987).

Une revue bibliographique sur la plasticité phénotypique chez les plantes montre à quel point ce phénomène est courant comme réponse à la variabilité environnementale pour des caractères aussi divers que l'architecture de la plante, le mode de reproduction, ou la production de mécanismes de défense (Franks *et al.* 2014). Par exemple, Sultan *et al.* (2013) ont montré grâce à une étude temporelle que la colonisation d'un environnement à ciel ouvert peut s'accompagner d'une augmentation de la plasticité phénotypique sur l'allocation de ressources entre racines et organes photosynthétiques, permettant de réduire l'évapotranspiration.

Au contraire, une contrainte sélective stable et une migration restreinte favorisent l'adaptation locale. Toujours chez les plantes, une méta-analyse de 35 études portant sur la mise en évidence d'adaptation locale montre que 45 à 71% des paires de populations comparées (native et non-native) présentent de l'adaptation locale (Leimu & Fischer 2008).

Comme attendu, cette adaptation dépend des différences entre les environnements mais aussi de la taille des populations, l'adaptation locale étant trouvée plus fréquemment dans les grandes populations (Leimu & Fischer 2008). Ainsi, une sélection forte présente dans un habitat hétérogène et une dérive limitée favorisent l'installation de l'adaptation locale. Les flux de gènes et la dispersion des individus sont alors contre-sélectionnés (Yeaman & Otto 2011).

#### **I.1.1.e Adaptation locale, un processus dynamique dans le temps et l'espace**

L'adaptation locale est un mécanisme qui peut se faire très rapidement à l'échelle de quelques générations. Le cas de la phalène du boulot *Biston betularia*, un papillon de nuit qui le jour adopte une stratégie de mimétisme pour échapper à la prédation des oiseaux, est un des exemples les plus célèbres d'adaptation rapide. A l'origine très rare dans les populations naturelles, le phénotype mélanique a fortement augmenté en fréquence en quelques dizaines d'années dans les populations situées dans les zones très industrialisées du nord de l'Angleterre, alors que le phénotype clair est resté en fréquence majoritaire dans les zones non industrielles. En effet, la pollution se dépose sur les troncs des arbres et les assombrit. Le phénotype mélanique constitue alors un dispositif de camouflage plus efficace pour contrer la pression de prédation (Bishop & Cook 1980). Il est intéressant de noter qu'avec le relâchement récent de la pression de sélection dû à une baisse de la pollution atmosphérique, un phénotype intermédiaire est en train de suppléer le phénotype mélanique (Cook & Grant 2000).

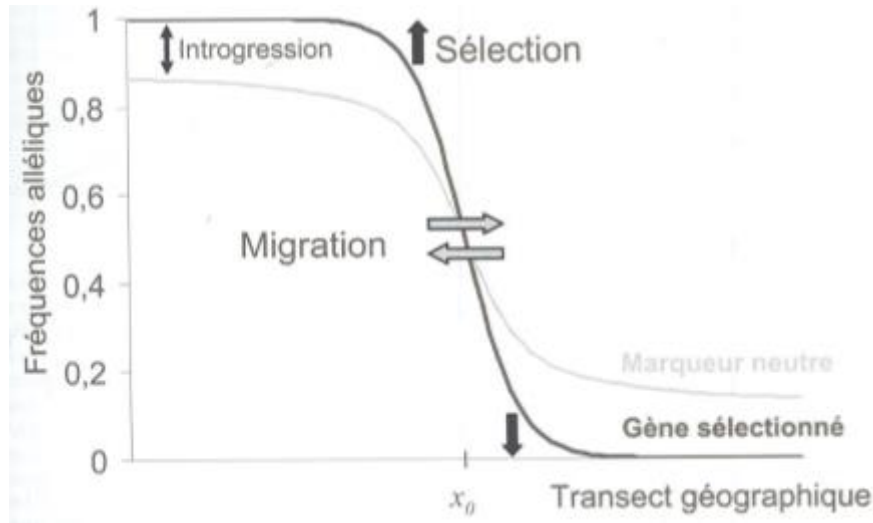
L'adaptation locale peut avoir lieu directement par l'installation d'un allèle persistant ou se faire au contraire en plusieurs étapes par l'acquisition d'une succession d'allèles adaptatifs. Lors d'une étude temporelle dans la région montpelliéraine, Labbé *et al.* (2009) ont ainsi montré de la résistance aux insecticides chez *Culex pipiens* : les premiers allèles sélectionnés étaient généralistes, puis l'adaptation locale a eu lieu via l'émergence d'allèles spécialistes qui présentaient un fort avantage sélectif dans les zones traitées, mais causant des performances moindres dans les zones non-traitées. La composition phénolique du thym *Thymus vulgaris* est un compromis évolutif entre résistance au froid (morphes non phénoliques) et résistance à la sécheresse (morphes phénoliques). Thompson *et al.* (2013) ont suivi pendant 40 ans l'évolution de populations de thym dans la région méditerranéenne. Ils ont pu mettre en évidence qu'avec le changement climatique, les morphes résistants au froid



disparaissaient au profit de ceux résistants à la sécheresse.

Les facteurs sélectifs conduisant à la mise en place d'adaptation locale peuvent concerner une échelle géographique large, comme par exemple les grands types de climat. La drosophile *Drosophila melanogaster*, à l'origine une espèce africaine, est récemment devenue cosmopolite. Plusieurs études comparant des populations tropicales aux populations tempérées ont mis en évidence une adaptation au climat (Gonzalez *et al.* 2010 ; Hoffmann & Weeks 2007). Gonzalez *et al.* (2010) ont notamment souligné l'importance des éléments transposables dans ces adaptations. Les auteurs ont identifié huit insertions d'éléments transposables impliquées dans l'adaptation au climat tempéré, dont un LTR, *Invader4*, qui est inséré dans le premier intron du gène *sra*. L'adaptation locale peut également se faire à une échelle beaucoup plus fine. C'est par exemple le cas lors d'une infection par le VIH. Les résistances aux rétrovirus émergent indépendamment chez les différents patients. Certaines substances actives des traitements contre le VIH peuvent ne pas atteindre tous les organes qui constituent alors des réserves où les virus peuvent se multiplier et les résistances peuvent émerger (Moreno-Gamez *et al.* 2015).

Dans les zones de transition entre présence ou non d'un facteur sélectif, il existe un continuum des fréquences alléliques, appelé cline de fréquence, généré par un équilibre entre sélection et migration (Figure 3). L'étude de ces cliniques permet d'une part de détecter des gènes et/ou des phénotypes impliqués dans l'adaptation locale, et d'autre part de mieux appréhender la force de la sélection et les flux de gènes entre populations. Un exemple provient du rongeur *Peromyscus polionotus* qui présente différentes colorations de pelage dues en grande partie au gène *Mclr*. L'analyse conjointe de ce gène avec des marqueurs neutres montre d'une part qu'il y a des groupes génétiques distincts chez ce rongeur, et d'autre part que le gène *Mclr* ne ségrège pas selon les groupes génétiques, mais qu'il est au contraire directement corrélé à la coloration du substrat de la population d'origine (Mullen *et al.* 2009).

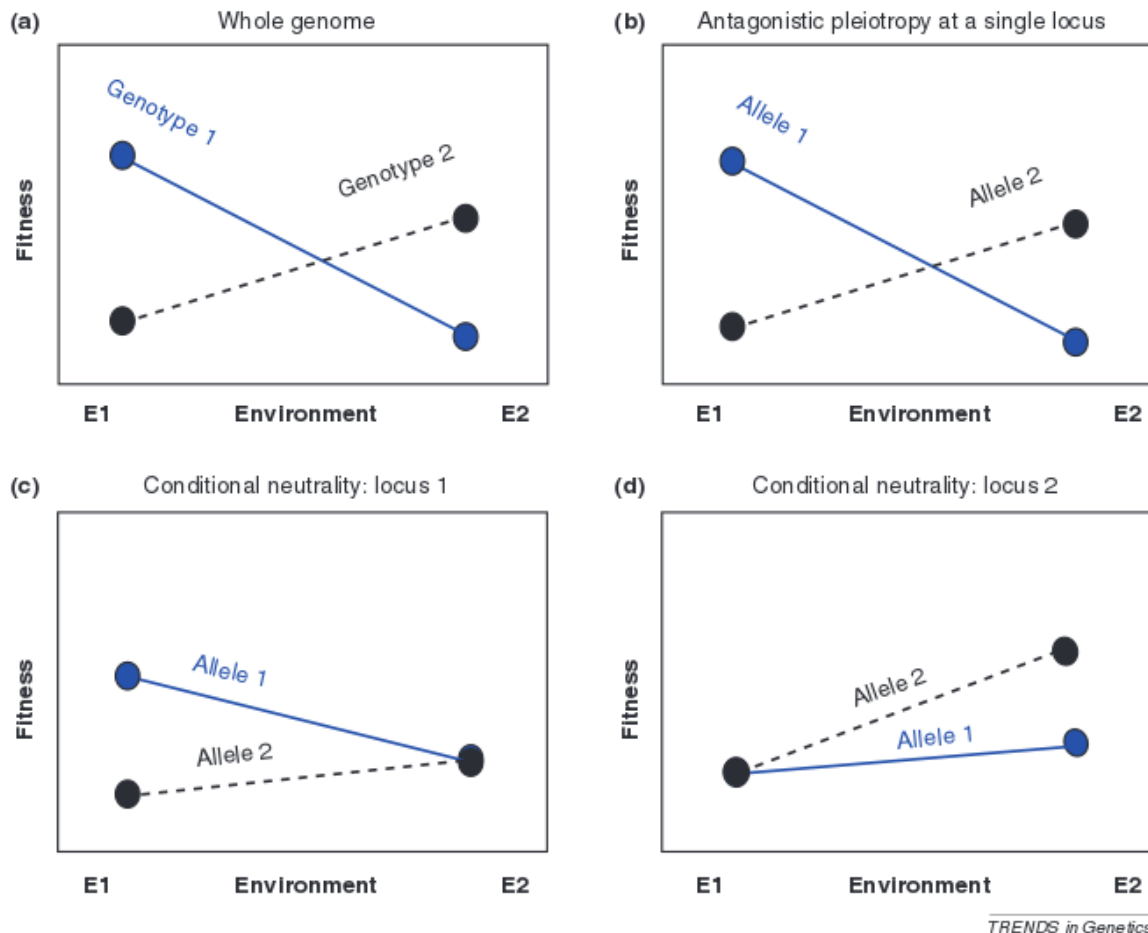


**Figure 3 : Cline de Fréquence.** La fréquence allélique le long du transect géographique varie en fonction de l'intensité de la sélection et de la migration. Les locus sous sélection présente un fort cline de fréquence, tandis que les allèles aux locus neutres diffusent de part et d'autre du cline. Figure de Ravigné *et al.* (2010).

#### **I.1.1.f Deux modèles d'adaptation locale : neutralité conditionnelle et pléiotropie antagoniste**

Comme indiqué dans l'exemple précédent, un allèle qui confère un avantage sélectif dans des conditions données peut avoir un coût dans d'autres conditions. Les moustiques porteurs de certains allèles conférant une résistance aux insecticides produisent un excès d'estérases qui piègent les insecticides organophosphorés. Ces derniers s'accumulent alors dans le corps des moustiques, ralentissant leurs fonctions métaboliques (Labbé *et al.* 2009 ; Weill *et al.* 2000).

Deux modèles d'adaptation locale sont classiquement définis (revue Anderson *et al.* 2011). Un modèle de pléiotropie antagoniste stipule que pour un locus codant un caractère adaptatif, des allèles différents sont sélectionnés dans chaque habitat conduisant, à l'échelle de l'espèce, à un maintien de polymorphisme sous l'action d'une sélection disruptive. Ces allèles procurent un avantage sélectif dans un habitat donné et un désavantage dans un autre (Figure 4-b).



**Figure 4 : Adaptation locale.** (a) Le génotype 1 présente une meilleure valeur sélective dans l'environnement 1, tandis que le génotype 2 présente une meilleure valeur sélective dans l'environnement 2. (b) L'adaptation locale de ces deux génotypes peut se faire par le biais de pléiotropie antagoniste où l'allèle local confère une meilleure valeur sélective que l'allèle étranger, et/ou (c et d) par le cas de neutralité conditionnelle à deux locus où à chaque locus un allèle confère une meilleure valeur sélective dans un environnement mais pas dans un autre. Figure provenant de Anderson *et al.* (2011).

Un deuxième modèle dit de neutralité conditionnelle propose qu'à un locus donné codant un caractère adaptatif, un allèle procure un avantage sélectif dans un habitat par rapport à l'autre allèle, tandis que dans un autre habitat, les individus portant l'un ou l'autre des allèles présentent des performances équivalentes (Figure 4-c, d). En d'autres termes, l'avantage sélectif d'un allèle ne s'exprime que dans un habitat. Sous ce modèle, le polymorphisme à l'échelle de l'espèce n'est pas nécessairement maintenu puisque l'allèle avantageux dans un habitat et neutre dans les autres peut éventuellement être fixé (revue Anderson *et al.* 2011).

Dans la littérature, les cas d'adaptation par neutralité conditionnelle sont plus fréquents (Colautti *et al.* 2012). Cependant, certains cas de pléiotropie antagoniste sont connus, notamment l'adaptation aux métaux lourds chez les plantes (Hörger *et al.* 2013). Certaines

plantes ont la faculté de pousser dans des environnements riches en métaux lourds, en adoptant une stratégie d'hyper-accumulation de ces métaux. Cette stratégie se retrouve fréquemment chez les Brassicacées. *Arabidopsis halleri* pousse dans des environnements riches en zinc et cadmium grâce à une augmentation de la cis-régulation de *HMA4* couplée à une augmentation du nombre de copies de ce gène impliqué dans l'évacuation de ces métaux hors de la cellule (Hanikenne *et al.* 2013 ; Hanikenne *et al.* 2008). Cependant, en l'absence de métaux lourds, cette bio-accumulatrice a des performances moindres en comparaison de l'espèce sœur *Arabidopsis thaliana* sensible aux métaux lourds. Plusieurs hypothèses peuvent expliquer ce phénomène. Les bio-accumulatrices présentent généralement un développement moindre de défenses immunitaires car l'accumulation de métaux lourds dans l'organisme aide à lutter contre les parasites. Les plantes tolérantes possèdent aussi un métabolisme plus lent (Hörger *et al.* 2013). Chez *Mimulus guttatus*, certaines populations originaires de mines de cuivres sont adaptées à de fortes concentrations en cuivre (Allen & Sheppard 1971). Ces populations montrent non seulement des performances moindres dans un milieu pauvre en cuivre par rapport aux génotypes sensibles (Allen & Sheppard 1971), mais aussi des degrés dans la tolérance au cuivre selon la distance aux mines, ce qui suggère un coût différent selon le degré de la tolérance (MacNair *et al.* 1993).

Dans une méta-analyse sur l'adaptation locale d'un grand nombre de taxons (plantes et animaux), Hereford (2009) a détecté un effet assez faible de pléiotropie antagoniste. Cependant dans le modèle de pléiotropie antagoniste, l'avantage sélectif des deux allèles dans leur environnement d'origine doit être significatif, ce qui demande une puissance plus élevée pour la détection statistique. Les cas d'adaptation par neutralité conditionnelle peuvent être aussi plus fréquents dans les expériences en laboratoire ou en environnement artificiel, qui ne reflèteraient pas correctement les conditions des milieux dans lesquels l'adaptation par pléiotropie antagoniste intervient. Cette plus grande difficulté à détecter les cas de pléiotropie antagoniste pourrait donc en partie expliquer le biais observé (Anderson *et al.* 2013).

#### **I.1.1.g Nature des forces de sélection et leur impact dans la mise en place de l'adaptation locale**

Connaître les pressions de sélection qui s'exercent sur les populations et savoir les expliciter facilite l'étude de l'adaptation locale (Kawecki & Ebert 2004). Les facteurs de sélection peuvent être d'origine biotique : la pression parasitaire (Daub *et al.* 2013), le parasitisme (Eizaguirre & Lenz 2010) ou encore la prédation (Åbjörnsson *et al.* 2004). Elles peuvent aussi

être de nature abiotique, comme par exemple la nature des sols, le pH ou la salinité du milieu. Le changement de latitude tout comme d'altitude s'accompagne généralement de grandes modifications du climat. La montée en altitude s'accompagne d'une baisse de température et d'une baisse du niveau d'oxygène, mais peut aussi être liée à une variation de la pluviométrie ou de la saisonnalité. Des réponses à la sélection sur des organismes divers ont été reportées pour un grand nombre de ces variables : adaptation à l'eau douce chez *Gasterosteus aculeatus* (Hohenlohe *et al.* 2010) ; au niveau d'oxygène chez l'homme (Yi *et al.* 2010) ; au climat chez la drosophile (Gonzalez *et al.* 2010 ; Hoffmann & Weeks 2007) ou chez *Arabidopsis thaliana* (Hancock *et al.* 2011 ; Long *et al.* 2013).

Une méta-analyse sur la détection de *Quantitative Trait Loci* (QTL) associés à l'adaptation à des facteurs biotiques chez les plantes montre que ces QTL ont des effets plus importants que les QTL associés à des facteurs abiotiques (Louthan & Kay 2011). Les auteurs ont avancé l'hypothèse que les facteurs biotiques ont une temporalité beaucoup plus importante en comparaison des facteurs abiotiques, ce qui induit un changement quasi-permanent de l'optimum phénotypique. En accord avec le modèle géométrique de Fischer qui prédit une diminution graduelle de l'effet des mutations fixées avant d'atteindre l'optimum phénotypique (Fisher 1930), l'adaptation en réponse à des facteurs biotiques aurait lieu par des mutations de larges effets (Fisher 1930 ; Louthan & Kay 2011).

## **I.1.2 Bases génétiques de l'adaptation**

### **I.1.2.a Nombre de locus impliqués**

Les traits phénotypiques peuvent être classés en deux catégories : les caractères qualitatifs et les caractères quantitatifs. Un caractère qualitatif présente des variants clairement distincts que l'on peut classer, comme par exemple les groupes sanguins ABO chez l'homme. Généralement, le déterminisme génétique de ces caractères peut être expliqué par un gène unique ou par peu de gènes qui se transmettent de façon mendélienne. Le phénotype « ridé » des pois de Mendel provient par exemple d'une insertion d'un élément transposable dans le gène *SBEI* qui modifie les processus métaboliques lors du développement de la graine (Bhattacharyya *et al.* 1990). L'allèle *r* est récessif : seuls les individus *rr* expriment le phénotype ridé, tandis que les individus *RR* ou *Rr* présentent un phénotype lisse.

Un caractère quantitatif présente au contraire une variation continue. Les caractères quantitatifs peuvent être codés par un nombre très variable de gènes ou locus, appelés QTL. Si

le phénotype est déterminé par un nombre important de gènes à effets faibles, on parle du modèle « infinitésimal » (Fisher 1930). Comme exemple de caractères quantitatifs on peut citer la date de floraison chez les plantes ou la réponse immunitaire chez l'homme. Ainsi, de nombreuses études chez le maïs *Zea mays* montrent que la date de floraison est déterminée par un grand nombre de locus (>80), et plus de 25 gènes ont été clairement identifiés (Hung *et al.* 2012 ; Weber *et al.* 2009). Ces gènes ont des effets variables. Le gène *ZmCCT* est celui qui affecte le plus la réponse à la photopériode, qui est une étape clé de la transition florale (Hung *et al.* 2012). D'autre part, une étude portant sur plus de 1000 réseaux de gènes chez 51 populations humaines a mis en avant l'importance de la sélection agissant sur les réseaux liés à l'immunité. Les gènes de ces réseaux présentent du déséquilibre de liaison (DL) de grande échelle, c'est-à-dire une association statistique entre des allèles de différents gènes qui ne sont pas liés physiquement, ce qui laisse supposer qu'il y a des relations épistasiques entre eux et que la sélection a agi sur ces gènes simultanément (Daub *et al.* 2013).

### **I.1.2.b.Nature des polymorphismes impliqués**

Les polymorphismes peuvent survenir dans des régions diverses du génome (régions codantes, introns, séquences régulatrices...) et être de nature différente, allant d'un changement d'une base de l'ADN à un réarrangement chromosomique.

Les polymorphismes ponctuels (*Single Nucleotide Polymorphism* ou SNP) concernent la substitution, l'insertion ou la délétion d'un seul nucléotide. Dans certaines populations humaines africaines pratiquant le pastoralisme et dont les produits laitiers sont une part importante de l'alimentation, la persistance de la tolérance au lactose à l'âge adulte, contrairement aux autres populations ne consommant pas ou peu de produits laitiers, est permise par la substitution d'une guanine par une cystéine en amont du gène *LCT* régulant ainsi son expression. (Tishkoff *et al.* 2007).

Les duplications de gènes peuvent aussi être des moteurs de l'adaptation : les contraintes de sélection peuvent ainsi s'amointrir sur une des copies des gènes, ce qui favorise l'émergence de nouvelles fonctions (Lynch & Conery 2000) malgré certaines contraintes (Lynch & Conery 2000 ; Tsai *et al.* 2012). Un exemple d'adaptation à un régime alimentaire végétarien ayant conduit à la spéciation est particulièrement frappant chez les Colobinés. Ceux-ci sont les seuls primates qui se nourrissent exclusivement de feuilles, ce qui entraîne une baisse de pH dans leur système digestif. Ce changement alimentaire s'accompagne d'une duplication de l'enzyme *RNase1* impliqué dans la digestion et la défense

immunitaire contre les virus. La copie *RNase1B* montre des traces de sélection positive, et possède une forte activité catalytique dans les environnements acides mais a perdu l'activité catalytique impliquée dans la digestion enzymatique des virus (Yu *et al.* 2010 ; Zhang *et al.* 2002). Les duplications peuvent également être impliquées dans l'adaptation locale, comme les duplications du gène *Ace1* responsables d'une résistance aux insecticides chez *Culex pipiens* (Lenormand *et al.* 1998).

Les variants du nombre de copies (*Copie Number Variant* ou CNV) sont une forme de polymorphisme qui, comme leur nom l'indique, concerne un nombre variable de copies d'un gène. Il existe chez la drosophile des séries alléliques pour le gène de résistance aux insecticides comportant du DDT, *Cyp6g1*. Les différentes formes de ce gène peuvent correspondre aussi bien à une unique version qu'à deux versions complètes du gène associées à de nombreux fragments de ce dernier. Les allèles porteurs d'un plus grand nombre de copies confèrent une plus grande résistance aux insecticides (Schmidt *et al.* 2010). Il s'agit donc ici d'une résistance quantitative qui varie en fonction du nombre de copies.

Les polymorphismes de présence/absence d'éléments transposables jouent aussi un rôle important dans l'adaptation locale. Les éléments transposables ont joué un rôle dans l'amélioration des espèces cultivées (revue de l'impact des éléments transposables dans l'évolution des génomes des espèces domestiques ; Vitte *et al.* 2014, Annexe I.1). Chez le maïs, l'insertion du transposon *Ds* dans un exon du locus *Wx* modifie l'activité enzymatique du gène et conduit à une modification de la composition de l'amidon des grains (Wessler 1989). Le pourcentage d'amylopectine par rapport à l'amylose augmente, une caractéristique recherchée par l'industrie agroalimentaire malgré un rendement moins important de ces maïs.

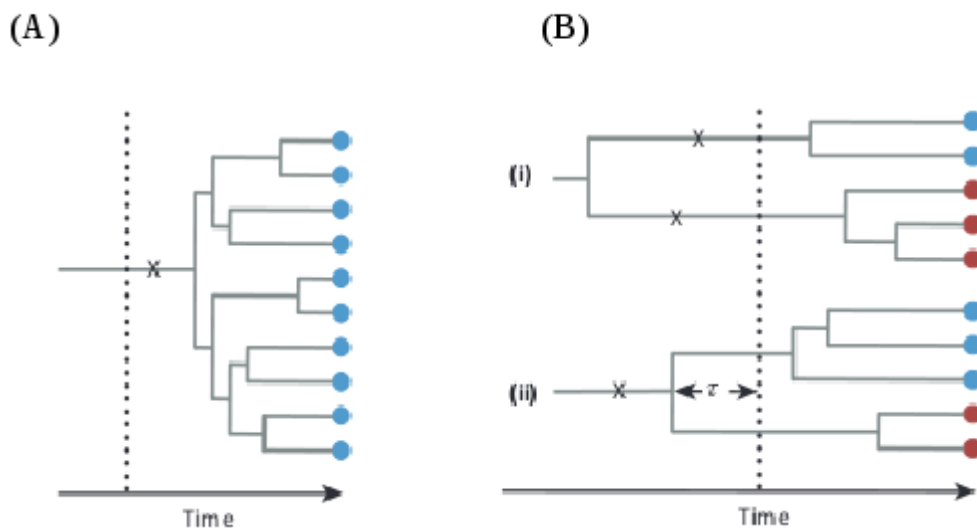
Les polymorphismes de structure du génome comme les inversions sont de bons candidats à l'adaptation locale. En effet, chez un individu hétérozygote, il n'y a pas de recombinaison au niveau de l'inversion. Ceci favorise le maintien de combinaisons favorables entre les gènes de l'haplotype de l'inversion (Kirkpatrick & Barton 2006). Il a été proposé qu'un regroupement génétique des gènes de domestication par le biais d'une suppression de la recombinaison puisse favoriser le maintien d'un phénotype domestiqué en dépit de flux de gènes récurrents avec les formes sauvages (Le Thierry d'Ennequin *et al.* 1999). Hoffmann & Weeks (2007) ont détecté une inversion dans le génome de la drosophile qui suit des clines de fréquence en relation avec l'adaptation aux régions tempérées sur le continent australien. Joron *et al.* (2006) ont découvert chez les papillons *Heliconius* l'existence d'un « super-gène », regroupement d'une trentaine de gènes protégés par des

réarrangements au sein du chromosome. Les différents haplotypes ainsi protégés codent pour des phénotypes reproduisant les patrons de coloration de papillons vénénéux de différentes espèces grâce aux relations épistatiques entre les éléments du « super-gène » (Joron *et al.* 2011).

### I.1.2.c Standing variation *versus* mutation *de novo*

Lorsqu'une mutation apparaît, elle est liée physiquement à l'haplotype dans lequel elle apparaît. Cette association peut être rompue par la recombinaison ou par l'apparition de nouvelles mutations. Lorsqu'un allèle est sélectionné et augmente en fréquence dans une population, il entraîne avec lui la région avoisinante liée par le déséquilibre de liaison avec une probabilité inversement proportionnelle au taux de recombinaison. C'est le modèle d'auto-stop génétique (Fay & Wu 2000 ; Kaplan *et al.* 1989 ; Maynard-Smith & Haigh 1974). La baisse de diversité locale ainsi que les changements associés (spectre de fréquence, diversité haplotypique) à cet événement de balayage sélectif permettent d'identifier la zone sous sélection.

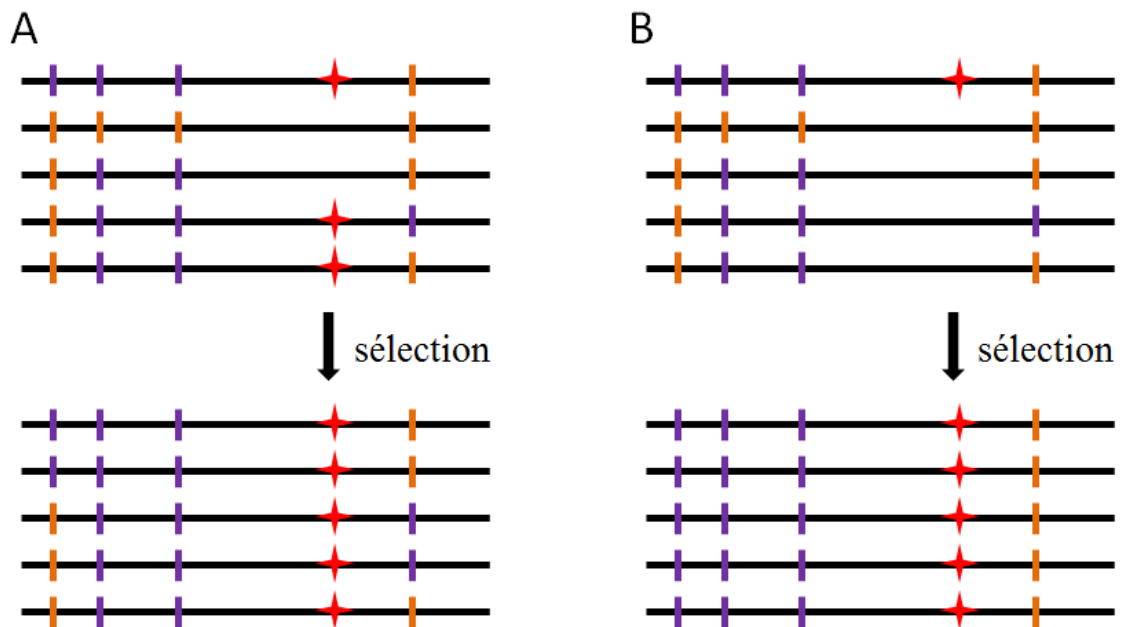
Deux cas de figures se présentent : l'adaptation *de novo* à partir d'une mutation nouvellement apparue dans la population, et l'adaptation à partir de variants déjà présents dans la population (standing variation ou variation préexistante). Ces deux cas de figures produisent des signatures sélectives différentes (Figure 5).



**Figure 5 : (A) Adaptation *de novo* versus (B) Adaptation à partir de variation préexistante.** Les croix symbolisent l'apparition de la mutation, les pointillés l'apparition de la pression de sélection. Figure adapté de Messer & Petrov (2013).



Lorsqu'un allèle avantageux apparaît *de novo* dans la population, il n'existe qu'une seule version de l'allèle avantageux (Figure 5 A, Figure 6 B). La sélection sur cet allèle produit alors un balayage sélectif fort : un seul haplotype est porteur de l'allèle avantageux. La baisse de la diversité locale est donc forte (Fay & Wu 2000 ; Maynard-Smith & Haigh 1974) et la diversité haplotypique faible (Figure 6 B). Après fixation de l'haplotype avantageux, l'action combinée de la recombinaison et l'apparition de nouvelles mutations conduit à une restauration de la variabilité génétique au fil du temps. Dans l'intervalle, le spectre de fréquence allélique est caractérisé par un excès de mutations à faible fréquence (celles récemment apparues par mutation) et de mutations à forte fréquence (les mutations avantageuses fixées ou celles associées au site sous sélection) (Fay & Wu 2000). Des méthodes basées sur l'analyse des spectres de fréquence ont été développées pour détecter de telles signatures (voir section I.1.3.e). Linnen *et al.* (2009) ont mis en évidence chez la souris sylvestre un allèle adaptatif responsable d'un changement de coloration dans les populations établies sur les collines de sable, ce qui leur confère un meilleur camouflage sur ce substrat clair. Une datation de cet allèle a permis d'établir qu'il est postérieur à la formation géologique de ces collines.



**Figure 6 : Mutation préexistante, (B) Mutation *de novo*.** La mutation avantageuse est signalée par une croix rouge.

Au contraire, un allèle déjà présent dans une population (standing variation) peut devenir adaptatif suite à l'apparition d'une nouvelle pression de sélection (Figure 5-B). Si la mutation à l'origine de l'allèle est suffisamment antérieure à l'apparition de la force sélective, l'allèle peut exister en plusieurs copies dans la population, qui ne sont pas nécessairement liées à un fond génétique unique car des événements de recombinaison ont pu l'associer à d'autres combinaisons d'allèles ou haplotypes (Figure 6-A). Dans ce cas, il se produit un balayage sélectif doux (ou soft sweep). La diversité n'est pas forcément réduite car plusieurs haplotypes coexistent (revue Messer & Petrov 2013). Le spectre de fréquence allélique est moins modifié que lors d'un balayage sélectif fort. Un cas connu d'adaptation basée sur la variation préexistante concerne le gène *teosinte branched 1* (*Tb1*) chez le maïs. *Tb1* est un locus impliqué dans la dominance apicale chez le maïs. L'élément transposable *Hopscotch* situé dans la région régulatrice de *Tb1* est un activateur qui est partiellement responsable de la différence d'expression du gène entre maïs et téosinte. L'insertion de cet élément transposable a été estimée à plus de 11 000 ans avant la domestication du maïs, ce qui indique que la sélection de ce gène a agit sur une variabilité préexistante chez la téosinte (Studer & Doebley 2011). L'existence de séries alléliques pour *Tb1* mise en avant par l'étude de Studer et Doebley (2012) appuie cette hypothèse. Cependant, la mutation de *Tb1* est présente à une si faible fréquence chez les téosintes que le signal produit ressemble à un balayage sélectif fort.

Un balayage sélectif faible peut aussi être dû à des mutations récurrentes si plusieurs mutations avantageuses apparaissent dans un laps de temps réduit. Dans ce cas, une mutation n'a pas encore été fixée dans la population quand une autre mutation adaptative apparaît. Plusieurs allèles avantageux sont donc sélectionnés en même temps et se propagent dans la population. C'est le cas, par exemple, des différents allèles de résistance chez le moustique *Culex pipiens* dans les zones géographiques traitées aux insecticides. L'émergence d'au moins trois allèles de résistance du gène de l'acétyl-cholinestérase a été facilitée par la potentialité du gène à se dupliquer (Labbé *et al.* 2007).

De nombreuses études se sont intéressées à la proportion des deux types de balayages sélectifs. Si la taille de la population et le taux de mutation sont faibles, les cibles de sélection sont peu réduites et il est peu probable de trouver une mutation adaptative préexistante dans la population. Dans ce cas, l'attendu est que le processus d'adaptation s'effectue surtout à partir de nouvelles mutations (Messer & Petrov 2013). Au contraire, l'adaptation à partir de variation préexistante est favorisée si l'avantage sélectif est faible, ou si le coefficient de sélection et le taux de mutation sont tous les deux élevés (Hermisson & Pennings 2005;

Messer & Petrov 2013). De part une plus grande facilité de détection, la majorité des adaptations étudiées sont des exemples de balayage sélectif fort (Messer & Petrov 2013; Pritchard & Di Rienzo 2010). Cependant, ces résultats ne reflètent pas forcément un processus biologique sous-jacent et sont en partie dus à la plus grande puissance nécessaire pour détecter les balayages sélectifs doux (Hermisson & Pennings 2005 ; Messer & Petrov 2013 ; Pritchard & Di Rienzo 2010). Détecter les signatures des balayages sélectifs doux dans un génome est un véritable défi. Les méthodes haplotypiques combinant l'information de plusieurs sites semblent être les plus appropriées (voir section I.1.3.e). Une étude comparant les séquences de micro-ARN chez *Arabidopsis thaliana* a mis en évidence un balayage sélectif doux sur le miR856 (de Meaux *et al.* 2008).

#### **I.1.2.d Convergence adaptative**

La réponse à une même contrainte sélective dans des populations distinctes peut engendrer des phénotypes différents ou au contraire des phénotypes convergents. Chez l'homme, la principale pression de sélection liée à l'altitude vient de la baisse de pression atmosphérique entraînant une diminution de l'oxygène disponible. Selon les populations humaines, différentes stratégies ont été développées pour répondre à cette pression. Beall *et al.* (1998) ont mesuré la concentration en hémoglobine dans les globules rouges chez les populations andines et tibétaines. Les populations andines présentent une augmentation de la concentration en hémoglobine comparées à la plupart des populations tibétaines qui présentent une concentration proche de celle des populations vivant au niveau de la mer (la concentration augmente chez les Tibétains vivant à plus de 4000 mètres d'altitude, Beall 2006). Chez les Tibétains, l'adaptation se fait via une très faible saturation en oxygène, tandis que chez les Andins une faible saturation en l'oxygène est couplée à une augmentation de la concentration en hémoglobine (Beall 2006). Les populations éthiopiennes vivant en altitude ne montre pas de différences notables pour ces deux caractères par rapport à des populations vivant au niveau de la mer, ce qui laisse supposer qu'il existe une troisième stratégie pour faire face à la baisse en disponibilité de l'oxygène (Beall *et al.* 2002).

Au contraire, face à des pressions de sélection similaires, les phénotypes adaptatifs peuvent être convergents. Par exemple, les plantes ayant adopté un système de reproduction autogame présentent une morphologie florale caractéristique appelée syndrome d'autogamie, et possèdent des fleurs plus petites et plus souvent fermées. Ce syndrome, adaptation importante sur l'efficacité de l'autogamie, apporte également d'autres avantages sélectifs. Il

permet de mieux allouer les ressources au sein de la plante en arrêtant la production de fleurs coûteuses en énergie pour attirer les pollinisateurs. Enfin, la production de fleurs moins voyantes et moins pourvues en substances nutritives permettrait de réduire l'herbivorie au niveau des apex floraux (revue Sicard & Lenhard 2011).

La question de convergence se pose aussi pour le génotype. Un grand nombre d'exemples concernant des cas d'adaptation via des gènes différents a été répertorié dans la littérature malgré un focus plus important sur les cas de convergences géniques. En effet, les cas de convergence sont souvent plus rapportés dans la littérature, car ils soulignent l'importance clé d'un gène sur un caractère adaptatif, et apportent des informations sur la prédictibilité de l'adaptation (Tiffin & Ross-Ibarra 2014). L'adaptation à la vie en altitude chez les mammifères est un mécanisme complexe impliquant plusieurs stratégies différentes et un grand nombre de gènes. Parmi ceux-ci, le gène *EPAS1* semble intervenir dans l'adaptation aux conditions hypoxiques dans deux espèces différentes, à la fois dans les populations humaines et canines des hauts plateaux tibétains (Wang *et al.* 2014). Le gène *EGLN1* est sélectionné, quant à lui, à la fois dans populations andines et tibétaines. Ce cas d'adaptation convergente implique cependant des haplotypes différents dans ces deux populations humaines (Scheinfeldt & Tishkoff 2013 ; Yi *et al.* 2010). Une expérience d'évolution expérimentale illustre particulièrement ce type de convergence. A partir d'une population initiale d'*Escherichia coli* maintenue à 37°C, Tenaillon *et al.* (2012) ont fait évoluer 115 populations sur 2000 générations à 42°C, maintenant une pression de sélection de haute température constante. Deux chemins adaptatifs principaux se distinguent, impliquant un grand nombre de gènes, dont *ybaL*. Sur ce gène, 38 mutations différentes impliquant des SNP, des changements de cadre de lecture et des insertions-délétions ont été trouvées dans 65 lignées. Ceci peut poser un problème lors de la recherche de convergence, notamment si la diversité génétique d'un locus est captée via des marqueurs différents. Plus généralement, cette étude montre que la convergence n'a pas nécessairement lieu au niveau de polymorphismes ponctuels mais plutôt au niveau d'un gène particulier ou au niveau d'une unité fonctionnelle comme les opérons ou des voies de biosynthèses (convergence de 2,6 % sur les polymorphismes ponctuels, de 20,3 % sur les gènes et de 24,5 % sur les opérons). Il est donc important de regarder des niveaux d'organisations élevés afin de ne pas omettre des cas d'adaptations convergentes.

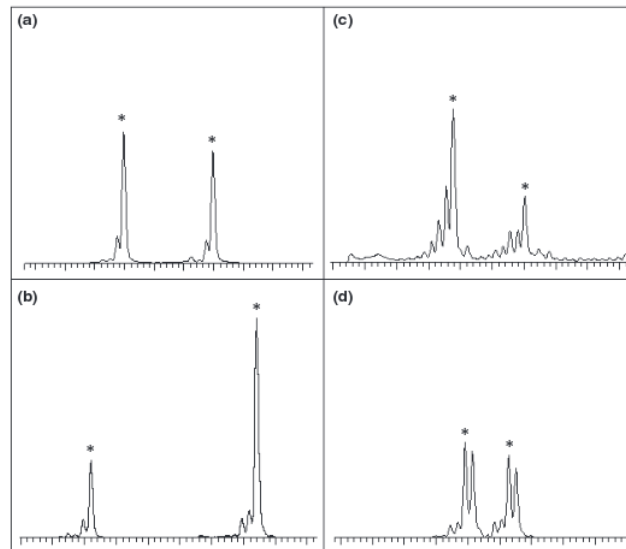
## **I.1.3 Méthodes moléculaires pour détecter l'adaptation**

### **I.1.3.a Génotypage haut-débit de positions prédéfinies**

Les marqueurs moléculaires sont des positions polymorphes du génome qui apportent des informations sur le génotype de l'individu. Ils peuvent renseigner directement sur l'histoire évolutive du locus concerné, ou sur les positions voisines. En effet des locus voisins sont en fort déséquilibre de liaison. Ce déséquilibre diminue avec la distance car la probabilité d'événements de recombinaison augmente. Les marqueurs moléculaires ont connus un succès énorme depuis la découverte de la PCR dans les années 70. Ils sont de nature variée et présentent des applications aussi diverses que les tests de paternité, les études de génétique des populations ou la cartographie génétique. Plus récemment, l'utilisation des marqueurs moléculaires a explosé grâce au développement des techniques haut-débit qui permettent d'utiliser ces marqueurs pour réaliser du génotypage haut-débit. Il existe plusieurs techniques de génotypage : les microsatellites, les puces à ADN, le KASPar et le GBS (*Genotyping by Sequencing*).

Les microsatellites sont de courts fragments d'ADN composés de motifs de 2 à 6 nucléotides répétés en tandem. Les microsatellites sont hautement mutables (taux de mutation estimé entre  $10^{-6}$  et  $10^{-2}$  par locus par génération selon le microsatellite) et sont relativement abondants dans la plupart de génomes (Ellegren 2000). Ils ont pour origine un glissement de la polymérase lors de la réplication de l'ADN dû à la répétition des motifs. La polymérase peut glisser vers l'avant et produire un microsatellite plus court que la version originelle, ou vers l'arrière, et dans ce cas produire une nouvelle version plus longue. Les microsatellites sont le plus souvent considérés comme neutres et sont codominants. Leur analyse nécessite peu d'ADN et ils sont reproductibles. Ils constituent donc de bons marqueurs moléculaires pour étudier les échelles évolutives courtes. Ils peuvent cependant présenter des allèles nuls, c'est à dire une version où les séquences flanquantes présentent une mutation et où les amorces ne peuvent se lier à la séquence d'ADN. Ces allèles nuls, détectables uniquement lorsqu'ils sont à l'état homozygote, peuvent biaiser les analyses. Le développement de marqueurs microsatellites nécessite un travail préalable important car il faut génotyper les séquences flanquantes des microsatellites et dessiner des amorces spécifiques au locus concerné. Une technique de génotypage des microsatellites couramment utilisée consiste à lier un fluorochrome aux amorces et à utiliser des désoxyribonucléotides marqués durant la phase d'élongation. Les pics de fluorescence sont ensuite lus et permettent de connaître le nombre

de répétitions du motif (Figure 7). Cette étape de lecture, cruciale dans le génotypage des microsatellites, peut être néanmoins délicate (Figure 7-b, c, d) et nécessite un protocole rigoureux lors des lectures des pics de fluorescence. En employant des fluorochromes différents et/ou en analysant des microsatellites de tailles variables, il est possible de génotyper plusieurs microsatellites en même temps, ce qui augmente considérablement le débit de génotypage (on parle de multiplexage).

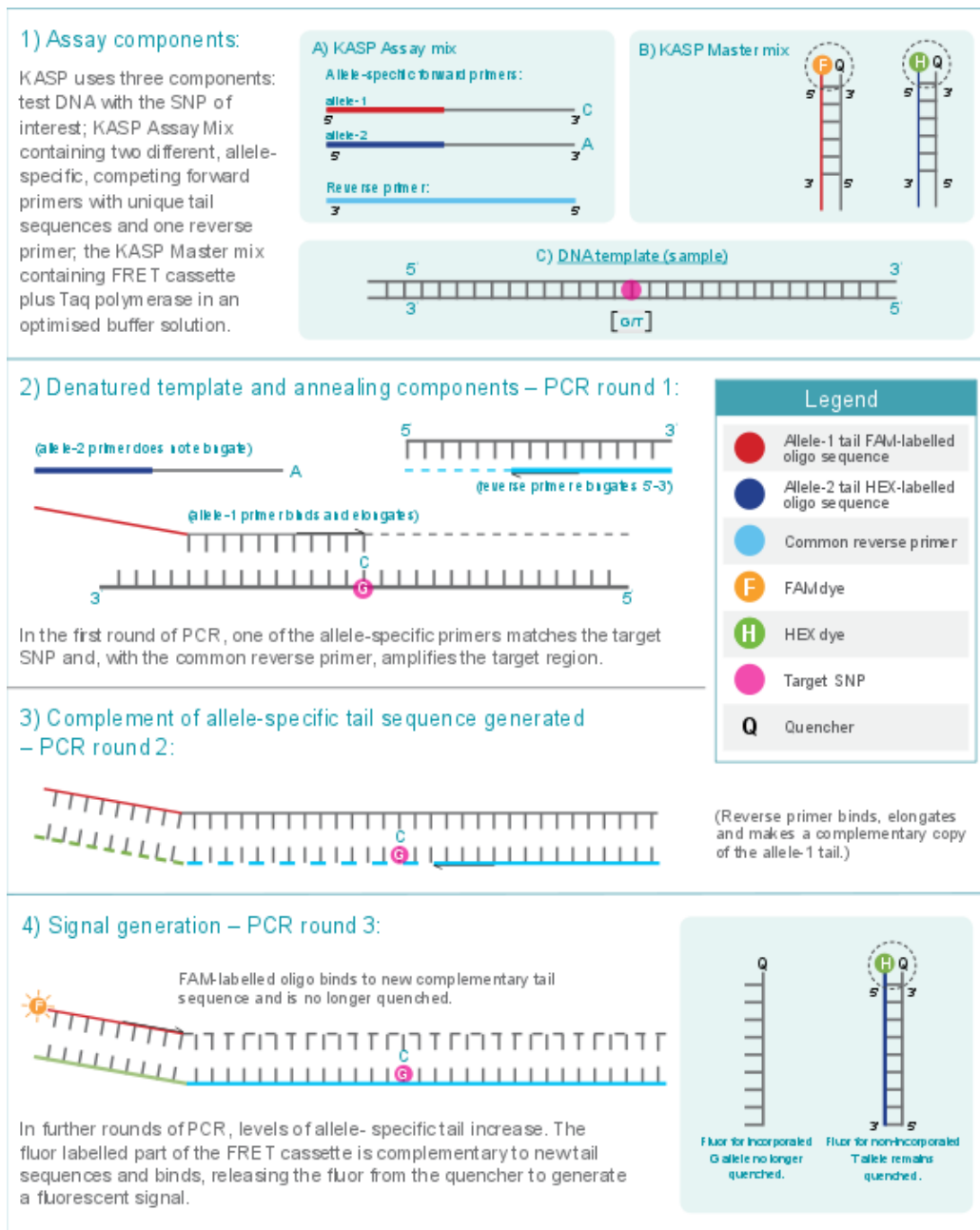


**Figure 7 : Profils de fluorescence générés lors du séquençage de microsatellites.** Les allèles véritables sont signalés par un astérisque. (a) Profil correct. (b) Ratio différentiel de fluorescence entre les deux allèles. (c) « Stuttering » ou étalement du signal de fluorescences sur plusieurs bases. (d) Pics doubles. Figure de Guichoux *et al.* (2011).

Les SNP sont beaucoup plus nombreux que les microsatellites et ont un taux de mutation relativement faible (chez les animaux entre  $10^{-11}$  et  $10^{-9}$  ; Baer *et al.* 2007 ; Drake *et al.* 1998, peu de rapports sur les plantes mais estimation chez le maïs pour le gène *Tb1* aux alentours de  $10^{-8}$  ; Clark *et al.* 2005). Ils peuvent être génotypés par différentes méthodes, comme par exemple via une puce à ADN ou via la technologie KASPar. Le principe des puces à ADN repose sur l'hybridation de fragments d'ADN dénaturés à des oligonucléotides fixés sur un support (appelés « sondes »). Chaque position couverte par la puce est dédoublée en deux versions d'oligonucléotides, complémentaires des allèles existants pour le SNP. La lecture du génotype consiste à repérer au moyen de fluorochromes à quels oligonucléotides les morceaux d'ADN se sont hybridés. Les puces à ADN se basent donc sur des fragments du génome de séquence connue et pour lesquels un polymorphisme a été au préalable découvert. Elles permettent de génotyper un grand nombre de SNP pour un coût réduit. L'usage d'une

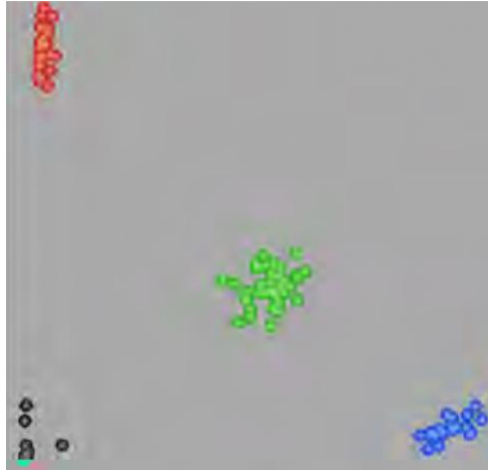
puce unique à ADN permet la comparaison d'études utilisant cette dernière sur des individus différents.

La technologie KASPar permet de génotyper des SNP en se basant sur le principe d'hybridation compétitive lors de la PCR. Deux amorces sont dessinées, chacune complémentaire d'un allèle et dont une des extrémités contient une séquence distincte (Figure 8-1). Le brin complémentaire de cette séquence est formé lors de l'amplification de l'ADN si l'amorce s'est hybridée (Figure 8-2, 3). Des fragments d'ADN, semblables à ceux des extrémités des amorces et porteurs d'un fluorochrome, s'hybrident alors avec ce brin complémentaire lors de la phase de dénaturation, libérant le fluorochrome qui émet un signal lumineux (8-4). Si un seul signal de fluorescence est capté, l'individu est homozygote pour l'allèle correspondant à la longueur d'onde du fluorochrome correspondant. Si deux signaux sont captés, l'individu est hétérozygote (Figure 9). La technologie KASPar est plus chère que les puces à ADN mais aussi plus fiable.



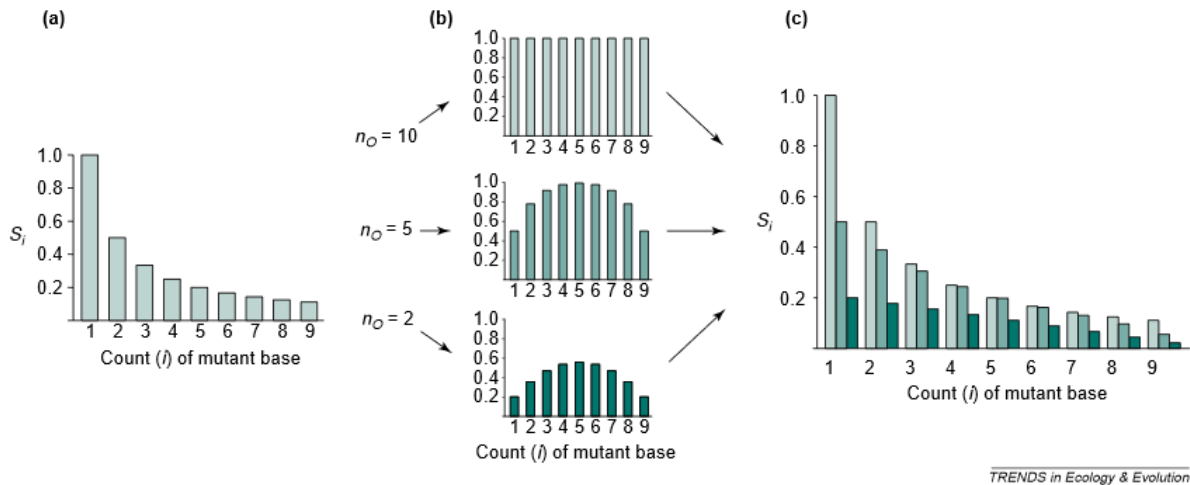
**Figure 8 : Génotypage par KASPar.** (1) Eléments utilisés durant le génotypage. (2) Hybridation d'une des deux amorces en compétition. (3) Elongation des séquences complémentaires au brin généré au premier cycle. (4) Libération du signal fluorescent. Figure provenant de [www.lgcgroup.com](http://www.lgcgroup.com).





**Figure 9 : Exemple de signal obtenu par la technologie KASPar.** En rouge et bleu, des individus homozygotes. En vert, des individus hétérozygotes. Chaque axe représente l'intensité de fluorescence d'un des deux marqueurs.

Les microsatellites, les puces à ADN et le génotypage KASPar sont des technologies qui nécessitent la découverte préalable de polymorphismes dont la séquence génomique environnante est connue pour pouvoir dessiner les amorces/sondes. Ces polymorphismes sont découverts à partir d'un échantillon d'individus (échantillon de découverte) avant d'être génotypés à plus large échelle. Les biais liés à la découverte des SNP (ou *ascertainment bias* en anglais) sont de deux natures différentes : l'échantillon de polymorphismes n'est pas forcément représentatif de l'espèce étudiée (certains polymorphismes peuvent être absents de l'échantillon de découverte), et selon la taille de l'échantillon de découverte, seuls les allèles à forte fréquence pourront être découverts (plus l'échantillon de découverte est petit, plus la fréquence des polymorphismes dans l'espèce devra être élevée pour qu'ils aient une forte probabilité d'être découverts). Le spectre de fréquence des marqueurs découverts est donc biaisé vers un excès de marqueurs à forte fréquence par rapport à l'attendu neutre (Figure 10). Ces problèmes introduisent des biais, notamment pour les études de génétique des populations et la détection des signaux de sélection (Albrechtsen *et al.* 2010). L'*ascertainment bias* est plus marqué pour les SNP que pour les marqueurs microsatellites qui présentent un polymorphisme plus important (Rogers & Jorde 1996).



**Figure 10 : Ascertainment bias sur le spectre de fréquence de polymorphisme de SNP sur un échantillon de 10 individus en fonction de la taille de l'échantillonnage ( $n_0$ ) sur lequel s'est basée la découverte des SNPs.** (a) Spectre de fréquence obtenu sans *ascertainment bias*, c'est-à-dire si les positions polymorphes sont obtenues directement de la population ( $n_0=10$ ). (b) Probabilités des SNP d'être découverts selon leur proportion et la taille de l'échantillonnage sur lequel s'est basée la découverte de SNP ( $n_0$ ). (c) Spectres de fréquence obtenus selon les *ascertainment bias* correspondant aux différents  $n_0$ . Figure de Morin *et al.* (2004).

### I.1.3.b (Re-)séquençage haut-débit de génomes complets

Le séquençage Sanger, inventé en 1977 par Frederick Sanger, est un ingénieux procédé pour séquencer de l'ADN. Des didésoxyribonucléotides sont ajoutés en petite quantité à un mélange réactionnel pour répliquer de l'ADN. Lorsqu'un didésoxyribonucléotide est incorporé à la séquence en croissance, l'élongation du brin d'ADN ne peut se poursuivre, ce qui conduit à une séquence interrompue. L'usage successif de quatre mélanges réactionnels (un pour chacun des nucléotides A, T, G, C) et la migration par électrophorèse des fragments d'ADN de tailles différentes ainsi obtenus permettent de lire la séquence d'ADN. Cette technique, qui valut le prix Nobel de Chimie en 1980 à Sanger, a permis d'accéder directement à l'information reine : la séquence d'ADN. Cette méthode a été perfectionnée par la suite pour obtenir plus rapidement des fragments plus longs, plus nombreux, et surtout pour un coût moins important. Cependant, malgré toutes les améliorations apportées, les biologistes étaient toujours en besoin d'une technologie capable de séquencer des génomes entiers en peu de temps et pour un coût limité.

Les nouvelles générations de séquençage (*Next Generation Sequencing* ou NGS) sont un ensemble de techniques permettant de séquencer en masse des dizaines de millions de fragments simultanément et pour un coût relativement bas. Les NGS peuvent être utilisées pour du séquençage *de novo* pour les espèces ne possédant pas de génome de référence

(espèces non modèles) ou pour du re-séquençage, c'est-à-dire pour des espèces pour lesquelles un génome de référence est disponible pour exploiter ensuite les données obtenues. La matrice utilisée peut être de l'ADN natif (séquençage génomique ou DNA-Seq), de l'ARN (séquençage de transcriptome ou RNA-Seq), de l'ADN modifié par traitement bisulfite (Bisulfite-Seq)... Quel que soit le type de matrice utilisée, elle est fixée à un support ce qui permet le séquençage en parallèle de millions de séquences.

Les NGS ont révolutionné les études écologiques en permettant d'étudier des espèces non modèles, et en permettant un changement d'échelle : au niveau génomique et non plus du gène seul (Hudson 2008). Les NGS se basent sur des technologies variées, mais reposent sur les mêmes principes : la préparation des bibliothèques, l'amplification parallèle des molécules à analyser, puis la réaction de séquençage. Plusieurs technologies de séquençage existent : le pyroséquençage – Roche 454, le séquençage à l'aide de terminaux réversibles – Illumina, le séquençage par ligation – SOLiD –, le séquençage par semi-conduction – Ion Torrent – et le séquençage de troisième génération – PacBio – (Tableau 1).

Selon la technologie employée, les lectures obtenues ont des tailles différentes, allant de la dizaine au millier de bases nucléotidiques. Les prix sont généralement croissants avec l'augmentation de la longueur des lectures, mais à ces prix technologiques s'ajoute le prix du traitement de données. En effet l'énorme quantité de données générées nécessite un travail bio-informatique de développement d'algorithmes permettant d'automatiser leur gestion. Le choix d'une technologie par rapport à une autre va se faire sur le budget disponible, mais aussi l'utilisation du biologiste. Pour du séquençage *de novo*, qui consiste à séquencer des génomes inconnus, plusieurs technologies de séquençage peuvent être combinées, permettant d'allier leurs forces et de pallier à leurs faiblesses respectives (notamment au niveau des erreurs de séquençage et des longueurs des séquences obtenues) afin d'obtenir un génome de référence de qualité. Les technologies de séquençage peuvent aussi être utilisées quand un génome de référence est disponible et permettent de couvrir un très grand nombre de positions (Tableau 1, Liu *et al.* 2012 ; Mardis 2008).

Le GBS offre une alternative intéressante pour réduire les coûts et augmenter le nombre d'échantillons. Le GBS consiste à génotyper simultanément un grand nombre d'individus (multiplexage) en réduisant la complexité du génome étudié. Cette réduction de la complexité se fait par des enzymes de restriction ciblant par exemple les régions géniques. Le GBS ne nécessite pas forcément de génome de référence ce qui en fait une technique intéressante pour les organismes non modèles. Les traitements informatiques sont cependant

très lourds : les génomes sont séquencés avec une assez faible profondeur, il y a du polymorphisme de sites de restriction et les fragments séquencés ne sont donc pas tous les mêmes entre individus, ce qui conduit à un grand nombre de données manquantes.

**Tableau 1 : Comparaison des différentes technologies de séquençage.** Adapté de Liu *et al.* (2012) et de Quail *et al.* (2012).

Technologie de séquençage	Roche 454 GS FLX Titanium	Illumina HiSeq 2000	SOLiD v4	Sanger 3730xl	IonTorrent NGS system	PacBio
Longueur des séquences	700 bp	100 PE	50+ 35 bp 50+ 50 bp	400~900	200-400 bp	1300 bases en moyenne
Précision (%)	99,9	98	99,94	99,99	98,5	95,95
Nombre de lectures	1 M	3G	1200~1400 M	96	400-550 mille à 5.5 million	50 000
Données par run	0,7Gb	600Gb	120Gb	1,9~84 Kb	30 Mb à 1Gb	300 - 500 Mb
Durée d'un run	24 heures	3 ~ 10 jours	14 jours	20 mn ~3 heures	2 heures	30 mn à 6 heures
Avantage	Longueur des séquences rapidité	Volume des données	Précision	Qualité et longueur des lectures	Rapidité, Coût peu élevé	Pas d'amplification Rapidité Longueur des lectures
Inconvénient	Taux d'erreur, coût élevé, faible volume de données générées	Longueur des lectures	Longueur des lectures	Coût élevé, faible volume de données générées.	Préparation en amont Coût	Taux d'erreur élevé Coût
Coût / million bases	\$10	\$0,07	\$0,13	\$2400	\$1	\$2
Re-séquençage		Oui	Oui	Oui		Oui
<i>De novo</i>	Oui	Oui		Oui	Génome microbien	Oui
Grand génome	Oui	Oui			Non	Oui
Détection de polymorphisme	Oui	Oui	Oui	Oui	Oui	Oui

### I.1.3.c Détection des polymorphismes par re-séquençage NGS

Les données générées par les NGS permettent d'identifier les SNP, mais aussi les INDEL (insertions-délétions), les réarrangements génomiques et les polymorphismes de nombre de copies. La détection des SNP a particulièrement été développée, d'une part parce que techniquement plus facile, et d'autre part parce que ces marqueurs sont généralement bialléliques et constituent la variation la plus commune dans les génomes (Mardis 2008). L'analyse des données de re-séquençage de génomes complets nécessite un génome de référence. Les principales étapes sont le nettoyage des données brutes, l'alignement des lectures sur le génome de référence, le *SNP calling* et le *genotype calling*.

Le nettoyage de données permet d'éliminer certaines lectures et/ou bases, en fonction du score de qualité de séquençage fourni avec les données brutes. Le score de qualité de séquençage dépend de nombreux paramètres comme la longueur des lectures par rapport à l'attendu ou l'intensité du signal de la base (par exemple la fluorescence). Les plateformes de séquençage adoptent des scores de qualité de séquençage différents, mais chacun peut être transformé en *Phred score* standard, pour lequel une valeur de 20 correspond à une probabilité de 1% d'être en présence d'une erreur de séquençage (Ewing & Green 1998).

L'alignement des lectures sur le génome de référence est une étape capitale du génotypage par re-séquençage de génomes complets, car des lectures mal-alignées conduisent à des erreurs dans les étapes ultérieures. Un bon aligneur doit savoir évaluer la probabilité d'erreur de séquençage par rapport à une différence réelle (SNP, indel) entre l'échantillon considéré et le génome de référence. Il existe deux grands types d'aligneurs : ceux basés sur les tables de « hashage » qui est une adaptation des algorithmes de BLAST (type Stampy) : ils découpent le génome en Kmer de taille fixée et référencent les positions associées dans la table de hashage. Ils positionnent ensuite à partir d'un ou plusieurs Kmers la séquence à aligner. Le deuxième type d'aligneurs concerne ceux basés sur des algorithmes qui reposent sur des arbres de suffixes/préfixes utilisé dans la recherche de motif (type Bowtie) : à partir de la séquence de référence ils construisent un arbre de suffixe, à partir d'un parcours de cet arbre ils essaient de positionner la séquence dans l'arbre (Li & Homer 2010 ; Nielsen *et al.* 2011). Ces derniers sont plus rapides et particulièrement efficaces pour les séquences répétées, mais sont moins sensibles que ceux basés sur les tables de « hashage ». L'aligneur Stampy représente un bon compromis entre efficacité, performance et rapidité (Lunter & Goodson 2011 ; Nielsen *et al.* 2011).

Le *SNP calling* consiste à déterminer les positions variantes du génome, le *genotype*

*calling* à déterminer le génotype de l'échantillon aux positions variantes. Il existe plusieurs méthodes de *SNP* et *genotype calling*. Des méthodes « classiques » ou empiriques de *genotype calling* consistent par exemple à compter à une position pour un individu donné le nombre de fois qu'un allèle est observé et présente un *Phred score* supérieur à une valeur déterminée, puis de définir le génotype en fonction de la fréquence d'observation de l'allèle alternatif et d'un cut-off. Cette méthode est raisonnable pour une couverture supérieure à 20x (Nielsen *et al.* 2011).

Les méthodes probabilistes sont généralement plus performantes, surtout pour les faibles couvertures. Elles calculent la probabilité de chaque génotype en fonction d'un prior qui intègre l'erreur de séquençage, des informations sur la qualité de l'alignement, sur l'hétérozygotie et/ou sur le déséquilibre de liaison. Il existe plusieurs façons de déterminer ensuite le génotype : un génotype exact ou bien des probabilités d'appartenance aux différents génotypes peuvent être proposés. SAMtools propose un ensemble d'outils permettant de calculer les probabilités de chaque génotype (SAMtools), puis d'effectuer le *SNP calling* et *genotype calling* (BCFtools, Li *et al.* 2009).

L'un des principaux inconvénients du génotypage par séquençage NGS vient du taux d'erreur relativement important dû d'une part aux erreurs de séquençage et d'autre part aux erreurs d'alignement. Ces variants erronés posent des problèmes d'analyse notamment parce qu'ils engendrent des singletons (variants présents une seule fois dans l'échantillon) qui ont ensuite un poids important dans certaines analyses de génétique des populations. Pour pallier à ce problème, Han *et al.* (2013) préconisent de recourir à des données ayant au minimum 10x de profondeur. En théorie, avec cette profondeur, il est possible de distinguer les erreurs de séquençage des vrais polymorphismes car chaque base est couverte plusieurs fois par des lectures. Il y a peu de chance en effet qu'une erreur se reproduise plusieurs fois à la même base. SAMtools suppose l'indépendance à chaque base de toutes les lectures séquencées et prend en compte les erreurs d'alignement avant d'inférer les génotypes.

#### **I.1.3.d Séquençage en mélange**

Bien que les coûts de séquençage aient drastiquement chuté au cours des dernières années avec l'avènement des NGS, séquencer plusieurs individus nécessite un budget conséquent. Une alternative au séquençage individuel est le séquençage en mélange, c'est à dire le séquençage de plusieurs individus en même temps. Le séquençage en mélange permet d'obtenir une estimation des fréquences alléliques à chaque site comme un tirage aléatoire des

allèles présents dans l'échantillon. Il existe cependant plusieurs inconvénients, par exemple des biais liés à la contribution hétérogène des individus à un site donné, la plus faible profondeur de séquençage individuelle à volume constant et le fait que l'on ne peut pas avoir accès aux haplotypes. Zhu *et al.* (2012) ont estimé ces biais en re-séquençant en mélange des populations de *Drosophila melanogaster* créées à partir de lignées isogéniques aux génotypes connus. Les auteurs ont comparé les fréquences alléliques obtenues avec les deux méthodes de séquençage (en mélange *versus* individuelle). Ils ont montré que le bruit provenant du séquençage en mélange pouvait être approximé par un échantillonnage dans une distribution binomiale. La variance créée par une contribution inégale des individus pouvait être grandement diminuée en mélangeant un nombre suffisant de chromosomes (dans leur cas, provenant de 42 lignées) et en mélangeant *a posteriori* les ADN extraits séparément et non en extrayant l'ADN à partir des mélanges de tissus. Si la qualité des données est un critère essentiel pour du séquençage individuel, ceci est d'autant plus vrai pour un séquençage en mélange. L'étude de Zhu *et al.* (2012) indique qu'il faut une profondeur minimale de 20x par pool pour 42 chromosomes. Bien que plus complexe à traiter que le séquençage individuel à cause des biais associés, le séquençage en pool a fait ses preuves dans de nombreuses études, comme celle de Turner *et al.* (2010) sur l'adaptation aux sols serpentiniques chez *Arabidopsis lyrata*.

### **I.1.3.e Méthodes statistiques de détection de la sélection**

Dans leur très grande majorité, les marqueurs ne sont pas impliqués dans des processus adaptatifs. Cependant ils peuvent être proches d'un locus sous sélection, *i.e.* en DL, avec un marqueur sous sélection, et être entraînés avec lui par le phénomène d'autostop génétique. De nombreuses méthodes de détection de la sélection se basent sur ce principe pour identifier des régions du génome sous sélection. Les données de séquence sont analysées grâce à des outils de génomique des populations pour réaliser ce que l'on appelle des scans génomiques. Il existe quatre grands types de méthodes détaillées ci-dessous : les méthodes basées sur des statistiques résumées du spectre de fréquence, qui peuvent être employées entre populations ou au sein d'une population ; les méthodes basées sur les haplotypes, donc le DL ; les méthodes de corrélation avec des variables environnementales et les méthodes de clines de fréquence.

#### *Détection intra-populationnelle*

Un spectre de fréquence correspond à la distribution des variants observés dans une population. Les tests basés sur le spectre de fréquences alléliques comparent la distribution des variants à une distribution neutre selon un modèle défini (Figure 11; Nielsen 2005). Lors d'un balayage sélectif on observe un excès de variants rares ainsi qu'une forte fréquence de variants dérivés qui se font « entraîner » par le phénomène (Figure 11). Sous sélection négative, les nouveaux polymorphismes sont contre-sélectionnés par la sélection naturelle, causant ainsi un excès de variant rares (Figure 11). Les balayages sélectifs liés à l'adaptation locale peuvent être mis en évidence en comparant au sein d'une population le spectre de fréquence d'une région du génome à celui observé dans d'autres régions. Boitard *et al.* (2012) ont développé une méthode pour détecter ces balayages sélectifs en partant de données séquencées en mélange et l'ont implémentée dans le programme PoolHMM (Boitard *et al.* 2013). Les probabilités des Spectres de Fréquences Alléliques (SFA) sont calculées en attribuant des poids différents selon la qualité des données. Les SFA sont ensuite assignés à trois états possibles: sous sélection, neutre ou intermédiaire. La transition entre ces états se fait selon une probabilité intégrée au modèle.

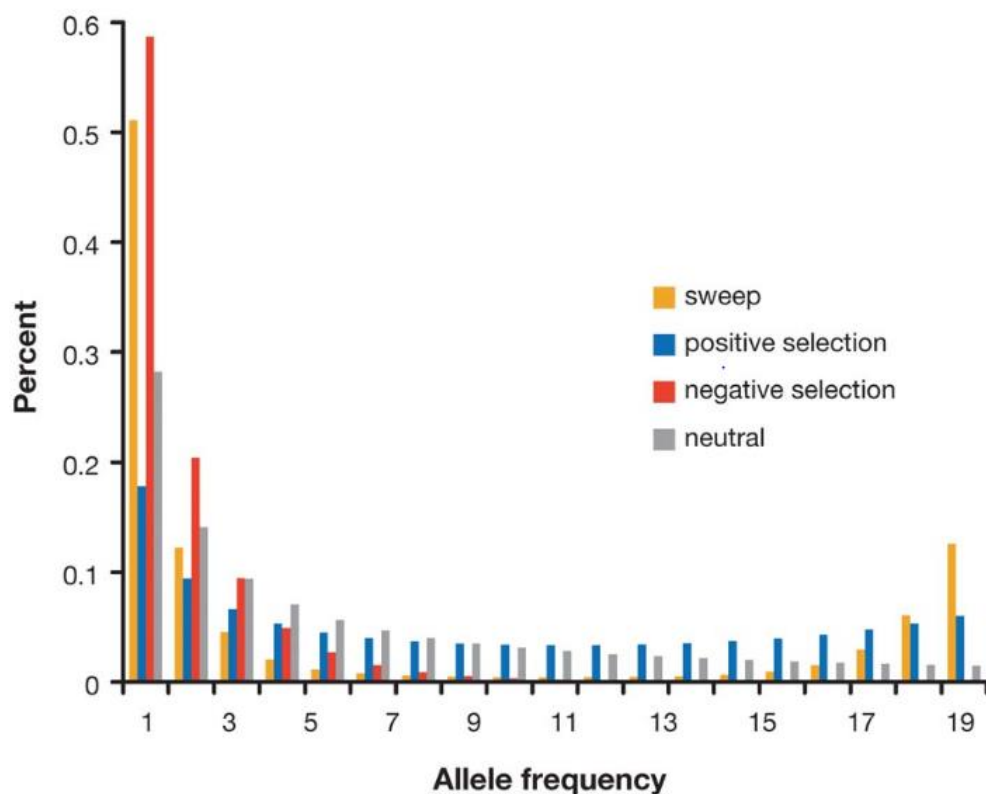
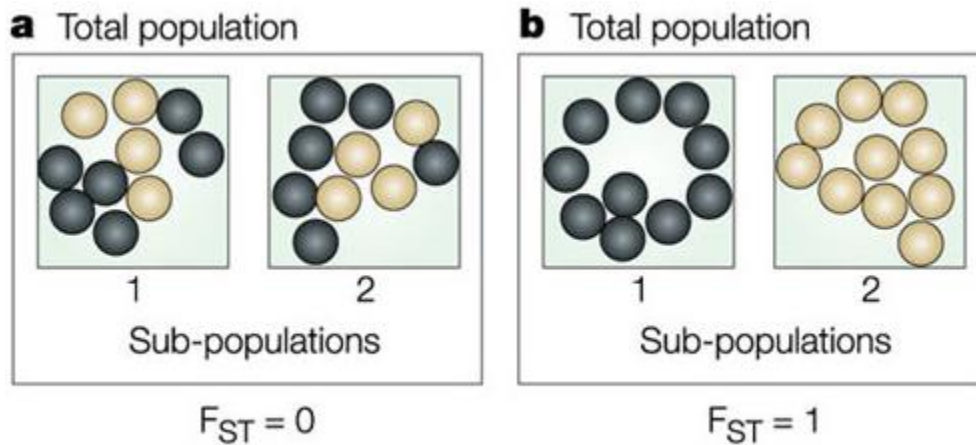


Figure 11 : Spectre de fréquence allélique sous un balayage sélectif (orange), de la sélection positive (bleu), de la sélection négative (rouge) ou sous un modèle neutre (gris). Figure de Nielsen (2005).



### *Détection de la sélection inter-populationnelle*

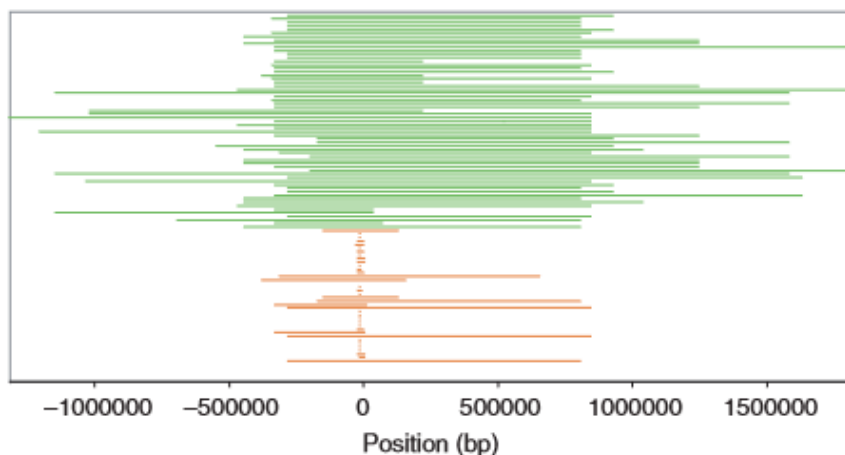
Le  $F_{st}$  est une statistique de différenciation des populations. En cas de sélection entre populations pour des allèles différents, on s'attend à ce que les locus sous-sélection soient plus différenciés que les locus neutres. Pour un locus donné, si les populations ne sont pas différenciées, le  $F_{st}$  vaut 0 ; si au contraire les populations sont fixées chacune sur un allèle particulier, le  $F_{st}$  vaut 1 (Figure 12). Comparer les valeurs de  $F_{st}$  des locus permet d'identifier des locus hors-types ou « outliers », c'est-à-dire des locus présentant une différenciation génétique plus importante que celle attendue sous un modèle neutre. Cependant, l'histoire démographique et la structuration spatiale des populations peuvent fausser l'interprétation de la différenciation allélique entre populations (Excoffier *et al.* 2009). Certaines méthodes ont donc intégré une structuration spatiale dans leur décomposition de la variance allélique. BayeScan, une méthode développée par Foll & Gaggiotti (2008), mesure la différenciation de chaque population dans un contexte de structuration des populations. Cette différenciation est appréciée par un coefficient  $F_{st}$  propre à chaque locus et chaque population. Le  $F_{st}$  est décomposé en une composante démographique  $\beta$ , partagée par tous les locus d'une même population, et une composante propre au locus,  $\alpha$ . Si la composante  $\alpha$  est nécessaire pour décrire le patron de diversité d'un locus, alors ce locus présente une déviation à la neutralité. Les probabilités postérieures d'appartenance à un modèle incluant  $\alpha$  (modèle avec sélection) par rapport au modèle sans  $\alpha$  (modèle neutre) sont ensuite calculées pour chaque locus. Un ensemble de SNP peut être testé directement sans *a priori*. Il est également possible d'identifier au préalable des SNP candidats et un jeu de SNP neutres pour calibrer le composant  $\beta$ . Keller *et al.* (2012) ont analysé 335 SNP provenant de gènes candidats sur un total de 747 SNP chez le peuplier. Ils ont mis en évidence avec cette méthode 14 SNP, dont 11 porteurs d'un signal très important (probabilité du modèle sous sélection et concordance avec d'autres méthodes).



**Figure 12 : Fst entre deux populations.** (a) Les populations ne sont pas différenciées, le Fst est de 0. (b) Les populations sont totalement différenciées, un allèle différent est fixé dans chaque population, le Fst est de 1. Figure de Bamshad *et al.* (2004).

### *Les méthodes haplotypiques*

Un haplotype est une combinaison donnée de nucléotides pour une région chromosomique donnée. Un allèle sous sélection qui augmente rapidement en fréquence emporte avec lui ses voisins par un effet d'entraînement (autostop) génétique. Cette liaison conduit à la formation d'haplotypes plus longs qu'attendus (Figure 13). Les haplotypes combinent ainsi plusieurs locus en un seul marqueur, et constituent un fort signal de sélection. L'accès à l'information haplotypique est cependant restreint au séquençage individuel, et nécessite une bonne couverture puisqu'il est nécessaire de phaser les données, c'est à dire de reconstruire les combinaisons de SNP appartenant à chaque allèle. Sous ces conditions, la longueur des haplotypes peut être testée afin de mettre en évidence des régions du génome sous sélection. Fariello *et al.* (2013) ont développé une méthode qui détecte la sélection en se basant sur les différences de fréquences haplotypiques entre les populations. Cette méthode est développée dans le logiciel hapFLK (Fariello *et al.* 2013). Elle a été élaborée à partir de la méthode FLK, proposée par Bonhomme *et al.* (2010), qui prend en compte une structuration hiérarchique entre populations via la construction d'une matrice d'apparentement estimée à partir de la distance de Reynolds entre paires de populations. La méthode hapFLK généralise ce test aux données haplotypiques recelant plus d'informations, en utilisant un modèle de clustering des haplotypes. En utilisant cette méthode, Fariello *et al.* (2014) ont mis en évidence de nouvelles régions montrant des signatures de sélection récentes chez le mouton domestique et associées à des caractères d'intérêt (morphologie, coloration).



**Figure 13 : Longueur des haplotypes autour de la position générant la persistance de la lactase chez l'humain après la période adolescente.** Dans certaines populations européennes et asiatiques cette persistance a été sélectionnée : les haplotypes associés à la persistance de la lactase (vert) sont plus longs que les haplotypes associés à la non persistance de la lactase (orange). Figure de Tishkoff *et al.* (2007).

#### *Les méthodes de corrélation avec l'environnement*

D'autres méthodes sont basées sur la recherche de corrélations entre la variation environnementale et la variation génétique entre populations. Cependant, des populations voisines possèdent souvent une histoire évolutive commune et peuvent être soumises à des conditions environnementales similaires. Des méthodes ont été développées afin de prendre en compte cette structure des populations pour éviter les faux positifs. BAYENV par exemple est un outil développé par Coop *et al.* (2010) qui teste l'association entre les variations de fréquences alléliques et les variations environnementales dans un cadre bayésien. Dans un premier temps, un modèle nul prenant en compte la covariance des fréquences alléliques entre les populations est construit. Pour chaque locus, un deuxième modèle autorisant une déviation linéaire de la matrice de covariance par une composante environnementale est testé. Les probabilités postérieures d'appartenance à l'un ou l'autre des modèles sont ensuite estimées. Ces facteurs de Bayes ne sont pas à proprement parler une mesure de la signification et ne peuvent être comparés entre variables environnementales testées. Ils permettent cependant d'identifier les locus hors-type par rapport à la distribution des facteurs de Bayes. La dernière version en date de BAYENV, BAYENV 2.0, permet également de prendre en compte des données de séquençage en mélange (Günther & Coop 2013). Ces méthodes de corrélation avec des variables environnementales replacent les études de l'adaptation dans un contexte écologique et aident à mieux comprendre quelles variables ont un impact important pour

l'adaptation locale (Tiffin & Ross-Ibarra 2014). Berg *et al.* (2015) ont utilisé cette méthode chez la morue d'Atlantique pour tester des associations entre les données de SNP d'une part, et la salinité, la température et la concentration en oxygène d'autre part. Parmi les SNP identifiés, deux SNP situés près d'un homologue de *AQPI*, un gène impliqué dans l'hydratation des oocytes, sont associés à la température et la salinité. Ce gène interviendrait également dans les mécanismes d'isolement reproducteur entre deux sous-groupes génétiques de la morue d'Atlantique.

### *Les clines de fréquences*

Lorsqu'une pression de sélection s'exerce localement sur une population, il peut se former des clines de fréquences (Figure 3). Ces clines sont générés par l'équilibre entre la sélection naturelle agissant localement sur une population et la migration apportant un lot de génotypes non exposés à la sélection (Huxley 1938). Si certains marqueurs se démarquent de la tendance générale du génome et suivent un cline, il se peut qu'ils soient soumis à la sélection. Cependant, la présence d'un cline ne traduit pas forcément une adaptation, car un cline peut se produire lors d'un contact secondaire entre deux populations ayant divergé, ou bien par un équilibre entre dérive génétique et migration entre populations. Dans ces cas-là, on recherche les marqueurs sortant de la distribution globale des clines obtenus pour l'ensemble des marqueurs et qui sont potentiellement soumis à la sélection.

Une méthode de visualisation du cline consiste à exprimer les fréquences alléliques en fonction d'une distance géographique ou génétique entre populations (Derryberry *et al.* 2014 ; Gay *et al.* 2008). Deux paramètres sont couramment utilisés pour étudier les clines : le point d'inflexion du cline et la pente. L'étude de ces deux éléments renseigne sur la force de sélection s'exerçant sur la population. L'analyse des clines a été particulièrement utilisée sur les zones hybrides entre deux espèces ou sous-espèces (Gay *et al.* 2008 ; Derryberry *et al.* 2014), mais aussi pour l'étude d'adaptation à des pressions de sélection suivant un gradient (Labbé *et al.* 2009 ; Mullen & Hoekstra 2008). Par exemple, le rongeur *Peromyscus polionotus* présente une gamme de variation phénotypique de la coloration et de la réflexion du pelage le long d'un transect de 155km où la nature et la coloration du sol changent graduellement. Mullen & Hoekstra (2008) ont éliminé les hypothèses de contact secondaire ou de dérive génétique en se servant de marqueurs microsatellites considérés comme neutres. L'analyse des fréquences alléliques de deux gènes de pigmentation suggère que l'un d'eux, *Agouti*, est impliqué dans l'adaptation via la pigmentation.

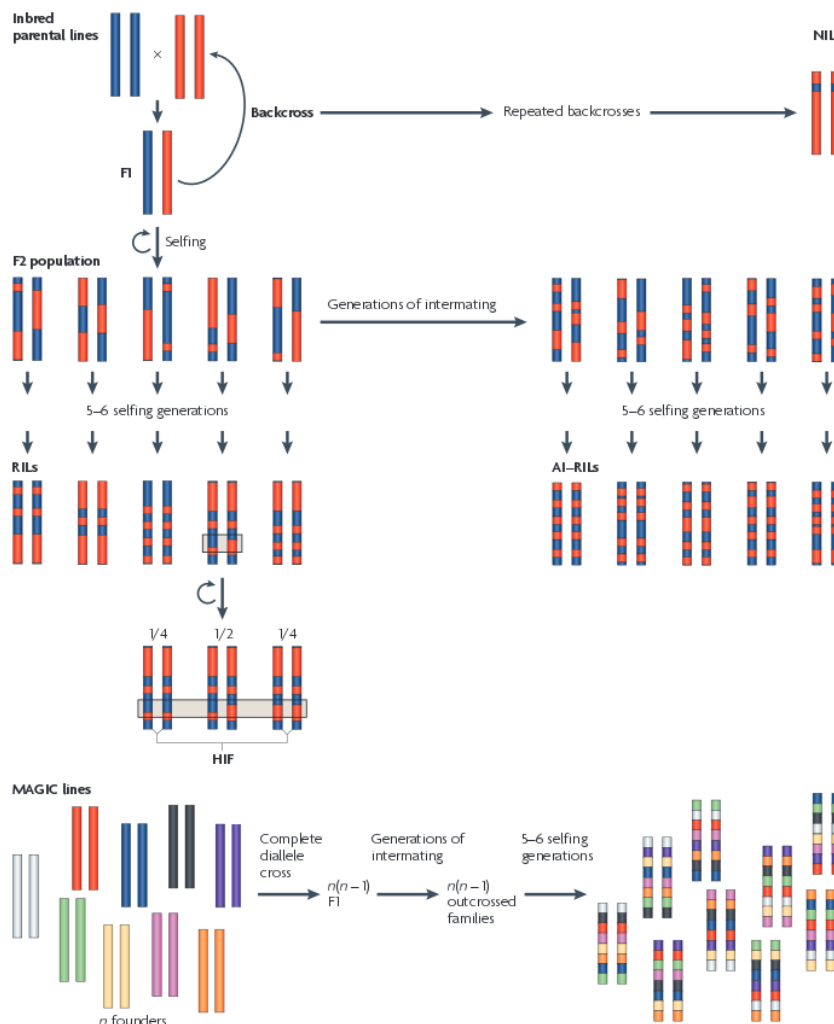
### **I.1.3.f Etudes d'association avec le phénotype**

Les méthodes de détection de la sélection sont une première étape dans l'identification des déterminants génomiques de l'adaptation. Elles servent de guide pour définir des régions candidates, potentiellement impliquées dans ces processus. Pour valider un cas d'adaptation locale, il est nécessaire de passer par des phénotypes. Il existe essentiellement deux types d'approche: les méthodes de cartographie de QTL et la génétique d'association.

#### *Les méthodes de cartographie de QTL*

Les méthodes de cartographie de QTL ont pour objectif de détecter les locus qui influencent la variation quantitative de caractères ou QTL, de les localiser avec un maximum de précision et d'estimer les effets de leurs différents allèles. Le principe consiste à lier des allèles en ségrégation à des variations phénotypiques dans une population issue de croisements contrôlés. Plusieurs types de populations peuvent être utilisés comme par exemple des populations F2 provenant du croisement de deux lignées pures (Lander & Botstein 1989). L'individu F1 ainsi obtenu est porteur d'un chromosome de chaque lignée, puis est autofécondé pour la production des F2 (Figure 14). Les événements de recombinaison entre les deux chromosomes permettent d'obtenir une génération F2 plurielle (Figure 14). Les analyses QTL peuvent alors porter soit directement sur les descendants F2 produits, soit sur les familles issues de l'autofécondation des individus F2 (familles F3), soit sur une population de lignées recombinantes (RIL) dérivées par plusieurs cycles d'autofécondation de chaque individu F2 (Figure 14). Ces cycles d'autofécondation sont réalisés par la méthode SSD (Single Seed Descend) jusqu'à disparition quasi-complète de l'hétérozygotie (Figure 14). La création d'haploïdes doublés, réalisée à partir de gamètes dont les chromosomes sont ensuite doublés artificiellement, est parfois possible. Cette méthode permet d'obtenir directement une population d'individus parfaitement homozygotes en ségrégation mais est assez lourde techniquement. L'analyse des populations biparentales présente cependant deux inconvénients majeurs. Le premier concerne la faible résolution physique obtenue. Cette résolution peut être améliorée en ajoutant quelques générations de croisements allogames entre individus de la F2 avant les cycles d'autofécondation. Les populations obtenues sont appelées AI-RIL (Advanced Intercross-Recombinant Inbred Lines, Figure 14). Cependant, les intervalles de confiance des QTL sont encore d'un ordre de grandeur beaucoup plus élevé que le gène. Le deuxième inconvénient est que seule une petite partie de la diversité génétique de l'espèce est capturée dans le dispositif, étant donné que, le plus souvent, seulement deux allèles ségrégent

à chaque locus dans la population. Pour pallier à cela, des dispositifs plus complexes ont été mis en place. Un des dispositifs les plus connus est le MAGIC (Multi-parent Advanced Generation InterCrosses), utilisé chez *Arabidopsis* (Kover *et al.* 2009) et de nombreuses espèces agronomiques (revue Huang *et al.* 2015 ; Figure 14). Les fondateurs des populations choisies sont croisés entre eux pendant plusieurs générations antérieurement à la création des RILs. Ceci a deux conséquences directes : la diversité génétique est plus élevée car plusieurs allèles peuvent ségréger pour le même QTL, et la résolution physique augmente avec le nombre de croisements allogames. A ce jour chez le maïs, deux dispositifs MAGIC ont été créés. Pea *et al.* (2013) ont utilisé 8 lignées fondatrices et obtenus 1300 RILs. Buet *et al.* (2013) ont croisés 16 lignées de maïs sur 3 générations avant de créer 543 lignées haploïdes-doublées. Quelles que soient les populations utilisées, les données sont ensuite traitées par des modèles plus ou moins complexes.



**Figure 14 : Création de différents types de populations en ségrégation contrôlée.** Figure de Bergelson and Roux (2010).

Dans tous ces types de populations en ségrégation, le pedigree est connu, et le déséquilibre de liaison entre deux marqueurs ou entre un marqueur et un QTL dépend uniquement de la distance génétique. Un des moyens les plus simples pour déterminer la position d'un QTL est de réaliser un test de comparaison de moyennes des phénotypes pour chacun des allèles/génotypes, marqueur par marqueur. La précision de localisation étant assez faible avec les méthodes de cartographie, des étapes supplémentaires peuvent être envisagées. La création des NIL (*Near Isogenic Lines*) obtenu après backcross successif du F2 et de ses descendants porteurs de l'allèle du QTL d'intérêt dans la lignée porteuse de l'autre allèle, ou des HIF (*Heterogeneous Inbred Lines*) qui exploitent les résidus d'hétérozygotie, permet une cartographie plus fine du QTL correspondant.

La méta-analyse de QTL est aussi utilisée pour combiner les informations d'études réalisées sur des dispositifs différents pour un même caractère. Ainsi, chez le maïs, une méta-analyse de QTL visant à étudier le déterminisme de la date de floraison a mis en évidence 62 régions consensus entre les études. Sur ces régions, seules deux sont liées à des gènes impliqués dans la transition florale. Cependant 19 associations ont été trouvées entre ces QTL et des gènes candidats impliqués dans la date de floraison, ce qui laisse supposer que beaucoup reste à découvrir (Chardon *et al.* 2004).

### *La génétique d'association*

La génétique d'association repose sur le même principe que la détection de QTL, mais utilise les événements de recombinaison accumulés sur de nombreuses générations. En effet, l'idée est de travailler sur des échantillons (panels) d'individus qui représentent une part importante de la diversité génétique d'une espèce. Le principe de la génétique d'association est de rechercher les associations statistiques entre la variation phénotypique d'un caractère et la ségrégation des allèles à des marqueurs moléculaires. L'utilisation des lignées (présentant peu de variabilité génétique intra-individuelle) permet une meilleure précision des évaluations phénotypiques grâce à la répétition des essais dans différents lieux/années et ainsi une distinction entre effets génétiques et environnementaux. Lors de l'utilisation de populations génétiquement hétérogènes, le contrôle environnemental sera moins optimal.

Originellement développée pour la génétique humaine et appliquée à l'échelle du génome (Genome-Wide Association Mapping, ou GWA), la génétique d'association permet une meilleure résolution que la cartographie de QTL à condition de disposer d'une densité de marqueurs importante. Cependant la structure des populations, et le déséquilibre de liaison qui

en découle entre les variants impliqués dans le déterminisme génétique du caractère et des locus non-liés, est un effet confondant important (Lander & Schork 1994). Les premières expériences de GWA chez les plantes se sont déroulées sur *Arabidopsis thaliana*, une plante autogame pour laquelle la structuration est particulièrement marquée, ce qui a permis de mettre en lumière l'importance de cet effet (Aranzana *et al.* 2005 ; Atwell *et al.* 2010).

Un choix judicieux des populations intégrées dans le dispositif permet de limiter ces effets et/ou de faciliter leur prise en compte par un modèle statistique approprié. Afin de minimiser les effets de structure, Bergelson & Roux (2010) préconisent l'emploi d'un panel régional, dont les populations proviennent de l'aire de distribution du facteur sélectif. Brachi *et al.* (2013) ont testé les mêmes variables sur des panels correspondant à des échelles géographiques différentes. Les régions détectées sont différentes bien que fortement soutenues statistiquement dans les deux cas, ce qui indique des mécanismes adaptatifs différents selon l'échelle spatiale considérée. D'une manière générale, l'utilisation d'un panel comportant des populations les moins structurées possibles et l'incorporation d'un grand nombre d'individus permet de limiter les effets confondants et donc les faux positifs (Rafalski 2010 ; Yu & Buckler 2006).

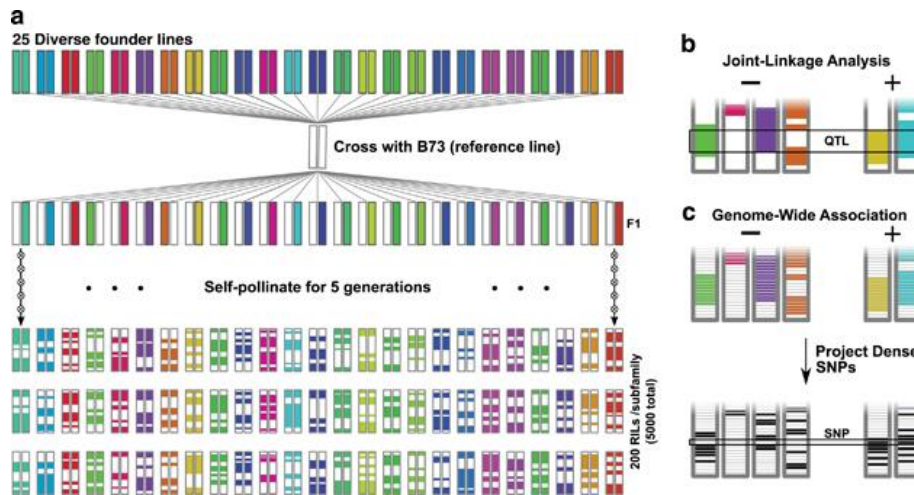
Les modèles statistiques utilisés en génétique d'association consistent à prendre en compte la structuration des populations et à chercher les associations statistiques entre marqueurs moléculaires et phénotype. Il existe différentes méthodes pour intégrer la structure au modèle d'association. Pritchard *et al.* (2000b) proposent d'utiliser une méthode d'assignement à des sous-groupes sans *a priori* en utilisant des marqueurs moléculaires non liés aux phénotypes d'intérêt, telle qu'effectuée par le logiciel STRUCTURE (Pritchard *et al.* 2000a). Price *et al.* (2006) ont recours à une analyse en composantes principales sur le même type de marqueurs moléculaires pour estimer et corriger les effets dus à la variation ancestrale partagée. Les associations entre phénotype et génotype sont ensuite testées en prenant en compte ces sous-groupes. Yu *et al.* (2006) proposent un modèle mixte permettant de prendre en compte différents niveaux de structuration. La structure des populations est intégrée au modèle sous la forme d'un effet fixe, alors que l'apparentement entre individus (matrice d'apparentement) est intégré comme effet aléatoire. Zhao *et al.* (2007) ont testé ces trois méthodes sur un jeu de données comprenant 95 accessions d'*Arabidopsis thaliana*, une plante présentant un haut niveau de structuration. Les auteurs ont trouvé que la méthode proposée par Yu *et al.* (2006) était la plus performante des trois citées ci-dessus. Ils proposent éventuellement de remplacer la structure des populations par les premiers axes de l'ACP sur



les marqueurs moléculaires. Ces méthodes permettent de réduire le nombre de faux positifs mais réduisent également la puissance de détection (Yu *et al.* 2006 ; Zhao *et al.* 2007). En effet, lorsque la variation phénotypique d'un caractère est concordante avec la structure des populations, comme par exemple pour la date de floraison, le fait de prendre en compte la structure dans les modèles statistiques peut aussi effacer une partie du signal biologique que l'on cherche à mettre en évidence (Flint-Garcia *et al.* 2003; Yu & Buckler 2006).

Dans une revue sur l'identification de gènes impliqués dans des caractères adaptatifs, Bergelson & Roux (2010) insistent sur la nécessité de replacer les études de génétique d'association dans un contexte écologique. Ils soulignent notamment l'importance de réaliser les essais dans des conditions naturelles. Dans une étude de génétique d'association sur la date de floraison chez *Arabidopsis* en milieu naturel, les gènes candidats répertoriés sont en grande partie différents de ceux isolés en conditions contrôlées (Brachi *et al.* 2010).

Ces dernières années, de nouveaux dispositifs permettant de combiner les avantages de la détection de QTL et de la génétique d'association ont été proposés sous le nom de *Nested Association Mapping* (NAM). Ce type de dispositif en étoile a été développé en premier chez le maïs, où 25 lignées pures ont été croisées avec la lignée B73 qui sert de lignée pivot (Yu *et al.* 2008). Ces croisements forment des populations biparentales de détection de QTL. L'analyse conjointe des différentes familles partageant un parent commun permet de localiser les QTL en s'appuyant sur le fort déséquilibre de liaison au sein des familles et les fortes fréquences alléliques (Figure 15 b) tandis que le faible déséquilibre de liaison entre les familles et la diversité des haplotypes confèrent une puissance élevée aux analyses de génétique d'association pour localiser avec une meilleure résolution le locus impliqué (Figure 15 c) (Wallace *et al.* 2014).



**Figure 15 : Dispositif NAM créé par Yu *et al.* (2008).** (a) 25 lignées ont été croisées avec B73. Les F1 ont ensuite été auto-fécondées pour créer 25 populations biparentales ayant B73 en parent commun et représentant un large panel de diversité. (b) Le fort déséquilibre de liaison au sein de familles et les fortes fréquences alléliques permettent de cartographier les QTL. (c) Le faible déséquilibre de liaison entre les différentes familles et la diversité haplotypique permettent ensuite de réaliser des études de génétique d'association. Figure de Wallace *et al.* (2014).

### I.1.3.g Etudes *Qst/Fst*

L'approche *Qst/Fst* mêle génétique des populations et données phénotypiques. Ces approches évaluent la variation d'un caractère quantitatif en fonction de la différenciation génétique neutre. Ceci permet d'évaluer l'influence de la sélection naturelle par rapport à la dérive génétique (Merila & Crnokrak 2001). Dans le cas où les *Qst* dévient des *Fst*, la dérive ne suffit plus à expliquer la valeur du *Qst* et un facteur sélectif est invoqué (revue Leinonen *et al.* 2013).

## I.2 ZEA MAYS COMME MODELE BIOLOGIQUE

### I.2.1 Système *Zea mays*

Le maïs cultivé (*Zea mays* ssp. *mays*) est le taxon le plus connu du genre *Zea*, qui comporte d'autres espèces et sous-espèces dénommées téosintes. Les espèces *Zea perennis*, *Zea diploperennis*, *Zea luxurians* et *Zea nicaraguensis* ont émergé récemment, il y a 100 000 à 300 000 ans (Ross-Ibarra *et al.* 2009). Elles présentent des cycles de vie et des aires de répartitions différentes. La question de l'origine du maïs semble avoir été élucidée avec l'accumulation d'éléments archéologiques et génétiques en faveur d'un unique événement de domestication à partir de la sous-espèce *Zea mays* ssp. *parviglumis* il y a 9 000 ans environ dans la région de la rivière Balsas au Mexique (Matsuoka *et al.* 2002 ; Piperno *et al.* 2009 ; van Heerwaarden *et al.* 2011). Le maïs se serait ensuite rapidement propagé à travers

l'Amérique centrale (Tenailon & Charcosset 2011).

Les téosintes les plus apparentées au maïs et aujourd'hui classées dans la même espèce sont les sous-espèces *Zea mays* ssp. *parviglumis*, *Zea mays* ssp. *mexicana* et *Zea mays* ssp. *huehuetenangensis* (Iltis & Doebley 1980). Elles sont annuelles. *Zea mays* ssp. *parviglumis* (désignée ici comme *parviglumis*) et *Zea mays* ssp. *mexicana* (désignée ici comme *mexicana*) poussent au Mexique et se ressemblent très fortement au point qu'il est difficile de les distinguer l'une de l'autre même pour un œil connaisseur. Leur date de divergence est récente, estimée à environ 60 000 ans (Ross-Ibarra *et al.* 2009). Elles présentent cependant des aires de répartition distinctes (Buckler *et al.* 2006 ; Hufford *et al.* 2012a, Figure 16). La sous-espèce *mexicana* a colonisé les régions du Mexique situées entre 1 600 mètres et 2 700 mètres d'altitude avec une altitude moyenne d'environ 2 100 mètres, tandis que *parviglumis* pousse à des altitudes inférieures à 1 800 mètres avec une altitude moyenne de 1 050 mètres (Hufford *et al.* 2012a), (Figure 16). Ces différences d'altitude se traduisent par des variations climatiques prononcées. *parviglumis*, qui nécessite une température plus stable le long de l'année et des précipitations plus importantes que *mexicana*, se trouve dans des régions où la saisonnalité est plus marquée, et où la température annuelle moyenne et les précipitations sont moins élevées (Hufford *et al.* 2012a). Ainsi, *parviglumis* se trouve essentiellement dans des climats tropicaux, alors que *mexicana* pousse le plus souvent dans des environnements semi-tempérés arides. Ces deux sous-espèces de téosintes possèdent un effectif efficace important (de l'ordre de 120 000 à 160 000 ; Ross-Ibarra *et al.* 2009) ainsi qu'une diversité génétique élevée (Fukunaga *et al.* 2005 ; Hufford *et al.* 2012a). Les populations de téosintes présentent une structuration spatiale (Fukunaga *et al.* 2005). Un premier niveau de structuration sépare les deux sous-espèces *mexicana* et *parviglumis*, qui présentent à leur tour de la structuration au niveau des populations. Les téosintes *parviglumis* et *mexicana* présentent une démographie complexe qui complique l'identification d'allèles localement adaptés. Depuis leur divergence il y a 60 000 ans, les deux sous-espèces ont connu une phase d'expansion (Buckler *et al.* 2006; Ross-Ibarra *et al.* 2009). La sous-espèce *mexicana* dériverait de *parviglumis*, et aurait colonisé les hauts-plateaux mexicains grâce à des adaptations à des environnements plus froids (Buckler *et al.* 2006). Les deux sous-espèces présentent une structure hiérarchisée. L'étude de Pyhäjärvi *et al.* (2013) à partir de données de génotypage de la puce 50K du maïs a permis de mieux comprendre cette structuration. Un premier niveau de structuration correspond aux deux sous-espèces ( $F_{st}$  global de 0,33), mais il existe également une structuration à l'intérieur de chaque sous-espèce ( $F_{st}$  de 0,24 pour *parviglumis* et  $F_{st}$  de 0,23 pour *mexicana*). Cette structuration est plus forte chez *mexicana* que chez *parviglumis* (Ross-

Ibarra *et al.* 2009), et l'isolement par la distance en explique une partie (Pyhäjärvi *et al.* 2013). Il est intéressant de noter que le niveau de structuration varie chez *parviglumis*. Les populations présentes dans la région de Jalisco sont plus structurées que celles provenant d'autres régions (Moeller *et al.* 2007, Jonas Aguirre communication personnelle). Ce patron de structuration peut être expliqué par le fait que les populations de Jalisco dériveraient des populations de la région de Balsas qui constituait un refuge lors de la dernière glaciation (Buckler *et al.* 2006). La colonisation de la région de Jalisco se serait accompagnée de forts effets démographiques (Fukunaga *et al.* 2005 ; Hufford *et al.* 2012a ; Moeller *et al.* 2007). Moeller *et al.* (2007) montrent que la structuration de *parviglumis* entraîne une déviation de spectre de fréquence neutre, et peut produire des signaux similaires à ceux attendus sous sélection balancée, c'est-à-dire un déficit de variants rares. Ils montrent également qu'il est important d'étudier l'espèce à l'échelle des populations et d'éviter les échantillonnages d'un individu par population sur l'ère de répartition de l'espèce. Bien que *parviglumis* et *mexicana* soient des plantes allogames, elles montrent des niveaux de structure relativement importants, peut-être dus à au fait qu'elles poussent dans un relief montagneux. Dans ce contexte, les flux de gènes peuvent être limités par la topographie de l'environnement, comme cela a été montré à faible échelle chez *parviglumis* (Van Heerwaarden *et al.* 2010).

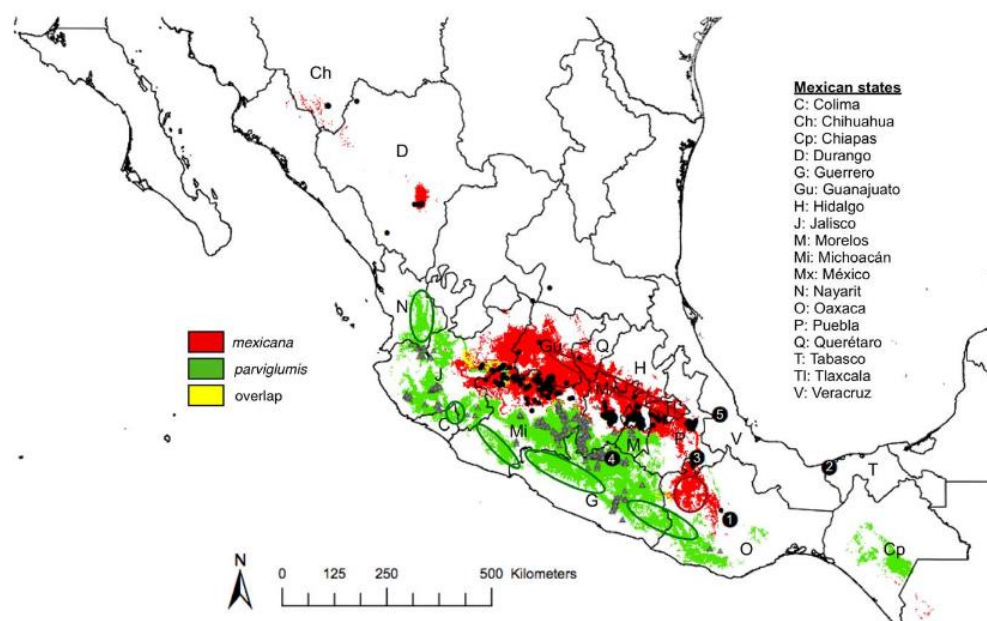
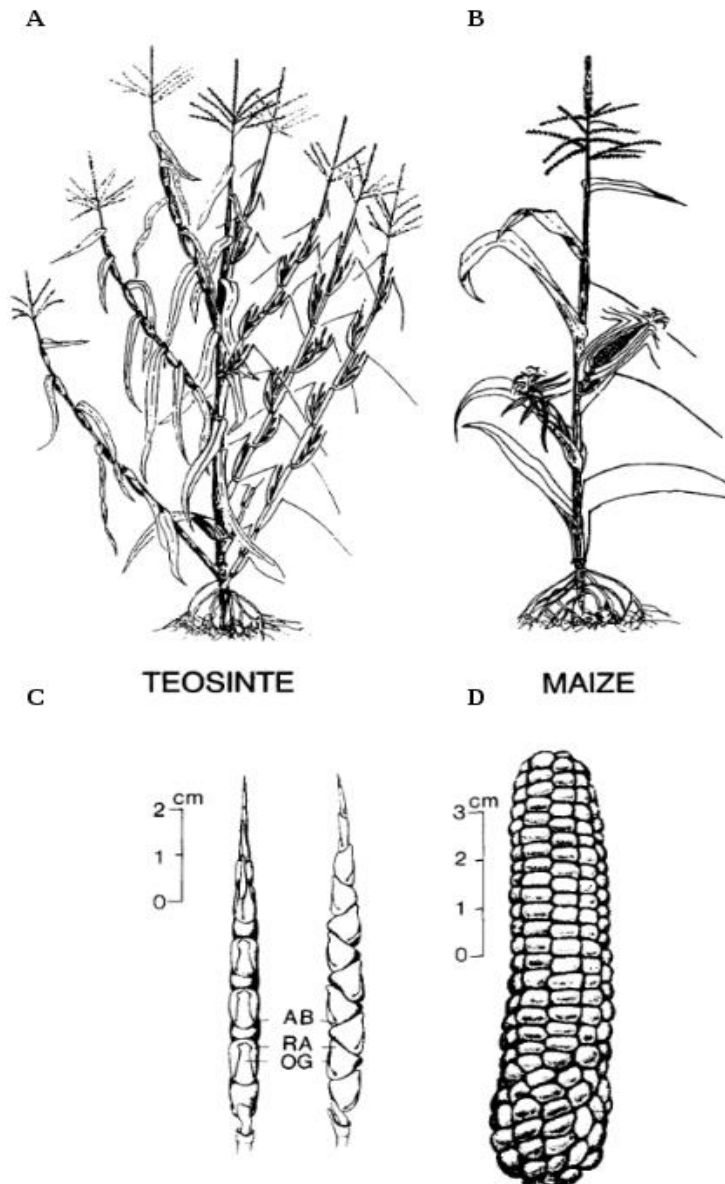


Figure 16 : Distribution des aires de répartition des téosintes *Zea mays mexicana* et *Zea mays parviglumis* prédites par un modèle de niche. Figure de Hufford *et al.* (2012a).

Le maïs présente quant à lui une diversité génétique plus faible que ses apparentés sauvages (environ 83% de la diversité trouvée chez les téosintes), ayant subi une sélection forte depuis sa domestication et au moins deux goulots d'étranglement : un ayant accompagné la domestication, et un ayant accompagné l'amélioration variétale plus récente (Yamasaki *et al.* 2005). Des études suggèrent que les flux de gènes entre *parviglumis* et le maïs se sont interrompus, contrairement aux flux de gènes entre *parviglumis* et *mexicana* d'une part, et entre *mexicana* et le maïs d'autre part (Ross-Ibarra *et al.* 2009 ; Wilkes 1977). Ainsi, des résultats ont montré jusqu'à 20% d'introgression entre les maïs d'altitude (highland) et *mexicana*, alors qu'il y en a au maximum 3% entre le maïs et *parviglumis* (van Heerwaarden *et al.* 2011). Les flux de gènes de *mexicana* vers le maïs auraient largement contribué à l'adaptation du maïs aux hautes altitudes (van Heerwaarden *et al.* 2011 ; Takuno *et al.* 2015). D'une manière générale, les flux de gènes sont bien plus forts des téosintes vers le maïs qu'inversement (Baltazar *et al.* 2005). Le taux de fécondation réussie est plus important chez les croisements impliquant des femelles maïs de près d'un facteur 10, et les F1 des téosintes pollinisées par du pollen de maïs sont le plus souvent stériles, au contraire des maïs pollinisés par des téosintes (Baltazar *et al.* 2005). De plus, les téosintes produisent plus de pollen que les maïs et leur soies sont moins persistantes dans le temps que celles des maïs, ce qui contribue à cette asymétrie en laissant ainsi moins de possibilité de fécondation si les deux sous-espèces ne fleurissent pas de manière synchrone cette asymétrie (Baltazar *et al.* 2005).

Les différences morphologiques entre téosintes et maïs sont particulièrement prononcées, que ce soit au niveau de l'inflorescence, des grains, ou de l'architecture générale de la plante (Figure 17, Doebley 1992). Les maïs présentent généralement une seule talle, tandis que les téosintes ont généralement une allure buissonnante, bien que certaines puissent aussi présenter une unique talle (Figure 17-A, B). Les talles portent de nombreuses branches latérales chez la téosinte qui se terminent par une inflorescence mâle. Chez le maïs, seul l'axe principal se termine par une inflorescence mâle, les branches latérales, courtes et peu nombreuses, sont porteuses d'un épi femelle. Les inflorescences femelles sont nombreuses chez la téosinte et situées à la fois sur les talles et les branches secondaires (Figure 17-A, B). Les différences les plus marquantes se situent au niveau des épis (Figure 17-C, D). Chez la téosinte, les épis sont petits, porteurs de moins d'une dizaine de grains disposés en quinconce, alors que les épis de maïs sont gros, porteurs de plus d'une centaine à quelques centaines de grains pour les plus gros, disposés en rangées le long d'un rachis (Figure 17-C, D). Les grains sont caducs chez la téosinte tandis qu'ils restent fixés au rachis de l'épi chez le maïs. Chez la téosinte, ils ont une morphologie triangulaire, sont entourés de glumes extrêmement indurées

dont la coloration peut prendre un grand nombre de nuance dans les tons brun-noir-vert (Figure 17-C). Chez le maïs, les grains sont plus grands, présentent une forme plus arrondie. Les glumes sont réduites, non indurées, et n'enserrent pas le grain. Le péricarpe peut prendre des colorations extrêmement variables, allant dans les blancs, les jaunes, les orangés, les rouges, les violets, les verts ou bien les bleus.



**Figure 17 : Différence morphologiques entre la téosinte (A, C) et le maïs (B, D).** (A, B) Architecture de la plante. (C, D) Organisation de l'épi. Adapté de Doebley (1992).

Les téosintes *parviglumis* et *mexicana* présentent une démographie complexe qui complique l'identification d'allèles localement adaptés. Depuis leur divergence il y a 60 000

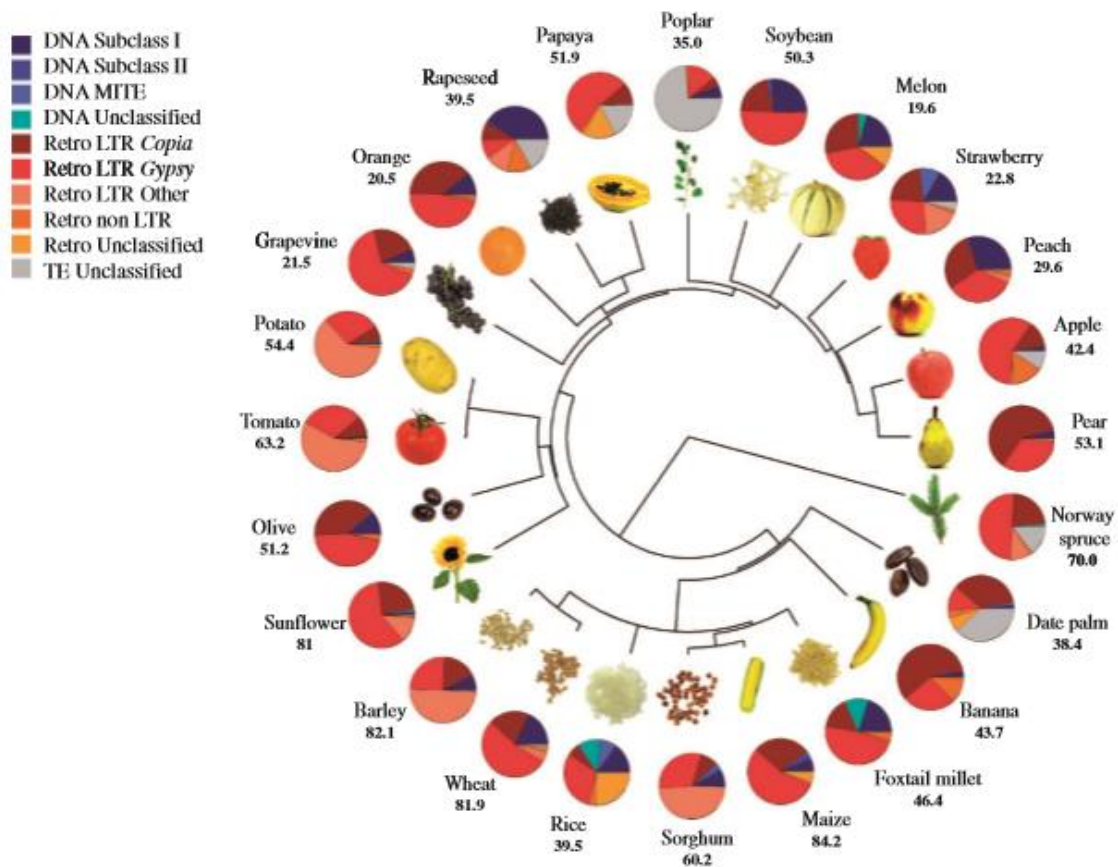
ans, les deux sous-espèces ont connu une phase d'expansion (Buckler *et al.* 2006; Ross-Ibarra *et al.* 2009). La sous-espèce *mexicana* dérive de *parviglumis*, et aurait colonisé les hauts-plateaux mexicains grâce à des adaptations à des environnements plus froids (Buckler *et al.* 2006). Les deux sous-espèces présentent une structure hiérarchisée. L'étude de Pyhäjärvi *et al.* (2013) à partir de données de génotypage de la puce 50K du maïs a permis de mieux comprendre cette structuration. Un premier niveau de structuration correspond aux deux sous-espèces (*Fst* global de 0,33), mais il existe également une structuration à l'intérieur de chaque sous-espèce (*Fst* de 0,24 pour *parviglumis* et *Fst* de 0,23 pour *mexicana*). Cette structuration est plus forte chez *mexicana* que chez *parviglumis* (Ross-Ibarra *et al.* 2009), et l'isolement par la distance en explique une partie (Pyhäjärvi *et al.* 2013). Il est intéressant de noter que le niveau de structuration varie chez *parviglumis*. Les populations présentes dans la région de Jalisco sont plus structurées que celles provenant d'autres régions (Moeller *et al.* 2007, Jonas Aguirre communication personnelle). Ce patron de structuration peut être expliqué par le fait que les populations de Jalisco dériveraient des populations de la région de Balsas qui constituait un refuge lors de la dernière glaciation. La colonisation de la région de Jalisco se serait accompagnée de forts effets démographiques (Fukunaga *et al.* 2005 ; Hufford *et al.* 2012a ; Moeller *et al.* 2007). Bien que *parviglumis* et *mexicana* soient des plantes allogames, elles montrent des niveaux de structure relativement importants, peut-être dus à au fait qu'elles poussent dans un relief montagneux. Dans ce contexte, les flux de gènes peuvent être influencés par la topographie de l'environnement, comme montré à faible échelle chez *parviglumis* (Van Heerwaarden *et al.* 2010).

## **I.2.2 Génome de *Zea mays***

### **I.2.2.a Génome du maïs**

Le maïs est un ancien allopolyploïde. Il a subi deux événements de duplication totale du génome, le premier il y a environ 70 millions d'années (Paterson *et al.* 2004) et le second il y a entre 5 et 12 millions d'années (Swigonova *et al.* 2004). Le retour à la diploïdie a mené à d'importants remaniements chromosomiques dont un nombre important d'inversions (Fei *et al.* 2007). Le génome de la lignée B73 est de 2,3 Gb et contient 32 000 gènes prédits, dont 99,8 % ont pu être localisés sur un des 10 chromosomes (Schnable *et al.* 2009). Presque 85% du génome est composé d'éléments transposables provenant de centaines de familles différentes (Schnable *et al.* 2009). Parmi ceux-ci, les rétrotransposons occupent 75% du génome du maïs (Figure 18, Schnable *et al.* 2009 ; Vitte *et al.* 2014). Leur amplification au

cours des trois derniers millions d'années a mené au doublement de la taille du génome de *Zea mays* (SanMiguel *et al.* 1998). Les éléments transposables ont également un impact sur la taille des gènes : bien que les exons soient sensiblement de même longueur que dans les orthologues du riz et du sorgho, les introns sont plus longs chez le maïs du fait de l'insertion de séquences répétées (Schnable *et al.* 2009).



**Figure 18 : Proportion des différents éléments transposables dans le génome des espèces cultivées dont le génome de référence est disponible.** Les relations phylogéniques sont montrées. Les nombres en dessous des noms indiquent la proportion d'éléments transposable dans les génomes respectifs. Les camemberts indiquent les proportions des différentes classes d'éléments transposables. Figure de Vitte *et al.* (2014).

Chez les maïs, la taille du génome est extrêmement variable, pouvant aller jusqu'à un coefficient de variation de 30% (Muñoz Diez *et al.* 2013). Muñoz Diez *et al.* (2013) ont confirmé cette variation dans l'ensemble des sous-espèces de *Zea mays*. Ils ont mesuré la taille du génome de plusieurs individus de 22 variétés traditionnelles et 21 téosintes (*parviglumis* et *mexicana*) et ont montré un fort effet de la population d'origine. La variation de taille entre individus de l'espèce *Zea mays* n'est cependant pas due aux éléments



transposables mais à la présence de knob (séquences hétérochromatiques répétées) (Chia *et al.* 2012 ; Muñoz Diez *et al.* 2013). Bennett (1987) propose une vision adaptative de la taille du génome, en avançant une relation entre taille du génome et durée du cycle de vie. Les résultats d'une étude de sélection sur l'avancée de la date de floraison semblent appuyer cette hypothèse : Rayburn *et al.* (1994) trouvent une relation entre avancement de la floraison et diminution de la taille du génome chez le maïs. Sur une étude comportant des variétés traditionnelles de maïs et des téosintes, Muñoz Diez *et al.* (2013) soutiennent le caractère adaptatif de la taille du génome, mais pour des facteurs sélectifs différents entre maïs (facteurs liés à l'altitude) et téosinte (facteurs liés à des variables climatiques).

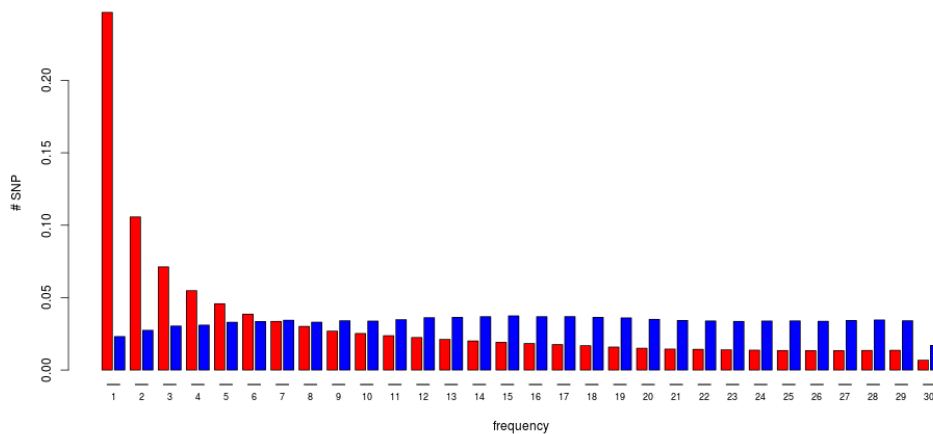
### **I.2.2.b Outils de géotypage disponibles**

Le maïs est une espèce céréalière d'intérêt économique majeur. Les organismes de recherche ainsi que l'industrie agroalimentaire ont investi beaucoup d'argent afin de mieux connaître son génome. Une grande partie de ces efforts a porté sur le séquençage de son génome, et l'annotation des gènes. Le génome de référence du maïs est celui de B73, une lignée élite dérivée de la population synthétique des Iowa Stiff Stalks qui a été largement utilisée dans la construction d'hybrides. La première version, sortie en 2009 (Schnable *et al.* 2009), continue d'être améliorée (actuellement la version 4 est en cours, [www.maizegdb.org/assembly](http://www.maizegdb.org/assembly)).

D'autres projets ont permis d'appréhender l'importante diversité génétique du maïs, comme les projets HapMap, dans lequel un grand nombre de lignées de maïs et de téosintes a été séquençé (Chia *et al.* 2012 ; Gore *et al.* 2009). Ces projets permettent de mieux capter l'incroyable diversité nucléotidique (10 à 20 fois supérieure à celle observée chez l'homme) et structurale (Chia *et al.* 2012) de *Zea mays*.

L'existence d'un génome de référence facilite drastiquement les études de génétique. Plusieurs puces à ADN ont été développées afin de capter des variations de SNP à moindre coût. Chez le maïs, deux puces ont été développées. La puce MaizeSNP50 d'Illumina (Ganal *et al.* 2011) permet de géotyper plus de 50 000 SNP. Plus récemment, une puce comportant plus de 600 000 positions de SNP a été commercialisée sous le nom de Affymetrix® Axiom® Maize Genotyping Array (Unterseer *et al.* 2014). Ces puces permettent de géotyper un grand nombre de SNP répartis de manière quasi-uniforme. Les biais d'échantillonnage liés à la construction des puces mentionnés en section I.1.3.a sont particulièrement importants chez le maïs (Figure 19). En effet, la puce 50K a été dessinée à partir d'un panel de 274 individus, dont les géotypes B73 et Mo17, des lignées parents et des F1 du dispositif NAM, ainsi que

des lignées de téosinte. L'incorporation de nombreuses lignées permet d'atténuer l'*ascertainment bias*, qui reste cependant très fort, en particulier parce que la moitié des SNP de la puce a été découvert par comparaison des séquences de B73 et Mo17 (Ganal *et al.* 2011). Les répercussions de cet échantillonnage sur les spectres de fréquences alléliques sont importantes et peuvent entraver la détection de polymorphismes adaptatifs. La puce 600K a été conçue comme une amélioration de la puce 50K, autant par le nombre de SNP génotypés qu'au niveau de l'*ascertainment bias*. Elle a été construite à partir de 57 millions de SNP découverts à partir du séquençage de 30 lignées de maïs tempérés, dont 1,2 millions ont été pré-sélectionnés selon des critères de qualité stringents. Le génotypage de ces 1,2 millions de SNP sur 285 échantillons d'origine diverse (hybride, lignées pures, tropicales...) a permis de sélectionner 616 201 variants (Unterseer *et al.* 2014).

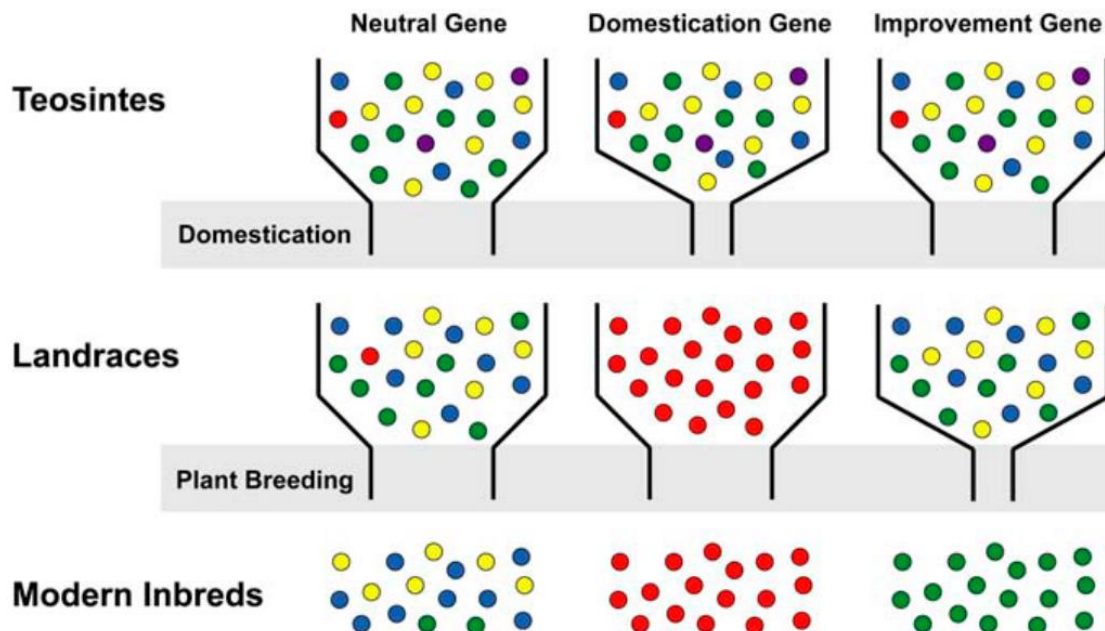


**Figure 19 :** Comparaison des SFS pliés de 67 lignées échantillonnées dans les groupes génétiques des Dentés du Corn Belt (14), Cornés du nord (13), Cornés Européen (13), Italiens/Espagnols (8) et Tropicaux (19) à partir de 32 338 positions de la puce 50 K (Bleue) et 10 264 202 positions obtenues sur des données SNP de re-séquencage (Rouge) à partir des mêmes lignées (Brandenburg *et al.* en préparation)

### I.2.2.c Identification de gènes ou régions d'intérêt chez le maïs

L'approche classique dite de bottom-up (du génotype vers le phénotype), employée dans les scans génomiques (cf section I.1.3.e) pour identifier des régions impliquées dans la domestication/amélioration, consiste à comparer les patrons d'évolution moléculaire entre des échantillons de téosintes *versus* des maïs *et/ou* des populations traditionnelles *versus* des lignées élites. On entend ici par populations traditionnelles des cultivars locaux soumis à une sélection massale douce qui sont supposément plus proches des premiers maïs domestiques que ne le sont les lignées de cycle avancé qui ont subi une sélection récente et intense (amélioration). Partant de ce principe, Yamasaki *et al.* (2005) proposent de comparer des

lignées à des variétés populationnelles de maïs et des téosintes. On s'attend à ce que les gènes de domestication soient fixés chez les lignées et les variétés traditionnelles par rapport à la variabilité existante chez les téosintes, alors qu'une baisse de variabilité des gènes d'amélioration au niveau des lignées est attendue (Figure 20). Par cette méthode, les auteurs ont mis en évidence 8 gènes de domestication et/ou d'amélioration sur 35 candidats. Dans une étude ultérieure, Wright *et al.* (2005) ont analysé le polymorphisme de 774 gènes chez 14 lignées pures de maïs et 16 lignées pures de téosintes, et ont trouvé que 2 à 4% des gènes du maïs présentent des traces de sélection cohérentes avec un goulot d'étranglement durant la domestication. Enfin, plus récemment, Hufford *et al.* (2012b) ont analysé des données de reséquençage de 35 lignées élités et 23 cultivars traditionnels de maïs ainsi que 17 populations de téosintes produites par le projet HaMap2 (Chia *et al.* 2012). Ces auteurs ont mis en évidence un épisode de sélection plus fort pendant la domestication que pendant l'amélioration, et ont montré que 23% des gènes de domestication ont aussi été impliqués dans l'amélioration du maïs. Ces études permettent d'établir des listes de gènes ou régions génomiques candidats, mais le lien avec le phénotype n'est pas établi, d'où la difficulté de procéder ensuite à une validation fonctionnelle, et d'aller donc plus loin dans la caractérisation des régions identifiées.



**Figure 20 : Impact de la domestication et de l'amélioration sur le polymorphisme.** Les gènes neutres ne sont que modérément affectés par le goulot d'étranglement lors de la domestication et de l'amélioration. Les gènes de domestication ont un polymorphisme important dans le compartiment téosinte mais présentent une chute drastique de cette diversité dans les lignées traditionnelles. Les gènes d'amélioration présentent une baisse importante de polymorphisme entre le compartiment des lignées traditionnelles et celui des lignées élités. Figure de Yamasaki *et al.* (2005)

Un autre grand type d'approche est celle dite de top-down, du phénotype vers le génotype. L'idée est de s'intéresser aux bases génétiques d'un phénotype caractérisé. Ce type d'approche regroupe les études de QTL et les études GWA réalisées notamment sur le dispositif NAM. De nombreux caractères d'intérêt agronomique sont déterminés par un grand nombre de locus à faibles effets. Le dispositif NAM a notamment permis de détecter des QTL liés à la précocité de floraison (Buckler *et al.* 2009), la photopériode (Hung *et al.* 2012), l'architecture de la feuille (Tian *et al.* 2011) et la composition des grains (Whitt *et al.* 2002 ; Cook *et al.* 2012).

Les différences morphologiques entre le maïs et la téosinte sont frappantes (Figure 17). En effectuant un nombre conséquent de croisements entre maïs et téosinte, Beadle a montré dans les années 1970 que cinq régions du génome sont suffisantes pour effectuer la majorité des transitions morphologiques entre les deux sous-espèces. Des gènes de domestication responsables de profondes modifications de l'architecture de la plante ont par la suite été identifiés comme le gène *Tb1* (*teosinte-branched-1*), responsable de la dominance apicale du maïs avec un phénotype monotalle à la différence des téosintes (Doebley *et al.* 1997). D'autres gènes, comme le gène *Tgal* (*teosinte glume architecture 1*) conférant un phénotype « nu » aux grains (Wang *et al.* 2005) et le gène *Bal* (*barren stalk 1*) intervenant dans la formation des méristèmes végétatifs (Doebley *et al.* 1995 ; Gallavotti, 2004) ont également été primordiaux dans la domestication du maïs.

Certaines approches se sont concentrées sur des gènes candidats dont les mutants ont un effet phénotypique majeur ou dont les orthologues ont été identifiés chez d'autres espèces domestiquées. C'est par exemple le cas du gène *Ral* (*ramosal*) qui contrôle l'architecture des floraisons mâles et femelles (Sigmon & Vollbrecht 2010). Briggs *et al.* (2007) ont aussi identifié, dans un dispositif impliquant des hybrides téosintes-maïs, plusieurs QTL impliqués, entre autres, dans l'architecture de l'inflorescence dont *Zea agamous Zag2* et *Tasselseed4 Ts4*. Le gène *Gtl* (*grassy tiller 1*) intervient dans le contrôle de la production des talles et des branches secondaires avec *Tb1* (Whipple *et al.* 2011). Des travaux ont été consacrés à l'identification de locus affectant le temps de floraison chez le maïs, et ont conduit à l'identification de plusieurs gènes dont *ZmCCT* et *Vgt1* responsables d'une part importante de la précocité de floraison et contribuant à l'adaptation des maïs aux climats tempérés (Buckler *et al.* 2009 ; Hung *et al.* 2012 ; Salvi *et al.* 2007). Une étude de génétique d'association sur la date de floraison a également permis de mettre en évidence de nombreux allèles à faible effet, et a montré que les QTL impliqués variaient selon les lignées considérées (Buckler *et al.*

2009). Des travaux se sont également intéressés à la composition et l'aspect des grains et ont mené à l'identification des gènes *Bt2* (*brittle endosperme2*) (Whitt *et al.* 2002) ; *C1* (*colored aleurone1*) et *R1* (*colored1*), (Hanson *et al.* 1996) ; *Sh2* (*shrunken2*), (Bhave *et al.* 1990) ; *Su1* (*sugary1*) (Tracy *et al.* 2006 ; Whitt *et al.* 2002) ; *Ae1* (*amylose extender*) (Whitt *et al.* 2002) ; *Y1* (*yellow endosperm1*), (Coque & Gallais 2008 ; Palaisa *et al.* 2003) ; *Lyce1* (*lycopene epsilon cyclasse1*) (Harjes *et al.* 2008) ; *O2* (*opaque2*) et *CyPPDK* (*cytosilic pyruvate orthophosphate dikinase*) (Manicacci *et al.* 2009) ; et *Frk2* (*fructokinase2*) (Coque & Gallais 2008).

Certaines études ont été dédiées à la recherche de QTL pouvant agir dans des conditions extrêmement humides (Qiu *et al.* 2007 ; Malosetti *et al.* 2008). Au contraire, Agrama & Moussa (1996) ont mis en évidence un certain nombre de QTL impliqués dans la tolérance à la sécheresse. Plusieurs études ont porté sur la recherche de QTL permettant à la plante d'être cultivée dans des conditions relativement peu favorables. Par exemple, Reiter *et al.* (1991) se sont concentrés à la découverte de QTL de résistance à un déficit en éléments phosphatés, alors que Ninamango-Cárdenas *et al.* (2003) et Sibov *et al.* (1999) ont identifié des QTL de tolérance à l'aluminium, comme *Alm1* et *Alm2*. De très nombreuses études ont porté sur la recherche de QTL de résistance aux virus, comme les locus *maize streak virus tolerance1* (Kyetere *et al.* 1999; McMullen *et al.* 1994; Welz *et al.* 1999), *wheat streak mosaic virus resistance2* (McMullen *et al.* 1994 ; Xia *et al.* 1999), *resistance to maize mosaic virus I* (*corn stripe1*) (Ming *et al.* 1997) et *heat streak mosaic virus resistance1* et *heat streak mosaic virus resistance3* (McMullen & Louie 1989).

### **I.2.3 Adaptation locale chez les téosintes**

Le maïs a divergé de la téosinte il y a seulement 9 000 ans. La diversité du génome de la téosinte qui pousse dans des milieux contrastés a facilité la dispersion du maïs et sa culture dans des environnements variés. Lors de sa domestication le maïs a subi un goulot d'étranglement qui lui a fait perdre une partie de la diversité originelle des téosintes (Yamasaki *et al.* 2005). L'amélioration des lignées modernes a conduit également à une perte de diversité (Yamasaki *et al.* 2005). Cette amélioration s'est concentrée sur des génotypes poussant dans des conditions riches en intrants et avec une irrigation soutenue. Cette diversité perdue est cependant encore présente chez la téosinte qui constitue une réserve intéressante d'allèles qui pourraient être récupérés et intégrés au maïs dans le cadre d'une agriculture durable. Un premier cas vient d'être reporté dans la littérature où des allèles provenant de *parviglumis* et intégrés au maïs ont permis de conférer à ce dernier une résistance à la

cercosporiose du maïs (Lennon *et al.* 2016). La connaissance approfondie de l'adaptation locale chez les téosintes, aussi bien au niveau des caractères morphologiques que des bases génétiques, est indispensable pour exploiter toutes les possibilités qu'offre cet apparenté sauvage.

Les téosintes montrent des différences phénotypiques selon l'altitude de la population dont elles proviennent. Les téosintes de haute altitude présentent généralement une architecture de type maïs, avec un (ou peu de) talle(s). Les gaines des feuilles peuvent avoir une coloration violette, et les structures végétatives sont recouvertes de nombreux poils. Au contraire, les populations de basse altitude ont tendance à présenter une architecture typiquement buissonnante, les gaines des feuilles sont vertes et la plante est globalement glabre (Lauter *et al.* 2004 ; Takuno *et al.* 2015).

Une seule étude de grande ampleur a été réalisée sur l'adaptation locale des téosintes. Pyhäjärvi *et al.* (2013) ont analysé 21 populations de téosintes sur des données de la puce 50K. Ils ont réalisé des études de structure et cherché des SNP sous sélection. L'analyse des fréquences alléliques couplée à des méthodes corrélatives a permis de mettre en évidence respectivement 1 363 et 1 598 SNP candidats. Cette étude soulève le rôle des SNP non-géniques dans l'adaptation de la téosinte. Certains de ces SNP sont associés à l'architecture des inflorescences mâle et femelle et à la précocité de floraison. Certains gènes montrent également de forts signaux de sélection : *B1*, un gène appartenant à la voie de synthèse des anthocyanines, *Sh11* qui contrôle la synthèse de sucrose et *Tb1*, le gène de domestication contrôlant la dominance apicale. Les auteurs ont aussi appuyé le rôle adaptatif de quatre inversions déjà identifiées (Hufford *et al.* 2012a ; Fang *et al.* 2012). Une inversion d'environ 50 Mb située sur le chromosome 1 (*Inv1n*) et comportant plus de 700 gènes a été particulièrement étudiée (Hufford *et al.* 2012b ; Fang *et al.* 2012). *Inv1-I* présente un fort cline altitudinal, est statistiquement associée avec l'altitude, la température et la précipitation, et présente un déficit en variants haplotypiques rares (Fang *et al.* 2012). Cette inversion serait aussi associée à des variables phénotypiques notamment au niveau des appareils floraux. *Inv1-I* est très présente chez *parviglumis* (de 18 à 90%), beaucoup moins chez *mexicana* (10%) et absente chez le maïs cultivé (Fang *et al.* 2012). L'incorporation d'autres espèces et sous-espèces du genre *Zea* et du genre *Trispsacum* suggère que l'inversion est antérieure à l'apparition de l'espèce *Zea mays*, et qu'elle a été perdue chez le maïs cultivé (Fang *et al.* 2012). Cette inversion est donc un bon candidat à l'adaptation.

Un certain nombre d'études a adopté une approche « top-down » et a cherché à

découvrir et localiser les gènes impliqués dans le déterminisme génétique des caractères phénotypiques impliqués dans l'adaptation locale des téosintes par des approches de cartographies ou de génétique d'association. Lauter *et al.* (2004) ont détecté deux QTL majeurs correspondant aux gènes *B1* (*colored plant 1*) et *A3* (*anthocyanin3*), appartenant à la voie de biosynthèse des anthocyanines et impliqués dans la coloration des feuilles, ainsi que le QTL *Mh1* (*macrohairless1*) impliqué dans la pilosité. Le locus *C2* (*colorless2*), à plus faible effet, a également été détecté pour la pigmentation. Une étude de génétique d'association sur des téosintes semble montrer qu'un certain nombre de gènes de domestication et d'amélioration présentent des séries alléliques chez la téosinte, comme *Bal*, *Elm1* (*elongated mesocotyl1*), *Ts6* (*tasselseed6*), *Ra1* (*ramosa1*) e *tRa2* (*ramosa2*), *Tb1*, *Te1* (*terminal ear1*), *Td1* (*thick tassel dwarf 1*), *Zfl1* (*zea floricaula 1*), *Zfl2* (*zea floricaula 2*), *Cca1* (*circadian clock associated1*), et *Gil* (*gigantea1*) (Studer & Doebley 2012 ; Weber *et al.* 2007 ; Weber *et al.* 2008)

### **I.3 OBJECTIFS DE LA THESE ET PRESENTATION DU TRAVAIL EFFECTUE**

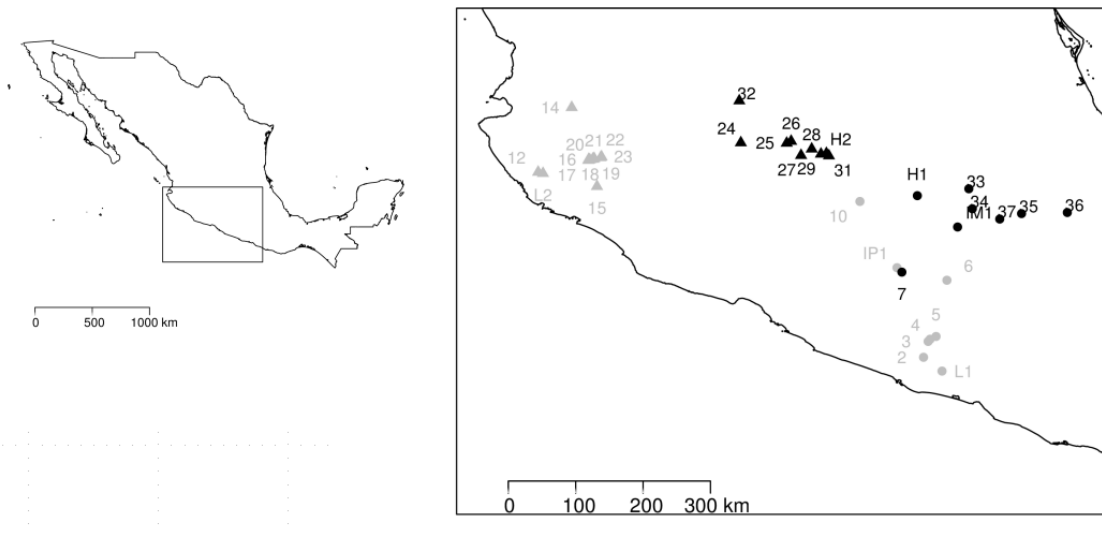
Au cours de mon doctorat je me suis intéressée aux mécanismes génétiques de l'adaptation locale. Ce type d'adaptation est intéressant à plusieurs égards. Tout d'abord, l'adaptation locale peut contribuer à une forte spécialisation écologique qui, à long terme, peut conduire à l'émergence de nouvelles espèces. Mesurer l'adaptation locale permet aussi de mieux prédire le devenir des populations dans le cadre des changements climatiques rapides actuels (TEAM 2001), et permet d'évaluer le potentiel d'une des trois réponses (adaptation locale, plasticité phénotypique, migration) à ces changements. Enfin, comprendre les déterminants génétiques de cette adaptation peut permettre de proposer des nouvelles pistes en amélioration des espèces cultivées, là encore dans un contexte de changements climatiques importants avec de nouvelles pressions biotiques et abiotiques.

Les téosintes sont un bon modèle pour étudier l'adaptation locale : elles sont annuelles, allogames, possèdent des tailles de population importantes, poussent dans des conditions environnementales très variées et présentent une grande variation phénotypique. Par ailleurs, les outils et les banques de données développés pour le maïs sont utilisables pour la téosinte, ce qui facilite grandement l'analyse des données de NGS. De telles ressources ne sont généralement pas disponibles pour les espèces non-modèles. De plus, les téosintes constituent un réservoir de diversité génétique intéressant pour la sélection. Ce travail a donc

également été motivé par une portée agronomique et pourra servir à dépister des allèles adaptatifs qui pourraient être introgressés dans le maïs par simples croisements.

Afin d'étudier les bases génétiques de ces adaptations, nous avons choisi d'échantillonner des téosintes le long de deux gradients d'altitude qui sont utilisés d'une part pour mimer une variation climatique continue et d'autre part comme deux pseudo-répétitions biologiques. Nous parlons de l'étude de l'adaptation à l'altitude en gardant en tête que l'altitude désigne ici un ensemble d'adaptations car elle co-varie avec de nombreuses autres variables environnementales biotiques et abiotiques. Notre échantillonnage se démarque de l'étude précédente de Pyhäjärvi *et al.* (2013) par la pseudo-répétabilité des gradients altitudinaux et aussi par le fait que nos populations proviennent d'un seul et même échantillonnage effectué dans un laps de temps restreint et représentatif d'une diversité génétique existante à un temps donné. En effet, l'étude de Pyhäjärvi *et al.* (2013) comprenait onze populations de *parviglumis* dont sept provenaient de banques de graines, les quatre autres provenant de populations naturelles échantillonnées en 2007, ainsi que dix populations de *mexicana* échantillonnées en 2008. Nous disposons de graines de 37 populations naturelles récoltées sur 16 à 47 plantes mères (Figure 21, Annexe I.2), et de données climatiques le long de ces gradients par l'intermédiaire de 19 variables climatiques portant sur la température, les précipitations et la saisonnalité. Parmi ces populations, 16 proviennent du gradient 1, principalement situé dans l'état de Guerrero, et 21 proviennent du gradient 2 principalement localisé dans l'état de Jalisco. Ces 37 populations ont été utilisées pour le chapitre 2. Pour les chapitres 1 et 3, nous ne disposons que de 31 populations : les 25 populations numérotées de 1 à 31 ainsi que les populations H1, L1, H2, L2, IM1 et IP1.





**Figure 21 : Carte du Mexique et zoom sur la zone d'échantillonnage.** Les 37 populations de téosintes ont été échantillonnées le long de deux gradients altitudinaux (gradient 1 : rond, gradient 2 : triangle). Les populations *mexicana* sont représentées en noir, les populations *parviglumis* en gris.

L'approche que nous avons employée comporte plusieurs volets, de la découverte des SNP par des approches de génomique des populations, à leur validation à travers la mesure de phénotypes dans des jardins communs. Nous avons aussi privilégié un échantillonnage plutôt local, des mesures de phénotype en conditions réalistes et différentes échelles : tout d'abord génomique pour l'identification de régions candidates puis l'échelle des régions candidates pour mener un travail plus avancé sur la variation des fréquences alléliques et des phénotypes. Le manuscrit est présenté sous la forme de trois chapitres portant sur des études à des degrés divers d'avancement.

Chapitre 1 : J'ai identifié des allèles candidats à l'adaptation en analysant les données de séquençage de génomes complets de six populations : quatre populations « extrêmes » de téosintes - une population de *parviglumis* de basse altitude et une population de *mexicana* de haute altitude pour chaque gradient - ainsi que deux populations « intermédiaires » - une population de *parviglumis* et de *mexicana* de moyenne altitude pour l'un des deux gradients. Cet échantillonnage permet de capter la variation nucléotidique entre populations extrêmes de chaque gradient et de prendre en compte la part de variation due à la différenciation entre les sous-espèces en utilisant les populations intermédiaires. Pour la détection de l'adaptation, j'ai couplé une démarche à l'aveugle (scan génomique) à une démarche « candidate » en m'appuyant sur l'annotation du génome et les études antérieures disponibles dans la littérature. Le dispositif de séquençage comprend des individus en mélange mais étiquetés, ce

qui m'a aussi permis de me poser des questions méthodologiques sur l'impact du séquençage en mélange pour les estimations des fréquences alléliques et la détection de sélection. Ce chapitre est présenté sous la forme d'un article.

Chapitre 2 : J'ai identifié des polymorphismes candidats puis choisi des SNP candidats à l'adaptation dans ce chapitre. Parmi ces polymorphismes, j'ai choisi 97 SNP identifiés par des scans génomiques *Fst*, 93 SNP identifiés par la démarche candidate du chapitre 1, et 80 SNP situés dans des régions identifiées dans des études précédentes comme présentant des gènes de domestication ou d'amélioration. L'ensemble des 270 SNP a fait l'objet de génotypage haut-débit pour étudier les variations de fréquences alléliques le long des deux gradients altitudinaux. Ce travail a nécessité un investissement technique important pour l'identification du jeu de SNP candidats, et les cultures en serre et extractions d'ADN (27x20 individus). A ce stade, je présente principalement le travail amont, les méthodes utilisées et quelques résultats préliminaires d'études de clines.

Chapitre 3 : Le Chapitre 3 est consacré à l'étude des relations entre les polymorphismes candidats et des caractères adaptatifs phénotypiques, ce qui constitue un réel pas dans la détection de l'adaptation locale. J'ai mis au point un dispositif expérimental dans deux lieux de moyenne altitude au Mexique, permettant d'évaluer onze populations dont les quatre populations extrêmes du chapitre 1. J'ai choisi de façon raisonnée 18 caractères en lien avec la valeur sélective et que je pensais également être les plus liés à des processus d'adaptation à l'altitude. J'ai contribué à mesurer ces caractères en effectuant deux séjours de deux mois au Mexique sur le terrain en 2013 et en 2014. Je présente ici ce dispositif et des résultats sur l'analyse des données phénotypiques. J'ai également largement contribué à la prise d'échantillons, leur envoi, et les extractions d'ADN (environ 2 600 plantes). Les individus ont été génotypés pour le jeu de SNP candidats identifiés dans le chapitre 2. Quarante-deux marqueurs microsatellites ont également été génotypés sur ces individus afin de prendre en compte la structuration des populations de téosinte et limiter ainsi le nombre de faux positifs. Les analyses de structure sont présentées dans ce chapitre, mais les analyses de génétique d'association dont je présente le principe sont encore en cours. A notre connaissance, aucune étude d'adaptation locale sur les téosintes n'a couplé les approches de génomique des populations à celle de génétique d'association.

## **REFERENCES BIBLIOGRAPHIQUES**

- Åbjörnsson K, Hansson L-A, Brönmark C (2004) Responses of prey from habitats with different predator regimes: local adaptation and heritability. *Ecology* **85**, 1859-1866.
- Agrama HA, Moussa ME (1996) Mapping QTLs in breeding for drought tolerance in maize (*Zea mays* L.). *Euphytica* **91**, 89-97.
- Ågren J, Schemske DW (2012) Reciprocal transplants demonstrate strong adaptive differentiation of the model organism *Arabidopsis thaliana* in its native range. *New Phytologist* **194**, 1112-1122.
- Albrechtsen A, Nielsen FC, Nielsen R (2010) Ascertainment biases in SNP chips affect measures of population divergence. *Molecular biology and evolution* **27**, 2534-2547.
- Allen WR, Sheppard PM (1971) Copper Tolerance in Some Californian Populations of the Monkey Flower, *Mimulus guttatus*. *Proceedings of the Royal Society of London B: Biological Sciences* **177**, 177-196.
- Anderson JT, Inouye DW, McKinney AM, Colautti RI, Mitchell-Olds T (2012) Phenotypic plasticity and adaptive evolution contribute to advancing flowering phenology in response to climate change. *Proceedings of the Royal Society of London B: Biological Sciences* **279**, 3843-3852.
- Anderson JT, Lee C-R, Rushworth CA, Colautti RI, Mitchell-Olds T (2013) Genetic trade-offs and conditional neutrality contribute to local adaptation. *Molecular Ecology* **22**, 699-708.
- Anderson JT, Willis JH, Mitchell-Olds T (2011) Evolutionary genetics of plant adaptation. *Trends in Genetics* **27**, 258-266.
- Aranzana MJ, Kim S, Zhao K, et al. (2005) Genome-wide association mapping in *Arabidopsis* identifies previously known flowering time and pathogen resistance genes. *PLoS Genet* **1**, e60.
- Atwell S, Huang YS, Vilhjálmsson BJ, et al. (2010) Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* **465**, 627-631.
- Baer CF, Miyamoto MM, Denver DR (2007) Mutation rate variation in multicellular eukaryotes: causes and consequences. *Nature Review Genetics* **8**, 619-631.
- Baltazar BM, Sanchez-Gonzalez JD, de la Cruz-Larios L, Schoper JB (2005) Pollination between maize and teosinte: an important determinant of gene flow in Mexico. *Theoretical and Applied Genetics* **110**, 519-526.
- Bamshad M, Wooding S, Salisbury BA, Stephens JC (2004) Deconstructing the relationship between genetics and race. *Nat Rev Genet* **5**, 598-609.
- Beall CM (2006) Andean, Tibetan, and Ethiopian patterns of adaptation to high-altitude hypoxia. *Integrative and Comparative Biology* **46**, 18-24.
- Beall CM, Brittenham GM, Strohl KP, et al. (1998) Hemoglobin concentration of high-altitude Tibetans and Bolivian Aymara. *American Journal of Physical Anthropology* **106**, 385-400.
- Beall CM, Decker MJ, Brittenham GM, et al. (2002) An Ethiopian pattern of human adaptation to high-altitude hypoxia. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 17215-17218.
- Bennett MD (1987) Variation in genomic form in plants and its ecological implications. *New Phytologist* **106**, 177-200.
- Bergelson J, Roux F (2010) Towards identifying genes underlying ecologically relevant traits in *Arabidopsis thaliana*. *Nature Reviews Genetics* **11**, 867-879.
- Bhattacharyya MK, Smith AM, Ellis TN, Hedley C, Martin C (1990) The wrinkled-seed character of pea described by Mendel is caused by a transposon-like insertion in a gene encoding starch-branching enzyme. *Cell* **60**, 115-122.
- Bhave MR, Lawrence S, Barton C, Hannah LC (1990) Identification and molecular characterization of Shrunken-2 cDNA of maize. *Plant Cell* **2**, 581-588.
- Bishop JA, Cook LM (1980) Industrial melanism and the urban environment. *Advances in ecological research* **11**, 373-404.
- Boitard S, Kofler R, Françoise P, et al. (2013) Pool-hmm: a Python program for estimating the allele frequency spectrum and detecting selective sweeps from next generation sequencing of pooled samples. *Molecular Ecology Resources* **13**, 337-340.
- Boitard S, Schlötterer C, Nolte V, Pandey RV, Futschik A (2012) Detecting selective sweeps from pooled next-generation sequencing samples. *Molecular Biology and Evolution* **29**, 2177-2186.
- Bonhomme M, Chevalet C, Servin B, et al. (2010) Detecting Selection in Population Trees: The Lewontin and Krakauer Test Extended. *Genetics* **186**, 241-262.
- Brachi B, Faure N, Horton M, et al. (2010) Linkage and association mapping of *Arabidopsis thaliana* flowering time in nature. *PLoS Genetics* **6**, e1000940.
- 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.

- Briggs WH, McMullen MD, Gaut BS, Doebley J (2007) Linkage mapping of domestication loci in a large maize-teosinte backcross resource. *Genetics* **177**, 1915-1928.
- Buckler ES, Goodman MM, Holtsford TP, Doebley JF, Sanchez J (2006) Phylogeography of the wild subspecies of *Zea mays*. *Maydica* **51**, 123-134.
- Buckler ES, Holland JB, Bradbury PJ, *et al.* (2009) The genetic architecture of maize flowering time. *Science* **325**, 714-718.
- Buet C, Dubreuil P, Tixier M-H, Durantin K, Praud S (2013) The molecular characterization of a MAGIC population reveals high potential for gene discovery. *Maize Genetics Conference* **55**, 314.
- Chardon F, Virlon B, Moreau L, *et al.* (2004) Genetic architecture of flowering time in maize as inferred from quantitative trait loci meta-analysis and synteny conservation with the rice genome. *Genetics* **168**, 2169-2185.
- Chevin LM, Lande R (2011) Adaptation to marginal habitats by evolution of increased phenotypic plasticity. *Journal of Evolutionary Biology* **24**, 1462-1476.
- Chia J-M, Song C, Bradbury PJ, *et al.* (2012) Maize HapMap2 identifies extant variation from a genome in flux. *Nature genetics* **44**, 803-807.
- Clark RM, Tavaré S, Doebley J (2005) Estimating a nucleotide substitution rate for maize from polymorphism at a major domestication locus. *Molecular Biology and Evolution* **22**, 2304-2312.
- Colautti RI, Lee C-R, Mitchell-Olds T (2012) Origin, fate, and architecture of ecologically relevant genetic variation. *Current opinion in plant biology* **15**, 199-204.
- Cook JP, McMullen MD, Holland JB, *et al.* (2012) Genetic architecture of maize kernel composition in the nested association mapping and inbred association panels. *Plant Physiology* **158**, 824-834.
- Cook L, Grant B (2000) Frequency of insularia during the decline in melanics in the peppered moth *Biston betularia* in Britain. *Heredity* **85**, 580-585.
- Coop G, Witonsky D, Di Rienzo A, Pritchard JK (2010) Using Environmental Correlations to Identify Loci Underlying Local Adaptation. *Genetics* **185**, 1411-1423.
- Coque M, Gallais A (2008) Genetic variation for N-remobilization and postsilking N-up-take in a set of maize recombinant inbred lines. 2. Line per se performance and comparison with testcross performance. *Maydica* **53**, 29-38.
- Darwin C (1859) On the origins of species by means of natural selection. *London: Murray*, 247.
- Daub JT, Hofer T, Cutivet E, *et al.* (2013) Evidence for polygenic adaptation to pathogens in the human genome. *Molecular Biology and Evolution*, mst080.
- de Meaux J, Hu J-Y, Tartler U, Goebel U (2008) Structurally different alleles of the *ath-MIR824* microRNA precursor are maintained at high frequency in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 8994-8999.
- Derryberry EP, Derryberry GE, Maley JM, Brumfield RT (2014) HZAR: hybrid zone analysis using an R software package. *Molecular Ecology Resources* **14**, 652-663.
- DeWitt TJ, Sih A, Wilson DS (1998) Costs and limits of phenotypic plasticity. *Trends in Ecology & Evolution* **13**, 77-81.
- Doebley J (1992) Mapping the genes that made maize. *Trends in Genetics* **8**, 302-307.
- Doebley J, Stec A, Gustus C (1995) Teosinte branched1 and the origin of maize: evidence for epistasis and the evolution of dominance. *Genetics* **141**, 333-346.
- Doebley J, Stec A, Hubbard L (1997) The evolution of apical dominance in maize. *Nature* **386**, 485-488.
- Drake JW, Charlesworth B, Charlesworth D, Crow JF (1998) Rates of Spontaneous Mutation. *Genetics* **148**, 1667-1686.
- Durand E, Tenaillon MI, Raffoux X, *et al.* (2015) Dearth of polymorphism associated with a sustained response to selection for flowering time in maize. *BMC Evolutionary Biology* **15**, 1-16.
- Eizaguirre C, Lenz TL (2010) Major histocompatibility complex polymorphism: dynamics and consequences of parasite-mediated local adaptation in fishes. *Journal of Fish Biology* **77**, 2023-2047.
- Ellegren H (2000) Microsatellite mutations in the germline: implications for evolutionary inference. *Trends in Genetics* **16**, 551-558.
- Ewing B, Green P (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Research* **8**, 186-194.
- Excoffier L, Hofer T, Foll M (2009) Detecting loci under selection in a hierarchically structured population. *Heredity* **103**, 285-298.
- Fang Z, Pyhäjärvi T, Weber AL, *et al.* (2012) Megabase-scale inversion polymorphism in the wild ancestor of maize. *Genetics* **191**, 883-894.
- Fariello M-I, Servin B, Tosser-Klopp G, *et al.* (2014) Selection signatures in worldwide Sheep populations. *PLoS ONE* **9**, e103813

- Fariello MI, Boitard S, Naya H, SanCristobal M, Servin B (2013) Detecting signatures of selection through haplotype differentiation among hierarchically structured populations. *Genetics* **193**, 929-941.
- Fay JC, Wu CI (2000) Hitchhiking under positive Darwinian selection. *Genetics* **155**, 1405-1413.
- Fisher RA (1930) *The genetical theory of natural selection: a complete variorum edition* Oxford University Press.
- Foll M, Gaggiotti O (2008) A Genome-Scan Method to Identify Selected Loci Appropriate for Both Dominant and Codominant Markers: A Bayesian Perspective. *Genetics* **180**, 977-993.
- Fournier-Level A, Korte A, Cooper MD, *et al.* (2011) A map of local adaptation in *Arabidopsis thaliana*. *Science* **334**, 86-89.
- Franks SJ, Weber JJ, Aitken SN (2014) Evolutionary and plastic responses to climate change in terrestrial plant populations. *Evolutionary Applications* **7**, 123-139.
- Fukunaga K, Hill J, Vigouroux Y, *et al.* (2005) Genetic diversity and population structure of teosinte. *Genetics* **169**, 2241-2254.
- Ganal MW, Durstewitz G, Polley A, *et al.* (2011) A large maize (*zea mays* l.) SNP genotyping array: development and germplasm genotyping, and genetic mapping to compare with the b73 reference genome. *PLoS ONE* **6** : e28334.
- Gay L, Crochet PA, Bell DA, Lenormand T (2008) Comparing clines on molecular and phenotypic traits in hybrid zones: a window on tension zone models. *Evolution* **62**, 2789-2806.
- Gonzalez J, Karasov TL, Messer PW, Petrov DA (2010) Genome-Wide Patterns of Adaptation to Temperate Environments Associated with Transposable Elements in *Drosophila*. *PLoS Genetics* **6**, 000905.
- Gore MA, Chia JM, Elshire RJ, *et al.* (2009) A First-Generation Haplotype Map of Maize. *Science* **326**, 1115-1117.
- Guichoux E, Lagache L, Wagner S, *et al.* (2011) Current trends in microsatellite genotyping. *Molecular Ecology Resources* **11**, 591-611.
- Günther T, Coop G (2013) Robust Identification of Local Adaptation from Allele Frequencies. *Genetics* **195**, 205-220.
- Han E, Sinsheimer JS, Novembre J (2013) Characterizing bias in population genetic inferences from low-coverage sequencing data. *Molecular Biology and Evolution* **31**, 723 - 735.
- Hancock AM, Brachi B, Faure N, *et al.* (2011) Adaptation to climate across the *Arabidopsis thaliana* genome. *Science* **334**, 83-86.
- Hanikenne M, Kroymann J, Trampczynska A, *et al.* (2013) Hard selective sweep and ectopic gene conversion in a gene cluster affording environmental adaptation. *PLoS Genetics* **9**, e1003707.
- Hanikenne M, Talke IN, Haydon MJ, *et al.* (2008) Evolution of metal hyperaccumulation required cis-regulatory changes and triplication of HMA4. *Nature* **453**, 391-395.
- Hanson MA, Gaut BS, Stec AO, *et al.* (1996) Evolution of anthocyanin biosynthesis in maize kernels: the role of regulatory and enzymatic loci. *Genetics* **143**, 1395-1407.
- Harjes CE, Rocheford TR, Bai L, *et al.* (2008) Natural genetic variation in lycopene epsilon cyclase tapped for maize biofortification. *Science* **319**, 330-333.
- Hereford J (2009) A quantitative survey of local adaptation and fitness trade-offs. *The American Naturalist* **173**, 579-588.
- Hermisson J, Pennings PS (2005) Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics* **169**, 2335-2352.
- Hoffmann AA, Weeks AR (2007) Climatic selection on genes and traits after a 100 year-old invasion: a critical look at the temperate-tropical clines in *Drosophila melanogaster* from eastern Australia. *Genetica* **129**, 133-147.
- Hohenlohe PA, Bassham S, Etter PD, *et al.* (2010) Population genomics of parallel adaptation in threespine sticklebacks using sequenced RAD tags. *PLoS Genetics* **6**, e1000862.
- Hörger AC, Fones HN, Preston GM (2013) The current status of the elemental defense hypothesis in relation to pathogens. *Frontiers in plant science* **4**, 395.
- Huang BE, Verbyla KL, Verbyla AP, *et al.* (2015) MAGIC populations in crops: current status and future prospects. *Theoretical and Applied Genetics* **128**, 999-1017.
- Hudson ME (2008) Sequencing breakthroughs for genomic ecology and evolutionary biology. *Molecular Ecology Resources* **8**, 3-17.
- Hufford M, Martinez-Meyer E, Gaut BS, Eguiarte LE, Tenaillon MI (2012a) Inferences from the historical distribution of wild and domesticated maize provide ecological and evolutionary insight. *PLoS ONE* **7**, e47659.
- Hufford MB, Xu X, van Heerwaarden J, *et al.* (2012b) Comparative population genomics of maize domestication and improvement. *Nature Genetics* **44**, 808-811.
- Hung H-Y, Shannon LM, Tian F, *et al.* (2012) ZmCCT and the genetic basis of day-length adaptation underlying the postdomestication spread of maize. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 1913-1921.

- Huxley JS (1938) Clines: an auxiliary taxonomic principle. *Nature* **142**, 219-220.
- Iltis HH, Doebley JF (1980) Taxonomy of *Zea* (*Gramineae*) .2. Subspecific categories in the *Zea mays* complex and a generic synopsis. *American Journal of Botany* **67**, 994-1004.
- Joron M, Frezal L, Jones RT, *et al.* (2011) Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature* **477**, 203-206.
- Joron M, Papa R, Beltrán M, *et al.* (2006) A conserved supergene locus controls colour pattern diversity in *Heliconius* butterflies. *PLoS Biology* **4**, e303.
- Kaplan NL, Hudson RR, Langley CH (1989) The "Hitchhiking Effect" revisited. *Genetics* **123**, 887-899.
- Kawecki TJ, Ebert D (2004) Conceptual issues in local adaptation. *Ecology Letters* **7**, 1225-1241.
- Keller SR, Levens N, Olson MS, Tiffin P (2012) Local adaptation in the flowering-time gene network of balsam poplar, *Populus balsamifera* L. *Molecular Biology and Evolution* **29**, 3143-3152.
- Kirkpatrick M, Barton N (2006) Chromosome inversions, local adaptation and speciation. *Genetics* **173**, 419-434.
- Kover PX, Valdar W, Trakalo J, *et al.* (2009) A multiparent advanced generation inter-cross to fine-map quantitative traits in *Arabidopsis thaliana*. *PLoS Genetics* **5**, e1000551.
- Kyete D, Ming R, McMullen M, *et al.* (1999) Genetic analysis of tolerance to maize streak virus in maize. *Genome* **42**, 20-26.
- Labbé P, Berthomieu A, Berticat C, *et al.* (2007) Independent duplications of the acetylcholinesterase gene conferring insecticide resistance in the mosquito *Culex pipiens*. *Molecular Biology and Evolution* **24**, 1056-1067.
- Labbé P, Sidos N, Raymond M, Lenormand T (2009) Resistance gene replacement in the mosquito *Culex pipiens*: fitness estimation from long-term cline series. *Genetics* **182**, 303-312.
- Lande R (2009) Adaptation to an extraordinary environment by evolution of phenotypic plasticity and genetic assimilation. *Journal of Evolutionary Biology* **22**, 1435-1446.
- Lander ES, Botstein D (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**, 185-199.
- Lander ES, Schork NJ (1994) Genetic dissection of complex traits. *Science* **265**, 2037-2048.
- Lauter N, Gustus C, Westerbergh A, Doebley J (2004) The inheritance and evolution of leaf pigmentation and pubescence in teosinte. *Genetics* **167**, 1949-1959.
- Le Thierry d'Ennequin M, Toupance B, Robert T, Godelle B, Gouyon PH (1999) Plant domestication: a model for studying the selection of linkage. *Journal of Evolutionary Biology* **12**, 1138-1147.
- Leimu R, Fischer M (2008) A meta-analysis of local adaptation in plants. *PLoS ONE* **3**, e4010.
- Leinonen T, McCairns RJS, O'Hara RB, Merila J (2013) QST-FST comparisons: evolutionary and ecological insights from genomic heterogeneity. *Nature Review Genetics* **14**, 179-190.
- Lennon JR, Krakowsky M, Goodman M, Flint-Garcia S, Balint-Kurti PJ (2016) Identification of Alleles Conferring Resistance to Gray Leaf Spot in Maize Derived from its Wild Progenitor Species Teosinte. *Crop Science* **56**, 209-218.
- Lenormand T, Guillemaud T, Bourguet D, Raymond M (1998) Appearance and sweep of a gene duplication: adaptive response and potential for new functions in the mosquito *Culex pipiens*. *Evolution*, 1705-1712.
- Li H, Handsaker B, Wysoker A, *et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079.
- Li H, Homer N (2010) A survey of sequence alignment algorithms for next-generation sequencing. *Briefings in Bioinformatics* **11**, 473-483.
- Lind MI, Ingvarsson PK, Johansson H, Hall D, Johansson F (2011) Gene flow and selection on phenotypic plasticity in an island system of *Rana temporaria*. *Evolution* **65**, 684-697.
- Linnen CR, Kingsley EP, Jensen JD, Hoekstra HE (2009) On the origin and spread of an adaptive allele in deer mice. *Science* **325**, 1095-1098.
- Liu L, Li Y, Li S, *et al.* (2012) Comparison of next-generation sequencing systems. *BioMed Research International* **2012**.
- Lohbeck KT, Riebesell U, Reusch TBH (2012) Adaptive evolution of a key phytoplankton species to ocean acidification. *Nature Geoscience* **5**, 346-351.
- Long Q, Rabanal FA, Meng D, *et al.* (2013) Massive genomic variation and strong selection in *Arabidopsis thaliana* lines from Sweden. *Nature genetics* **45**, 884-890.
- Louthan AM, Kay KM (2011) Comparing the adaptive landscape across trait types: larger QTL effect size in traits under biotic selection. *BMC Evolutionary Biology* **11**, 60.
- Lunter G, Goodson M (2011) Stampy: A statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Research* **21**, 936-939.
- Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. *Science* **290**, 1151-1155.

- MacNair MR, Smith SE, Cumbes QJ (1993) Heritability and distribution of variation in degree of copper tolerance in *Mimulus guttatus* at Copperopolis, California. *Heredity* **71**, 445-455.
- Malosetti M, Ribaut JM, Vargas M, Crossa J, Van Eeuwijk FA (2008) A multi-trait multi-environment QTL mixed model with an application to drought and nitrogen stress trials in maize (*Zea mays* L.). *Euphytica* **161**, 241-257.
- Manicacci D, Camus-Kulandaivelu L, Fourmann M, *et al.* (2009) Epistatic interactions between Opaque2 transcriptional activator and its target gene CyPPDK1 control kernel trait variation in maize. *Plant Physiology* **150**, 506-520.
- Mardis ER (2008) Next-generation DNA sequencing methods. *Annual Review of Genomics and Human Genetics* **9**, 387-402.
- Matsuoka Y, Vigouroux Y, Goodman MM, *et al.* (2002) A single domestication for maize shown by multilocus microsatellite genotyping. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 6080-6084.
- Maynard-Smith J, Haigh J (1974) The hitch-hiking effect of a favorable gene. *Genetics Research* **23**, 23-35.
- McMullen M, Louie R (1989) The linkage of molecular markers to a gene controlling the symptom response in maize to maize dwarf mosaic virus. *Molecular Plant-Microbe Interactions* **2**, 309-314.
- McMullen MD, Jones MW, Simcox KD, Louie R (1994) Three genetic loci control resistance to wheat streak mosaic virus in the maize inbred Pa405. *Molecular Plant Microbe Interactions* **7**, 708-712.
- Merila J, Crnokrak P (2001) Comparison of genetic differentiation at marker loci and quantitative traits. *Journal of Evolutionary Biology* **14**, 892-903.
- Merilä J, Hendry AP (2014) Climate change, adaptation, and phenotypic plasticity: the problem and the evidence. *Evolutionary Applications* **7**, 1-14.
- Merilä J, Laurila A, Lindgren B (2004) Variation in the degree and costs of adaptive phenotypic plasticity among *Rana temporaria* populations. *Journal of Evolutionary Biology* **17**, 1132-1140.
- Messer PW, Petrov DA (2013) Population genomics of rapid adaptation by soft selective sweeps. *Trends in Ecology & Evolution* **28**, 659-669.
- Ming R, Brewbaker J, Pratt R, Musket T, McMullen M (1997) Molecular mapping of a major gene conferring resistance to maize mosaic virus. *Theoretical and Applied Genetics* **95**, 271-275.
- Moeller DA, Tenailon MI, Tiffin P (2007) Population structure and its effects on patterns of nucleotide polymorphism in teosinte (*Zea mays* ssp. *parviglumis*). *Genetics* **176**, 1799-1809.
- Moreno-Gamez S, Hill AL, Rosenbloom DIS, *et al.* (2015) Imperfect drug penetration leads to spatial monotherapy and rapid evolution of multidrug resistance. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 2874-2883.
- Morin PA, Luikart G, Wayne RK, the SNPwg (2004) SNPs in ecology, evolution and conservation. *Trends in Ecology & Evolution* **19**, 208-216.
- Mullen LM, Hoekstra HE (2008) Natural selection along an environmental gradient: a classic cline in mouse pigmentation. *Evolution* **62**, 1555-1570.
- Mullen LM, Vignieri SN, Gore JA, Hoekstra HE (2009) Adaptive basis of geographic variation: genetic, phenotypic and environmental differences among beach mouse populations. *Proceedings of the Royal Society of London B: Biological Sciences* **278**, 3809-3818.
- Muñoz Diez C, Gaut BS, Meca E, *et al.* (2013) Genome size variation in wild and cultivated maize along altitudinal gradients. *New Phytologist* **199**, 264-276.
- Nielsen R (2005) Molecular signatures of natural selection. *Annual Review of Genetics* **39**, 197-218.
- Nielsen R, Paul JS, Albrechtsen A, Song YS (2011) Genotype and SNP calling from next-generation sequencing data. *Nature Reviews Genetics* **12**, 443-451.
- Ninamango-Cárdenas FE, Guimarães CT, Martins PR, *et al.* (2003) Mapping QTLs for aluminum tolerance in maize. *Euphytica* **130**, 223-232.
- Palaisa KA, Morgante M, Williams M, Rafalski A (2003) Contrasting effects of selection on sequence diversity and linkage disequilibrium at two phytoene synthase loci. *Plant Cell* **15**, 1795-1806.
- Paterson A, Bowers J, Chapman B (2004) Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 9903-9908.
- Pea G, Aung HH, Frascaroli E, Landi P, Pè ME (2013) Extensive genomic characterization of a set of near-isogenic lines for heterotic QTL in maize (*Zea mays* L.). *BMC genomics* **14**, 61.
- Pigliucci M (2005) Evolution of phenotypic plasticity: where are we going now? *Trends in Ecology & Evolution* **20**, 481-486.
- Piperno DR, Ranere AJ, Holst I, Iriarte J, Dickau R (2009) Starch grain and phytolith evidence for early ninth

- millennium BP maize from the Central Balsas River Valley, Mexico. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 5019-5024.
- Price AL, Patterson NJ, Plenge RM, *et al.* (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics* **38**, 904-909.
- Pritchard JK, Di Rienzo A (2010) Adaptation - not by sweeps alone. *Nature Reviews. Genetics*. **11**, 665-667.
- Pritchard JK, Stephens M, Donnelly P (2000a) Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-959.
- Pritchard JK, Stephens M, Rosenberg N, Donnelly P (2000b) Association mapping in structured populations. *American Journal of Human Genetics* **67**, 170-181.
- Pyhäjärvi T, Hufford MB, Mezouk S, Ross-Ibarra J (2013) Complex patterns of local adaptation in teosinte. *Genome biology and evolution* **5**, 1594-1609.
- Qiu F, Zheng Y, Zhang Z, Xu S (2007) Mapping of QTL associated with waterlogging tolerance during the seedling stage in maize. *Annals of Botany* **99**, 1067-1081.
- Quail MA, Smith M, Coupland P, *et al.* (2012) A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC genomics* **13**, 341.
- Rafalski JA (2010) Association genetics in crop improvement. *Current opinion in plant biology* **13**, 174-180.
- Ravigné V., Barberousse A. Bierne N, *et al.* La spéciation. In Thomas F, Lefèvre T, Raymond (2010) *Biologie évolutive* Edition De Boeck. **4** 87.
- Biradar DP (1994) Selection for early flowering results in simultaneous selection for reduced nuclear DNA content in maize. *Plant Breeding* **112**, 318-322.
- Reiter RS, Coors J, Sussman M, Gabelman W (1991) Genetic analysis of tolerance to low-phosphorus stress in maize using restriction fragment length polymorphisms. *Theoretical and Applied Genetics* **82**, 561-568.
- Rogers AR, Jorde LB (1996) Ascertainment bias in estimates of average heterozygosity. *American Journal of Human Genetics* **58**, 1033-1041.
- Ross-Ibarra J, Tenaillon MI, Gaut BS (2009) Historical divergence and gene flow in the genus *Zea*. *Genetics* **181**, 1399-1413.
- Salvi S, Sponza G, Morgante M, *et al.* (2007) Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 11376-11381.
- SanMiguel PJ, Gaut BS, Tikhonov A, Nakajima Y, Bennetzen JL (1998) The Paleontology of Intergene Retrotransposons of Maize. *Nature genetics* **20**, 43-45.
- Savolainen O, Lascoux M, Merilä J (2013) Ecological genomics of local adaptation. *Nature Reviews Genetics* **14**, 807-820.
- Scheinfeldt LB, Tishkoff SA (2013) Recent human adaptation: genomic approaches, interpretation and insights. *Nature Reviews Genetics* **14**, 692-702.
- Schmidt JM, Good RT, Appleton B, *et al.* (2010) Copy number variation and transposable elements feature in recent, ongoing adaptation at the *Cyp6g1* locus. *PLoS Genetics* **6**, e10009998.
- Schnable PS, Ware D, Fulton RS, *et al.* (2009) The B73 Maize Genome: Complexity, Diversity, and Dynamics. *Science* **326**, 1112-1115.
- Sheldon CC, Rouse DT, Finnegan EJ, Peacock WJ, Dennis ES (2000) The molecular basis of vernalization: The central role of flowering locus S C (FLC). *Proceedings of the National Academy of Sciences of the United States of America* **97**, 3753-3758.
- Sibov S, Gaspar M, Silva M, *et al.* (1999) Two genes control aluminum tolerance in maize: genetic and molecular mapping analyses. *Genome* **42**, 475-482.
- Sicard A, Lenhard M (2011) The selfing syndrome: a model for studying the genetic and evolutionary basis of morphological adaptation in plants. *Annals of Botany* **107**, 1433-1443.
- Sigmon B, Vollbrecht E (2010) Evidence of selection at the ramosal locus during maize domestication. *Molecular Ecology* **19**, 1296-1311.
- Slatkin M (1987) Gene flow and the geographic structure of natural populations. *Science* **236**, 787-792.
- Studer AJ, Doebley JF (2011) Do large effect QTL fractionate? a case study at the maize domestication QTL teosinte branched1. *Genetics* **188**, 673-681.
- Studer AJ, Doebley JF (2012) Evidence for a natural allelic series at the maize domestication locus teosinte branched1. *Genetics* **191**, 551-558.
- Sultan SE, Horgan-Kobelski T, Nichols LM, Riggs CE, Waples RK (2013) A resurrection study reveals rapid adaptive evolution within populations of an invasive plant. *Evolutionary Applications* **6**, 266-278.
- Sultan SE, Spencer HG (2002) Metapopulation structure favors plasticity over local adaptation. *The American Naturalist* **160**, 271-283.



- Swigonova Z, Lai J, Ma J, *et al.* (2004) On the tetraploid origin of the maize genome. *Comparative and functional genomics* **5**, 281-284.
- Takuno S, Ralph P, Swarts K, *et al.* (2015) Independent molecular basis of convergent highland adaptation in maize. *Genetics Early online* June 15, 2015
- TEAM C (2001) Climate change 2001: synthesis report.
- Tenaillon MI, Charcosset A (2011) A European perspective on maize history. *Comptes Rendus Biologies* **334**, 221-228.
- Tenaillon O, Rodríguez-Verdugo A, Gaut RL, *et al.* (2012) The Molecular Diversity of Adaptive Convergence. *Science* **335**, 457-461.
- Thompson J, Charpentier A, Bouguet G, *et al.* (2013) Evolution of a genetic polymorphism with climate change in a Mediterranean landscape. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 2893-2897.
- Tian F, Bradbury PJ, Brown PJ, *et al.* (2011) Genome-wide association study of leaf architecture in the maize nested association mapping population. *Nature genetics* **43**, 159-162.
- Tiffin P, Ross-Ibarra J (2014) Advances and limits of using population genetics to understand local adaptation. *Trends in Ecology & Evolution* **29**, 673-680.
- Tishkoff SA, Reed FA, Ranciaro A, *et al.* (2007) Convergent adaptation of human lactase persistence in Africa and Europe. *Nature genetics* **39**, 31-40.
- Tracy WF, Whitt SR, Buckler ES (2006) Recurrent Mutation and Genome Evolution: Example of and the Origin of Sweet Maize. *Crop Science* **46**, S-49-S-54.
- Tsai ZT-Y, Tsai H-K, Cheng J-H, *et al.* (2012) Evolution of cis-regulatory elements in yeast de novo and duplicated new genes. *BMC genomics* **13**, 717.
- Turner TL, Bourne EC, Von Wettberg EJ, Hu TT, Nuzhdin SV (2010) Population resquencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. *Nature genetics* **42**, 260-263.
- Unterseer S, Bauer E, Haberer G, *et al.* (2014) A powerful tool for genome analysis in maize: development and evaluation of the high density 600k SNP genotyping array. *BMC genomics* **15**, 823.
- van Heerwaarden J, Doebley J, Briggs WH, *et al.* (2011) Genetic signals of origin, spread, and introgression in a large sample of maize landraces. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 1088-1092.
- Van Heerwaarden J, ROSS-IBARRA J, Doebley J, *et al.* (2010) Fine scale genetic structure in the wild ancestor of maize (*Zea mays* ssp. *parviglumis*). *Molecular Ecology* **19**, 1162-1173.
- Van Kleunen M, Fischer M (2005) Constraints on the evolution of adaptive phenotypic plasticity in plants. *New Phytologist* **166**, 49-60.
- Via S, Gomulkiewicz R, De Jong G, *et al.* (1995) Adaptive phenotypic plasticity: consensus and controversy. *Trends in Ecology & Evolution* **10**, 212-217.
- Vitte C, Fustier MA, Alix K, Tenaillon MI (2014) The bright side of transposons in crop evolution. *Briefings in functional genomics* **13**, 276-295.
- Wallace J, The bright side of transposons in crop evolution. Larsson S, Buckler E (2014) Entering the second century of maize quantitative genetics. *Heredity* **112**, 30-38.
- Wang G-D, Fan R-X, Zhai W, *et al.* (2014) Genetic Convergence in the Adaptation of Dogs and Humans to the High-Altitude Environment of the Tibetan Plateau. *Genome biology and evolution* **6**, 2122-2128.
- Wang H, Nussbaum-Wagler T, Li B, *et al.* (2005) The origin of the naked grains of maize. *Nature* **436**, 714-719.
- Wang T, O'Neill GA, Aitken SN (2010) Integrating environmental and genetic effects to predict responses of tree populations to climate. *Ecological Applications* **20**, 153-163.
- Weber AL, Briggs WH, Rucker J, *et al.* (2008) The genetic architecture of complex traits in teosinte (*Zea mays* ssp. *parviglumis*): new evidence from association mapping. *Genetics* **180**, 1221-1232.
- Weber AL, Zhao Q, McMullen MD, Doebley JF (2009) Using association mapping in teosinte to investigate the function of maize selection-candidate genes. *PLoS ONE* **4**.
- Weill M, Berticat C, Raymond M, Chevillon C (2000) Quantitative polymerase chain reaction to estimate the number of amplified esterase genes in insecticide-resistant mosquitoes. *Analytical biochemistry* **285**, 267-270.
- Welz H, Xia X, Bassetti P, Melchinger A, Lübberstedt T (1999) QTLs for resistance to *Setosphaeria turcica* in an early maturing Dent × Flint maize population. *Theoretical and Applied Genetics* **99**, 649-655.
- Wessler SR (1989) The splicing of maize transposable elements from pre-mRNA—a minireview. *Gene* **82**, 127-133.
- Whipple CJ, Kebrom TH, Weber AL, *et al.* (2011) grassy tillers1 promotes apical dominance in maize and responds to shade signals in the grasses. *Proceedings of the National Academy of Sciences of the United States of America* **108**, e506-e512.
- Whitt SR, Wilson LM, Tenaillon MI, Gaut BS, Buckler IV ES (2002) Genetic diversity and selection in the maize starch pathway. *Proceedings of the National Academy of Sciences of the United States of America* **99**,

12959-12962.

- Wilkes HG (1977) Hybridization of maize and teosinte in Mexico and Guatemala and the improvement of maize. *Economic Botany* **31**, 254-293.
- Wolf JB, Wade MJ (2009) What are maternal effects (and what are they not)? *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **364**, 1107-1115.
- Wright SI, Bi IV, Schroeder SG, *et al.* (2005) The effects of artificial selection on the maize genome. *Science* **308**, 1310-1314.
- Xia X, Melchinger AE, Kuntze L, Lübberstedt T (1999) Quantitative trait loci mapping of resistance to sugarcane mosaic virus in maize. *Phytopathology* **89**, 660-667.
- Yamasaki M, Tenaillon MI, Bi IV, *et al.* (2005) A large-scale screen for artificial selection in maize identifies candidate agronomic loci for domestication and crop improvement. *Plant Cell* **17**, 2859-2872.
- Yeaman S, Otto SP (2011) Establishment and maintenance of adaptive genetic divergence under migration, selection, and drift. *Evolution* **65**, 2123-2129.
- Yi X, Liang Y, Huerta-Sanchez E, *et al.* (2010) Sequencing of 50 human exomes reveals adaptation to high altitude. *Science* **329**, 75-78.
- Yu J, Buckler ES (2006) Genetic association mapping and genome organization of maize. *Current Opinion in Biotechnology* **17**, 155-160.
- Yu J, Holland JB, McMullen MD, Buckler IV ES (2008) Genetic design and statistical power of nested association mapping in maize. *Genetics* **178**, 539-551.
- Yu J, Pressoir G, Briggs WH, *et al.* (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet* **38**, 203-208.
- Yu L, Wang X-y, Jin W, *et al.* (2010) Adaptive evolution of digestive RNASE1 genes in leaf-eating monkeys revisited: New insights from ten additional colobines. *Molecular Biology and Evolution* **27**, 121-131.
- Zhang J, Zhang Y-p, Rosenberg HF (2002) Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey. *Nature genetics* **30**, 411-415.
- Zhao K, Aranzana MJ, Kim S, *et al.* (2007) An Arabidopsis example of association mapping in structured samples. *PLoS Genetics* **3**, e4.
- Zhu Y, Bergland AO, González J, Petrov DA (2012) Empirical validation of pooled whole genome population re-sequencing in *Drosophila melanogaster*. *PLoS ONE* **7**, e41901.



## **II. CHAPITRE 1 : LOCAL ADAPTATION OF TEOSINTES ALONG ALTITUDINAL GRADIENTS USING WHOLE GENOME SEQUENCING OF POOLED SAMPLES**

M-A Fustier <sup>1</sup>, J-T Brandenburg <sup>1</sup>, S. Boitard <sup>2,3</sup>, J. Lapeyronnie <sup>2,3</sup>, L.E. Eguiarte <sup>4</sup>, Y. Vigouroux <sup>5</sup>, D. Manicacci <sup>1</sup>, M.I Tenailon <sup>1</sup>

<sup>1</sup> Génétique Quantitative et Evolution – Le Moulon, INRA, Univ. Paris-Sud, CNRS, AgroParisTech, Université Paris-Saclay, Ferme du Moulon, F-91190 Gif-sur-Yvette, France

<sup>2</sup> Institut de Systématique, Evolution, Biodiversité, ISYEB, UMR 7205, CNRS & MNHM & UPMC & EPHE, Ecole Pratique de Hautes Etudes, Sorbonne –Universités, 45 rue Bouffon, CP 50, 75005 Paris, France

<sup>3</sup> GABI, INRA, AgroParistech, Université Paris-Saclay, domaine de Vilvert, 78352 Jouy en Josas Cedex, France

<sup>4</sup> Departamento de Ecología Evolutiva, Instituto de Ecología, Universidad Nacional Autónoma de México, Apartado Postal 70–275, Coyoacán, 04510 México D.F., México

<sup>5</sup> Institut de Recherche pour le développement (IRD), UMR Diversité, Adaptation et Développement des plantes (DIADE), 911 avenue Agropolis DP 6450134394 Montpellier Cedex 5, France

Keywords: *Fst*-scan, Environmental variables, Haplotypes, Subspecies differentiation, Selective sweeps



## **II.1 ABSTRACT.**

We aimed at mining the determinants of local adaptation in the two closest maize wild relatives, the teosintes *Zea mays* ssp. *parviglumis* and ssp. *mexicana*. We sequenced 20 individuals from two lowland and two highland populations sampled along two altitudinal gradients, considered as biological replicates. Sequencing of two middle elevation *parviglumis* and *mexicana* populations in one gradient helped controlling for subspecies differentiation. We detected 8,479,581 SNPs common to the 6 populations with an average sequencing depth per-site-per-population of 17x to 22x for “intermediate” and “extreme” populations respectively. Population diversity ranged from 0.10 to 0.15 and differentiation was greater between subspecies (0.119 in gradient 1, 0.122 in gradient 2) than within subspecies (0.048 for *mexicana*, 0.058 for *parviglumis*). We combined differentiation- and diversity-based methods to detect outlier SNPs that were further tested for correlation of allele frequencies with environmental variables. Outliers defined 43 candidate regions. We modified a haplotype-based method to incorporate genotype uncertainties in haplotype calling. We found haplotype signal consistency in 58% of our candidate regions along with several instances of the same haplotype being selected in lowland or highland populations. We confirmed the adaptive role of an inversion located on chromosome 1. We found frequent co-localization between our candidate regions and loci involved in the variation of traits linked to plant-soil interactions such as root morphology, aluminium and low phosphorus tolerance indicating that soil is an important factor driving local adaptation in teosintes.

## **II.2 INTRODUCTION.**

Local adaptation designates situations in which a population of local native genotypes outperforms genotypes from non-native introduced populations (Kawecki and Ebert 2004). It is widespread in plants. Hence transplantation experiments in multiple species have revealed that local populations outperform introduced populations in 71% of the cases (Leimu & Fischer 2008). Local adaptation results from a balance between spatially varying selection, gene flow, and drift (Yeaman & Otto 2011). Spatially varying selection consists of a shifting mix of biotic and abiotic factors including herbivory (Crémieux *et al.* 2008; Moeller & Tiffin 2008), parasitism (Sicard *et al.* 2007), edaphic conditions (Baxter *et al.* 2010; Turner *et al.* 2010), temperature (Rhone *et al.* 2008), and water-stress (Fournier-Level *et al.* 2011). Selection driving local adaptation often operates on complex phenotypes such as flowering time (Le Corre *et al.* 2002), timing of budset (Savolainen *et al.* 2011), seed size (Volis *et al.* 2002), life-history traits (Johnson 2007). The complexity of both environmental factors and responding traits makes the identification of loci governing local adaptive processes difficult.

Biotic and abiotic changes along latitudinal or altitudinal gradients allow the identification of functionally important genes contributing to the adaptive potential of species at a restricted scale (Franks *et al.* 2014). Hence, reverse ecology that uses genome scans to target clinal adaptation has delivered interesting results. SNPs correlating with climatic variables across species ranges are significantly enriched in amino-acid changing SNPs in *Arabidopsis thaliana* (Hancock *et al.* 2011) or are localized in genic regions in *Medicago truncatula* (Yoder *et al.* 2014), suggesting that local adaptation modelled the distribution of functional alleles in both species. SNPs from homologues to *A. thaliana* flowering time genes showed signature of local adaptation in Balsam Poplar (Keller *et al.* 2012).

Understanding local adaptation and species ability to respond to future climate changes is a major concern in crop species that represent human staple food (Lobell *et al.* 2011). Scenarios of climate changes predict a bolstering of regional differences, and hence an increasing need for developing locally adapted varieties (Howden *et al.* 2007; Porter & Semenov 2005). In crops, one of the only described systems for local adaptation is maize (*Zea mays* ssp. *mays*) and its closest annual wild progenitors, the teosintes *Zea mays* ssp. *parviglumis* and *Zea mays* ssp. *mexicana*. Teosintes are a good model to study local adaptation (Hufford *et al.* 2012b): they present a high level of diversity and correspondingly substantial effective population size (Ross-Ibarra *et al.* 2009); they are distributed across a

wide range of climates spanning environmental gradients from mesic low elevation to high elevation of the Central Plateau; they grow in relatively undisturbed habitats. Because teosintes are genetically very close to maize, extensive genomic resources available in maize, including a reference genome and gene functional annotations, further facilitate the discovery of genetic determinants of local adaptation in teosintes. One critical issue however is that the two teosinte subspecies, *Zea mays parviglumis* and *Zea mays mexicana*, grow in distinct ecological niches defined by altitude (Buckler *et al.* 2006; Hufford *et al.* 2012a) thereby introducing a confusion between climatic adaptation and subspecies divergence.

Pyhäjärvi *et al.* (2013) conducted a reverse ecology approach using the MaizeSNP50 BeadChip and scanning over 40,000 SNPs across the genomes of 21 teosinte populations. They identified around 2,000 candidate SNPs. Interestingly several of these genes belonged to 4 inversions localized respectively in chromosomes one, four and nine. Inversion *Inv1n* exhibited clinal patterns of allele frequency variation within the two subspecies, and *Inv9d* and *Inv9e* displayed this pattern only within the *mexicana* subspecies making all three good candidates for local adaptation (Fang *et al.* 2012; Pyhäjärvi *et al.* 2013). In contrast, *Inv4m* displayed differentiation between the two subspecies.

By focusing on a restricted and mainly genic compartment of the genome, the MaizeSNP50 BeadChip likely misses a number of important loci, especially because maize exhibits restricted linkage disequilibrium (Tenailon *et al.* 2001). A Genome Wide Association (GWA) study by Yu *et al.* (2012) illustrates this point by reporting that non-genic variants contribute to half of phenotypic variation of traits related to flowering time and leaf architecture in maize. Cost reduction of genome wide approaches can be achieved through sequencing of pooled-individuals from the same population (Pool-seq). This strategy is increasingly popular for population genomic analyses (Fischer *et al.* 2013; Turner *et al.* 2010). In a recent review Schlötterer *et al.* (2014) indicated that by sampling many more chromosomes at each locus, Pool-seq actually delivers greater accuracy in allele frequency estimates than sequencing of a restricted set of individual chromosomes i.e tagged (barcoded) individuals. This is especially true when the number of individuals contributing to the pool exceeds the number of tagged individuals. But, access to individuals may be limiting, especially in natural plant populations sometimes constituted by only 20-30 mother plants. More importantly, Pool-seq does not provide access to haplotype information, which has proven to be useful to detect selection (Fariello *et al.* 2013; Voight *et al.* 2006).

Detection of outlier loci that evolved under spatially divergent selection is an active field of investigation. Tests that identify larger differentiation values than expected by drift



have received particular attention. Hence, from the original *Fst*-based test proposed by Lewontin & Krakauer (1975), a number of improvements that relax the original assumptions of independent populations with the same effective size deriving from an ancestral population have been proposed (Beaumont & Balding 2004; Vitalis *et al.* 2001; Weir & Hill 2002). Recent developments includes (i) the use of a bayesian framework that describes gene frequencies under a wide range of demographic models (BayeScan, Foll & Gaggiotti 2008), (ii) methods that test correlation of allele frequencies to environmental variables and account for population structure by estimating covariance or coancestry among populations (Coop *et al.* 2010; Gautier 2015; Günther & Coop 2013), (iii) the FLK test, an extension of the Lewontin-Krakauer test that integrates covariance of allele frequencies among populations to account for population structure (Bonhomme *et al.* 2010). The latter was extended to a haplotype-based test (hapFLK) that explicitly models non-independence of adjacent SNPs through local clustering (Fariello *et al.* 2013).

The two main drawbacks of these methods, i.e. their lack of power and their elevated rate of false positives, have been abundantly discussed in the literature (Lotterhos & Whitlock 2014; Mita *et al.* 2013). While false negatives are often difficult to control, there are currently two strategies to control for false positives. The first one is to assess consistency of signals among different methods (Mita *et al.* 2013; Narum & Hess 2011). The second is to assess consistency of signals across biological replicates (Conte *et al.* 2012; Kawecki & Ebert 2004; Tiffin & Ross-Ibarra 2014). Haplotype-based methods also clearly improve the ratio of true over false positives (Fariello *et al.* 2013).

In this study, we aimed at discovering the determinants of local adaptation in teosintes at a genome-wide scale. We used two altitudinal gradients to mimic continuous climate variation and considered them as biological replicates to better control for false positives. We performed whole genome sequencing of one highland and one lowland populations in each of the two gradients, and sequenced 20-tagged individuals per population (Tag-seq) at low-depth, thereby mimicking a Pool-seq protocol, but keeping the individual information. First, we assessed the impact of Pool-seq over Tag-seq on allelic frequency estimates. Second, we relied on consistencies across a panel of methods to identify candidate loci for local adaptation. We modified a previously published method, hapFLK (Fariello *et al.* 2013), to incorporate genotype uncertainties in haplotype calling procedure. Third, we compiled current literature to discuss the biological implication of our findings with regards to genes whose functions have been characterized, or genes known to be involved in phenotypic variation in

maize and/or teosintes for relevant traits, i.e. traits potentially involved in adaptation to altitude.

## **II.3 MATERIAL AND METHODS.**

### **II.3.1 Sampling information.**

We collected 31 populations along two altitudinal gradients used as biological replicates. Gradient 1 includes populations originating mostly from the Mexican state of Guerrero, while Gradient 2 was mainly located in the state of Jalisco (Figure S1, Supporting information). Populations were previously described in Muñoz Diez *et al.* (2013). The two gradients ranged between 343 and 2,581 meters, and included both *mexicana* and *parviglumis* subspecies (Table S1, Supporting information). For each population, we recorded altitudinal, latitudinal, longitudinal information (Table S1, Supporting information), and used the values for 19 climatic variables reported in (Hufford *et al.* 2012a; Table S2, Supporting information). Using these variables, we performed for each gradient a Principal Component analysis (PCA) using the FactoMineR package in R (Husson *et al.* 2007). Longitude, latitude and altitude were added to the PCA graphs *a posteriori*.

Based on PCA results we chose a restricted sample of 6 populations to generate whole genome sequencing and perform local adaptation analyses. We sampled 4 “extreme” populations with contrasted climatic conditions, 2 highlands – H1, H2 for Gradient 1 and 2 respectively – and 2 lowlands – L1, L2 for Gradient 1 and 2 respectively (Tables S1 and S2, Supporting information). In order to control for subspecies differentiation, we sampled two additional intermediate populations from Gradient 1: a *mexicana* population – IM1 – and a *parviglumis* population – IP1 (Tables S1 and S2, Supporting information). IM1 corresponded to a lowland *mexicana* and IP1 to a highland *parviglumis*, while both grew at similar altitude, 1669 and 1439 m for IM1 and IP1, respectively. Those populations were sampled at localities from which subspecies were characterized genetically (Fang *et al.* 2012).

### **II.3.2 Sequencing procedure.**

Seeds were chosen randomly from 20 different mother plants within each of the 6 populations. They were sown in a greenhouse at 25°C under natural light, supplemented by artificial light to provide a 12-h photoperiod. 100 mg fresh leaves tissues were sampled and frozen in liquid nitrogen then stored in a freezer at -80°C. For the 4 extreme populations, we

used a NucleoSpin Plant Kit (Macherey-Nagel) to extract and purify total genomic DNA. For each plant of the two intermediate populations, DNAs were extracted from 100 mg of fresh material following recommendations of the DNeasy 96 Plant Kit (QIAGEN, Valencia, CA, USA) manufacturer.

For each of the four extreme populations, 20 bar-coded DNA libraries were constructed by Montpellier GenomiX-Plateform MGX and sequenced using two Illumina HiSeq 2000 lanes per library (8 lanes total) following a paired-end 100-bp protocol. The 20 bar-coded libraries corresponding to the 20 individuals (40 chromosomes) of the same population were pooled prior to sequencing. For the intermediate populations, DNAs from 20 individuals of each population were pooled by Genotoul Genomic platform (Toulouse, France) prior to the construction of a single library per population. DNA content of each sample was measured by picogreen prior to pooling to ensure equimolarity. Each of the 2 libraries was then sequenced on a single Illumina HiSeq lane following a paired-end 100-bp protocol. In summary, the 4 extreme populations were sequenced at a higher depth and individuals were bar-coded (Tag-seq), while the 2 intermediate populations were sequenced in pool (Pool-seq). Each population was represented by twenty individuals (40 chromosomes) grown from 20 mother plants.

### **II.3.3 SNP detection, coverage and depth.**

We aligned all reads (raw data) to the maize reference genome ZmB73 version 2 (Schnable *et al.* 2009) using the program Stampy (Lunter & Goodson 2011) with default parameters. We conserved only properly paired reads and mapped reads with a mapping quality higher than 30 (uniquely mapped reads only). We trimmed bases with sequencing quality inferior to 20% with SAMtools mpileup (Li *et al.* 2009).

We evaluated the total coverage of the maize genome per population as well as the average depth per site. For each of the 4 extreme populations (H1, L1, H2, L2), we determined the individual relative contribution (number of reads per individual) to the sequencing effort (total number of reads for a given library) from raw data, and after mapping and filtering. We assessed significance of differences in individual contribution using a  $X^2$  test (multinomial distribution with 20 classes).

First, we carried out SNP calling from data of the 4 extreme populations separately. We applied one of two filters: we retained either positions covered by at least 12 individuals (SNP<sup>12</sup>) or positions covered by at least 12 reads (SNP<sup>12x</sup>), while filtering out positions

covered by more than 50 reads (50x). Our rationale here was to eliminate potential genome redundancy that may cause false calling of SNPs (upper bound 50), while providing a correct estimate of allele frequency (lower bound 12). Second, we performed SNP calling from data of the 4 extreme populations taken together while applying the 12 individuals-50x filter in each population (SNP<sup>4</sup>). Third, we filtered the SNP<sup>4</sup> positions with at least 12 reads (12x) in both intermediate populations (IM1 and IP1), leading to the SNP<sup>6</sup> dataset. Note that SNP<sup>12</sup> and SNP<sup>12x</sup> positions differed among populations since a given SNP may exist/pass our filters in one population but not in another. In contrast, the number of SNPs in SNP<sup>4</sup> and SNP<sup>6</sup> was limited to common positions, i.e. positions that passed our filters in all populations. We determined the coverage and per-site depth for all SNP datasets.

### **II.3.4 Allele frequency estimation.**

We employed the maximum likelihood approach of BCFtools (Li *et al.* 2009) to estimate allelic frequencies and folded Site Frequency Spectrum (SFS) in each of the 4 extreme populations from SNP<sup>12</sup> and SNP<sup>4</sup>. This methodology indeed required individual information to establish genotype probabilities. We discarded all SNPs with Minor Allele Frequency (MAF) below 5% to account for sequencing errors. In parallel, we established count-based allelic frequency estimations for all SNP datasets by dividing the alternative allele counts by the total number of counts per site (reference + alternative counts). Reference allele referred to the B73 sequence. We obtained corresponding folded SFS with MAF>5% using an R-script.

### **II.3.5 Site-based neutrality tests.**

In order to identify outlier SNPs related to altitudinal adaptation processes, we combined a variety of detection methods (differentiation-based, diversity-based and correlation-based methods) and used the two gradients as biological replicates. We applied all methods to SNP<sup>6</sup> and therefore used count-based allele frequencies estimations except for Pool-HMM (see below).

#### **II.3.5.a Differentiation-based methods.**

Because local adaptation creates genetic differentiation among populations, we computed pairwise differentiation between extreme populations within each gradient (H1-L1, H2-L2) using *Fst* estimates from Weir and Hill (2002; script kindly provided by R. Vitalis). We

retained the 5% highest values (outliers) common to the 2 gradients. This constituted our first set of *Fst*-outliers. As a control for subspecies differentiation, we also computed *Fst*-values for within-subspecies differentiation (H1-IM1 and L1-IP1). Considering that SNPs differentiated between subspecies (H-L comparisons) but not within subspecies may be involved in subspecies differentiation rather than altitudinal adaptation we excluded from the first set of outliers, SNPs not differentiated in the within-subspecies comparisons (95% lowest values of the H1-IM1 and L1-IP1 comparisons). This constituted our *Fst*-outlier set.

We used BayeScan (Foll & Gaggiotti 2008), a bayesian method that decomposes locus-population *Fst* into a population specific component shared by all loci (*beta*), and a locus specific component (*alpha*) shared by all populations. The population-specific component explicitly accounts for demographic effects, while the locus-specific component accounts for potential deviation from neutral expectations. For each locus and all populations taken together, two models were compared, one with the population component only and one in which the locus specific component was added. Superiority of the latter and a corresponding positive value of *alpha* indicated positive selection and its posterior probability is further computed. We applied BayeScan with default settings for priors and a FDR of  $10^{-4}$  on the SNP<sup>6</sup> dataset to define our BayeScan-outliers.

### **II.3.5.b Diversity-based method.**

In order to test for depleted levels of diversity within populations that accompany past and recent selection, we employed PoolHMM (Boitard *et al.* 2013). PoolHMM computes the probability of all possible allele frequencies at each genomic position, based on pool-sequencing data, and integrates over these probabilities to (i) estimate the global SFS and (ii) detect regions where the local SFS is more compatible with a selective sweep than with the global (and presumably neutral) SFS. For the sweep detection, it assumes that each SNP position is associated to one of the 3 states: “neutral”, “intermediate” and “selected”, and that these 3 states are associated to different SFS. These hidden states are estimated from the pool-sequencing data using the Viterbi algorithm. PoolHMM delivers 3 lists of windows of various sizes, neutral- intermediate- and selected-windows. We analysed the 6 populations (H1, L1, H2, L2, IM1, IP1) independently. We used the following settings: -a “unknown”, -c 20, -C 50, -q 20, -t 0.005, -k  $10^{-9}$ . Option -a indicates whether ancestral alleles are known, -c -C -q set the minimum and maximum depth and quality of bases to be analysed, -t defines a starting value for the proportion of polymorphic sites in the genome, which is then updated by an EM

algorithm, and  $-k$  defines transition probability between hidden states. The smallest the  $k$  value, the harder the transitions to the “selected” state. A value of  $10^{-9}$  is stringent with regards to detection of selection.

### **II.3.5.c Correlation-based method.**

Correlation between allele frequencies and environmental variables suggests local adaptation. BAYENV (Coop *et al.* 2010) is a Bayesian approach based on this principle with a recent modification that incorporates pool-sequencing data (Günther & Coop 2013). A covariance matrix of neutral allele frequency is first estimated and used as a null model to control for demographic effects. Second, BAYENV2 computes the posterior probability of a model with an additional environmental factor compared to the null model. To summarize information across environmental variables and reduce multiple testing issues, we performed a PCA on the environmental variables for the 6 populations, and used the first PCA axis as the environmental factor. Because BAYENV2 is computationally intensive, we applied it to SNPs surrounding outliers detected with *Fst* and BayeScan, including 10 SNPs - 5 on each side - for each outlier. We also applied BAYENV2 to all SNPs in regions detected using PoolHMM. We previously estimated the neutral covariance matrix using a set of 117,614 “neutral” control SNPs - not detected by any other method - taken randomly along the genome. We considered the top 1% highest values of Bayes Factor returned by BAYENV2 as selection candidates.

### **II.3.6 Haplotype-based neutrality tests.**

We looked for genomic regions showing large haplotype frequency differences between populations using the hapFLK method (Fariello *et al.*, 2013), which we modified in order to account for low coverage NGS data. HapFLK proceeds in two steps. First, it infers a pair of local haplotype clusters (which can be thought of as ancestral haplotypes) for each genomic position and each sampled individual, using the hidden Markov model of Scheet & Stephens (2006). In a second step, it scans the genome for regions where cluster frequency differences between populations are not consistent with neutral evolution. To do so, it models the neutral history through a kinship matrix  $F$ , which is estimated using genome-wide data and accounts for (i) differences in effective sizes between populations and (ii) possible shared ancestry between populations.

In the original hapFLK approach, local haplotype clusters are inferred from the genotype data observed in all individuals. However, in our study, individuals from H1, H2, L1 and L2 were sequenced at very low coverage, which resulted in a very high uncertainty concerning the genotypes. Therefore, we modified the hidden markov model of Scheet & Stephens (2006) in order to account for this uncertainty: when computing the emission probabilities of the model, we summed over all possible genotype pairs for each individual and genomic position, weighting these terms using the genotype probabilities provided by samtools.

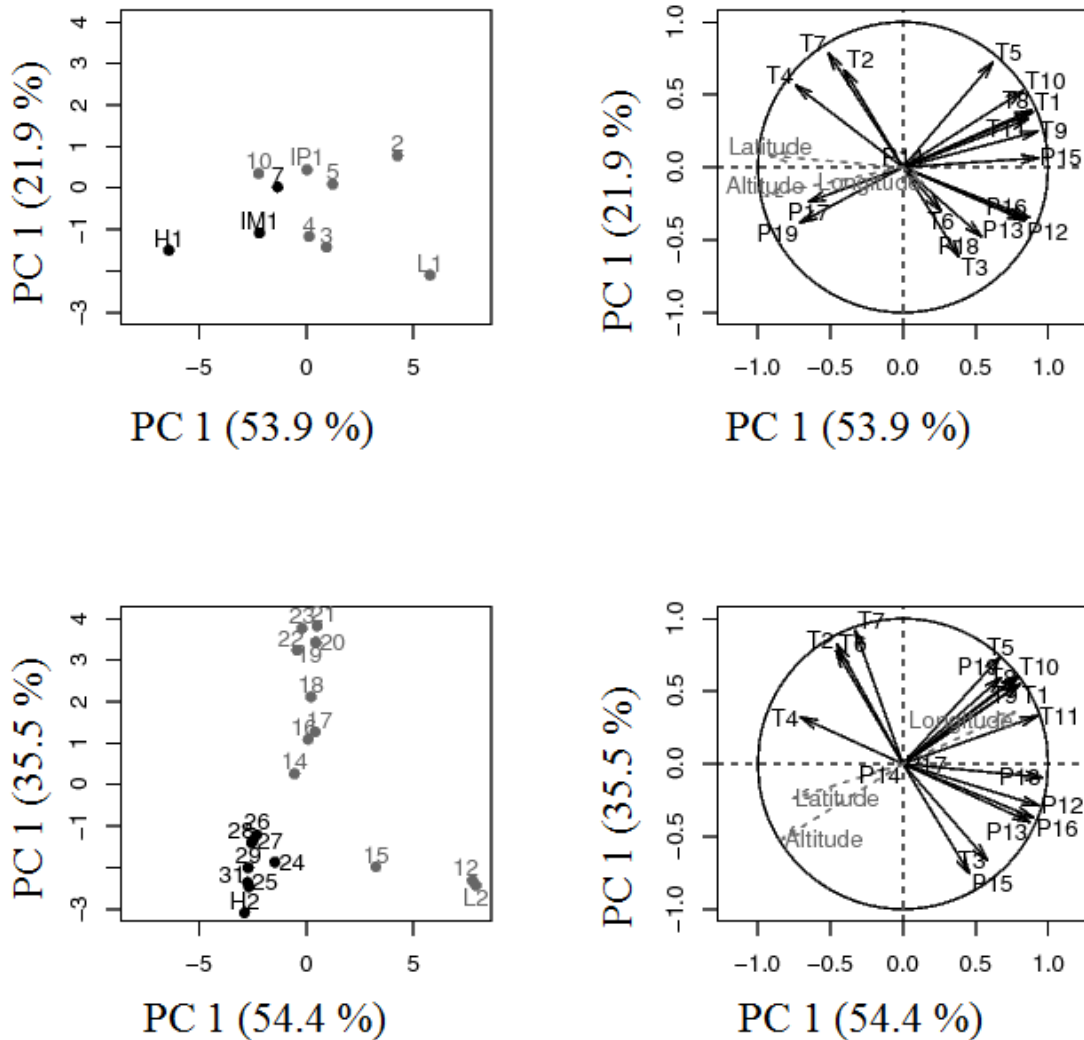
We applied this modified approach to Tag-seq data from H1, H2, L1 and L2. We estimated the kinship matrix using allele frequencies of SNP<sup>4</sup>. We then looked for selection signatures in candidate regions detected by site-based neutrality tests. We did not perform a complete genome scan, because the modifications brought to hapFLK made it computationally very intensive. For this analysis, we used 8 haplotype clusters per genomic position and inferred these clusters using 10 independent runs of the EM algorithm (see Fariello *et al.* 2013 for more details). We computed P-values (both with and without fixing degrees of freedom) by fitting a chi-square distribution to the empirical hapFLK values obtained from all analysed regions, using the code provided at <https://forge-dga.jouy.inra.fr/projects/hapflk/documents>.

## **II.4 RESULTS.**

### **II.4.1 Sampling 6 populations based on environmental analyses.**

The aim of our study was to investigate the genetic basis of local adaptation to altitude in teosintes, using 31 previously sampled natural populations (Table S1 and S2, Supporting information) along 2 altitudinal gradients (Figure S1, Supporting information) used as biological replicates. First a restricted number of populations was chosen for full genome sequencing, based on both seed availability (>20 mother plants per population) and PCA analyses on climatic variables. We aimed at sequencing 2 extreme populations from each gradient, one lowland (coded as L1 and L2 for gradients 1 and 2, respectively) and one highland (H1 and H2 for gradients 1 and 2, respectively). Because elevation was fully confounded with subspecies origin (lowlands being from the subspecies *parviglumis* and highlands from the subspecies *mexicana*), we sampled two additional populations at middle elevation (hereafter intermediate populations) along gradient 1, IP1 (a *parviglumis* of high

elevation given the subspecies range) and IM1 (a *mexicana* of low elevation given the subspecies range).



**Figure 1: Principal Component Analysis based on 19 climate variables on 11 teosinte populations sampled on gradient 1 (A and B) and 20 populations sampled on gradient 2 (C and D).** A, C: Projection of populations on the first PCA plan. *mexicana* populations are represented in black, *parviglumis* populations in grey. The four Extreme and two Intermediate populations are indicated as H1, L1, H2, L2, IM1, IP1. B, D: Correlation circles on the first two PCA axes. Population names and locations are described in Table S1, Supporting information. T-climatic variable are related to temperature, P-climatic variable are related to precipitations (Table S2, Supporting information).

The PCA analysis was performed for each gradient separately using 19 climatic variables (Table S2, Supporting information). The two first axes (PC1 and PC2) explained a large proportion of the variation both for gradient 1 (75.8%, Figure 1-A) and gradient 2 (89.5%, Figure 1-C). For Gradient 1 and 2, PC1 was linked to both temperature and precipitation, and PC2 mainly to temperature (Figure 1-B, 1-D). Among geographical



variables, altitude was the most correlated to PC1 (Gradient 1: -0.95 Gradient 2 : -0.83, Figures 1 -B and -D). We chose extreme populations (H1 and L1 from gradient 1 and H2 and L2 from Gradient 2, Figure 1-A, C) as the mostly differentiated. Along gradient 1, we chose the *parviglumis* and *mexicana* intermediate populations with close geographical locations and similar environments and altitude (1669 m and 1439 m for IM1 and IP1, respectively, Figure 1A).

## **II.4.2 Sequencing statistics.**

We applied different sequencing protocols for the extreme and intermediate populations. For the extreme populations, we tagged all individuals and sequenced them in pool using 2 lanes (Tag-seq). For the intermediate populations we performed pool-sequencing (without tags) in a single lane (Pool-seq).

For each population, we obtained roughly 313 to 437 million reads using Tag-seq and around 195 million reads using Pool-seq (Table 1), the differences being explained by the sequencing volume, i.e. 2 lanes in Tag-seq versus a single lane for Pool-seq. We assessed the relative contribution of each individual to the sequencing effort using the Tag-seq datasets (Figure S2, Supporting information). We found significant differences of the number of reads across individuals in all 4 populations ( $P\text{-value} < 2.10^{-16}$ ) despite attempt to obtain equimolarity when pooling the tagged samples.

**Table 1: Sequencing statistics for 6 populations**

Population	# reads <sup>a</sup>	Coverage <sup>b</sup> (%)	Depth <sup>c</sup>	mapped reads (%)
H1	377,317,315	58.83	22.1	35.1
L1	312,979,394	59.33	22.9	36.2
H2	317,647,367	55.73	21.5	43.3
L2	437,425,261	53.88	26.6	35.6
IM1	196,271,307	54.14	17.6	35.8
IP1	194,443,385	60.89	16.6	35.0

<sup>a</sup>Total Number of paired-end reads

<sup>b</sup>Coverage of maize reference genome V2

<sup>c</sup>Sequencing depth per site after mapping, filtering on base- & mapping- quality, and filtering on 12x-50x (SNP<sup>12x</sup>).

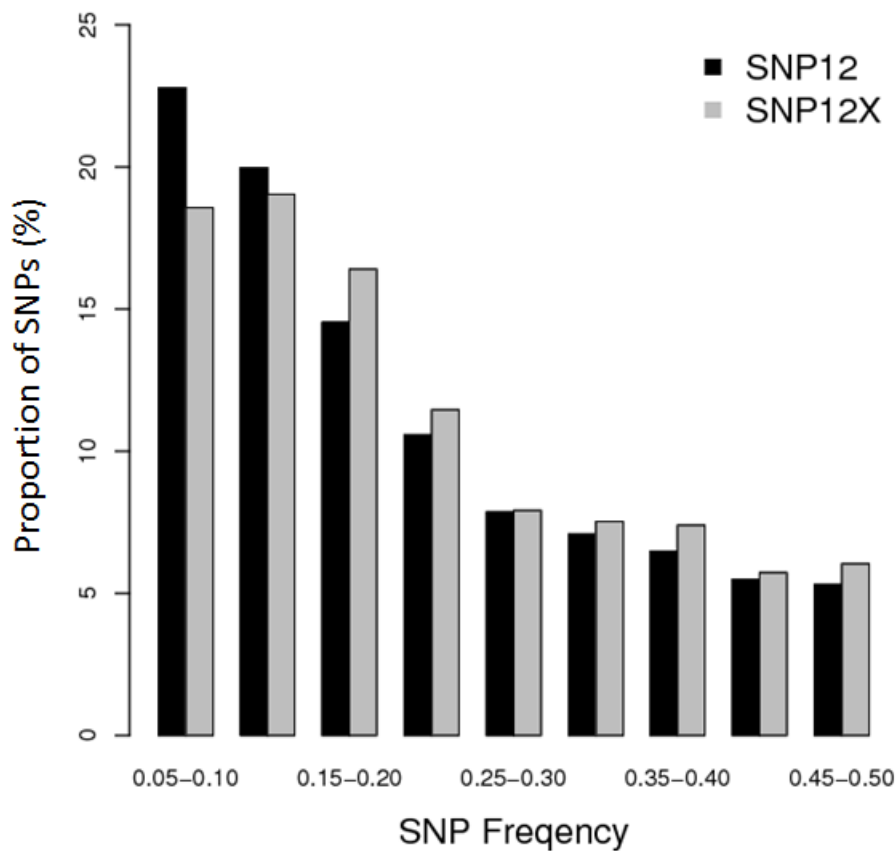
We mapped reads to the maize reference genome and applied base- and mapping-quality filters. The coverage percentage of the maize reference genome per population ranged from 53.9% to 60.9% (Table 1). Differences in coverage may be attributed to both sequencing quality and proximity with the reference. Note that *parviglumis* populations, the subspecies from which maize was domesticated, exhibited a slightly greater coverage (58.5% on average for L1, L2, IP1) than *mexicana* (56.3%). Variance in average depth per site per population across libraries was wide (Figure S3, Supporting information). After filtering, it ranged from 21.5 to 26.6 for Tag-seq and from 16.6 to 17.6 for Pool-seq. Note that differences in the number of reads across individuals in all 4 extreme populations remained significant after filtering (P-value  $< 2.10^{-16}$ , Figure S2, Supporting information).

Altogether we assembled 4 sets of SNPs: SNP<sup>12</sup>, SNP<sup>12x</sup>, SNP<sup>4</sup>, and SNP<sup>6</sup>. For each extreme population, SNP<sup>12</sup> included SNPs covered by at least 12 individuals and less than 50x depth, SNP<sup>12x</sup> included SNPs covered by a 12x-50x range depth. SNP<sup>4</sup> corresponded to common SNPs to the 4 extreme populations covered by at least 12 individuals and less than 50x. SNP<sup>6</sup> was a subset of SNP<sup>4</sup> for which SNPs were also covered in intermediate populations in a range of depth of 12x-50x. For the four extreme populations, we called from 12 to 15 million SNPs per population for SNP<sup>12</sup>, and from 22 to 27 million SNPs per population for SNP<sup>12x</sup>. The distribution of the number of individuals covering SNP positions in Tag-seq extreme populations is shown in Figure S4 (Supporting information). Median value of those distributions was 12 except for L2 (13 individuals) for which the greater number of reads sequenced (Table 1) translated into greater number of individuals covering each site (Figure S4, Supporting information). We observed that by applying a criterion of 12x-50x (see above), 95% of SNPs were covered by 7 to 17 individuals for populations H1 and H2, and by 7 to 18 individuals for populations L1 and L2. Because sequencing depth was lower in intermediate populations, by applying the same criterion of 12x-50x we expected the distribution of position coverage by individuals to be biased towards lower values obtained than those obtained from Tag-seq. SNP<sup>4</sup> and SNP<sup>6</sup> datasets encompassed 9,664,868 and 8,479,581 SNPs respectively.

### **II.4.3 Allele frequencies and differentiation between populations.**

We established folded SFS from count-based allelic frequencies estimations for SNP<sup>12</sup> and SNP<sup>12x</sup>. We found a deficit of rare alleles for SNP<sup>12x</sup> compared to SNP<sup>12</sup> in population H1 (Figure 2). A similar signal was observed in the three other populations (Figure S5,

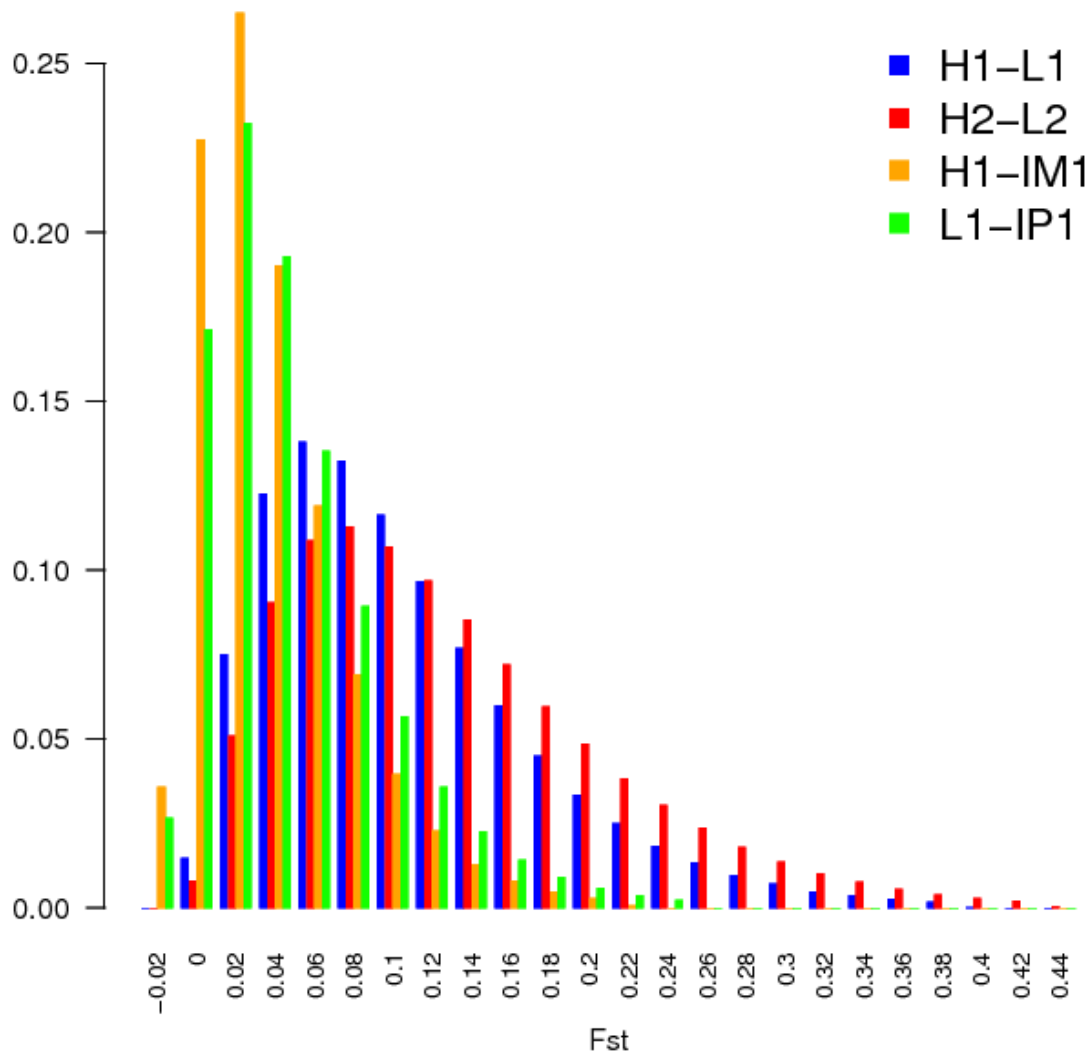
Supporting information). By filtering on a minimum depth threshold of 12x (SNP<sup>12x</sup>) rather than on a minimum number of 12 individuals (SNP<sup>12</sup>), we included estimations of allele frequencies at sites with a restricted number of individuals (from 1 to 11), leading to underestimation of rare alleles from SNP<sup>12x</sup>. Unequal contributions of individuals to sequencing (Figure S2, Supporting information) likely strengthened this effect. It was interesting to note that biases in the detection of rare alleles was less pronounced in the population L2 as compared to other populations (Figure S5, Supporting information), as expected given its greater individual coverage (Figure S4, Supporting information).



**Figure 2: SFS for SNP<sup>12</sup> and SNP<sup>12x</sup> with count-based estimates of allele frequencies in population H1.** SNPs with MAF<0.05 were discarded.

We compared SFS from our count-based allele frequencies estimates to BCF-tools estimates for SNP<sup>12</sup>. We observed a deficit of rare variants with BCF-tools estimates in all populations (Figure S6, Supporting information). Prior to estimate allele frequencies, BCF-tools assigns the most probable genotype to each individual. Since the per population coverage was around 23x after filtering, resulting in an average 1x coverage per individual, genotype inferences likely lead to underestimating rare variants.

Finally we evaluated the population-diversity (Nei 1973), and  $F_{st}$  distributions between pairs of populations in order to evaluate differentiation between subspecies within gradients (H1-L1, H2-L2) and within subspecies within Gradient 1 (H1-IM1, L1-IP1). We averaged diversity and  $F_{st}$  values across sliding windows of 20 SNPs with a step of 5 SNPs. We found similar diversity estimates (from 0.10 to 0.15) in all six populations. As expected (Figure 3), we observed a greater differentiation between subspecies (average  $F_{st_{H1-L1}} = 0.119$ , average  $F_{st_{H2-L2}} = 0.122$ ) than within subspecies (average  $F_{st_{H1-IM1}} = 0.048$ , average  $F_{st_{L1-IP1}} = 0.058$ ).



**Figure 3: Distributions of  $F_{st}$  between pairs of populations as evaluated on 8,479,581 SNPs (SNP<sup>6</sup> dataset).** For better visualization, distributions were truncated since >99% of the  $F_{st}$  values were comprised between -0.02 and 0.44.

#### **II.4.4 Combining genome-scan statistics to identify candidate loci.**

Several patterns are expected at loci involved in local adaptation processes, including enhanced differentiation between populations facing contrasting environmental challenges, depletion of genetic diversity within populations, and correlated allele frequencies with environmental variables driving selection. To detect such patterns and identify outlier loci contributing to local adaptation, we used *Fst*-scan and BayeScan to detect differentiation signals between populations, Pool-HMM to detect selective sweeps within populations, and BAYENV2 to associate allele frequencies to environmental gradients.

Using *Fst*-scan, we identified 423,983 outlier SNPs for Gradient 1 and 423,991 outliers SNP for Gradient 2. Among these, roughly 20% (86,148) were common to the two gradients. Because a number of these common loci may arise from subspecies divergence rather than local adaptation, we used results from intermediate populations – a lowland *mexicana* and a highland *parviglumis* – as controls. Hence, from the Gradient 1 outliers, we removed 407,068 SNPs that were not differentiated between populations growing at different elevation within subspecies (95% *Fst* values of H1-IM1 and L1-IP1 comparisons). Using the same rationale, we discarded 82,486 SNPs from the outlier SNPs of gradient 2. In total, we identified 3,662 outlier SNPs common to the two gradients that did not exhibit a pattern of subspecies differentiation (Table 2).

We also used a bayesian computation *Fst* approach that accounts for population structure on the six populations taken together (BayeScan). 1,028 SNPs were found to be under directional selection at an FDR of  $10^{-4}$  (Table 2). These SNPs exhibited a mean locus-specific component *alpha* of 2.156, the mean locus-specific component *alpha* for all other loci being 0.001. With Pool-HMM, we detected 53, 87, 51, 152, 50, 50 selective sweeps for populations H1, L1, H2, L2, IM1, IP1 respectively. Overall, Pool-HMM results revealed only 4 congruent selective sweeps among populations, two for the highland populations (H1 and H2) located on chromosomes 6 and 8, and two for the lowland populations on chromosomes 1 and 8.

**Table 2: Number of SNPs detected for each method and in combination with BAYENV2**

Method	# SNPs detected <sup>a</sup>	# SNPs detected with BAYENV2 <sup>b</sup>
Bayescan	1028 (596)	17 (1.65%)
<i>Fst</i>	3662 (3228)	55 (2.07%)
PoolHMM	50 windows, 161371 (80302)	229 (0.14%)
None	-	58 (0.05%)

<sup>a</sup> Number of SNPs detected using a single method, in bracket the number of genic SNPs

<sup>b</sup> In bracket the percent of SNPs significant with BAYENV2 among those detected by each method

*Fst* and BayeScan identified outliers from empirical distributions and cut-off values, with two main drawbacks. The first one was that because all distributions had values in their tails, a number of outliers may be false positives. The second, not independent from the first one, was that because we employed distinct cut-off values for each method (including PoolHMM) we were not able to assess congruence among methods. Such congruence however may be useful to actually distinguish false from true positives (Savolainen *et al.* 2013). In this context, we either considered consistency of signals within/across methods in a given genomic region as evidence for selection, or we conditioned our results from each of the 3 methods, *Fst*, BayeScan, and Pool-hmm, on BAYENV2 results. Our choice on BAYENV2 was motivated by (1) its power over other methods (Mita *et al.* 2013), and (2) the use of environmental variables that explicitly tests the role of environment on local adaptation (Mita *et al.* 2013).

We applied BAYENV2 to the complete set of 166,061 outlier SNPs detected by other methods and their surrounding regions (238,003 SNPs), as well as 117,614 control SNPs. We employed the PC1 on climatic variables run on all 6 populations as the correlative environmental variable. PC1 explained a large part of the variance (69.6%). We detected 359 SNPs using the BAYENV2 method (Table 2, Table S3, Supporting information) among which 17, 55, and 229 were first detected by *Fst*, BayeScan, and PoolHMM, respectively, and 58 were detected by BAYENV2 only. Proportionally, *Fst* and BayeScan had a greater congruence (2.07% and 1.65%) with BAYENV2 than Pool-HMM (0.14%, Table 2).

We determined the genic/non genic status of the complete set of outlier SNPs. We considered as genic the regions with an annotated gene within 10kb of the SNP. We found 88% of genic SNPs among those detected by *Fst* and 58% among those detected by BayeScan. As comparison, the SNP<sup>6</sup> dataset contained 58% genic SNPs. We found 29% of genic SNPs among those detected by PoolHMM (Table 2).

We compiled the literature to obtain a list of regions known to be associated with phenotypic variation in maize and/or teosintes (QTL and association mapping studies), as well as a list of genes that have been functionally validated. We focused on traits of adaptive relevance for our study including biotic and abiotic factors that may co-vary with altitude such as drought tolerance, water logging tolerance, pests and disease resistance, heavy metal tolerance. We also included regions/genes involved in phenotypic variation of traits that may be involved in adaptive responses including fitness-related traits, plant and ear architecture, flowering time. Finally we considered co-localization with the *Inv1* inversion that covaries with altitude (Fang *et al.* 2012; Pyhäjärvi *et al.* 2013).

Among the 359 SNPs detected with BAYENV2 (Table 2), 221 were located within 29 regions highlighted in the literature as potentially involved in adaptation or variation of agronomic traits (Table 3). We reported associated traits, discovery method and detected 37 annotated genes encompassing at least one of the detected SNPs in each region (Table 3). In addition, we defined candidate regions as groups of outliers distant by less than 10 kb when considered pairwise and detected 39 candidate regions encompassing between 10 and 61 outlier SNPs (Table 4). Three of these regions were detected by a combination of SNPs significant with *Fst* and BayeScan but the vast majority (36 out of 39) exhibited a consistent signal with only one method: *Fst* detected 26 regions, BayeScan, 9, while BAYENV2 detected only one region (Table 4). Because we could not access individual SNP signal with Pool-HMM, we focused on the consistency of signal across the 2 highlands or the 2 lowlands populations, and identified 4 additional candidate regions. Among the total 43 candidate regions, 23 were comprised in regions reported in our literature compilation, and some regions also encompassed SNPs belonging to annotated genes (Table 4).

**Table 3: Co-localization of BAYENV2 outliers with previously described candidate genes/regions and annotation of genes encompassing outliers.**

ID Region	Chro <sup>1</sup>	# of SNPs (ID) <sup>2</sup>	Position	Genes-SNP	Locus, gene function	Trait	Detection methods	Reference
1	1	8 (15, 16, 17, 18, 19, 20, 21, 22)	65,000,000-115,000,000	GRMZM2G137561** GRMZM2G424595** GRMZM2G149903**	Inversion	-	molecular evolution study	(Fang <i>et al.</i> 2012) (Pyhäjärvi <i>et al.</i> 2013)
2	1	8 (29, 30, 31, 32, 33, 34, 35, 36)	154,138,515-180,709,507	-	-	Grain yield	QTL mapping	(Malosetti <i>et al.</i> 2008)
3	1	9 (37, 38, 39, 40, 41, 42, 43, 44, 45)	195,068,415-209,513,848 203,700,000-240,600,000 194,300,000-250,900,000	GRMZM2G159975+ GRMZM2G388892+ AC195164.3_FG002 AC192362.3_FG002	-	Low phosphate Plant height Root morphology	QTL mapping	(Malosetti <i>et al.</i> 2008) (Azevedo <i>et al.</i> 2015) (Chen <i>et al.</i> 2014)
4	1	1 (46)	246,787,256-263,079,095	GRMZM2G088413	<i>abi24 - ABI3-VP1-transcription factor 24</i>	Tassel branch number Anthesis-silking interval	QTL mapping	(Malosetti <i>et al.</i> 2008)
5	2	14 (63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76)	66,240,000-119,270,000			Low phosphate tolerance	QTM meta-analysis	(Zhang <i>et al.</i> 2014)
6	2	5 (87, 88, 89, 90, 91)	152,207,394-163,405,828	-	-	Grain yield, Nitrogen use efficiency	QTL mapping	(Coque & Gallais 2006)
7	3	10 (100, 101, 102, 103, 104, 105, 106, 107, 108, 109)	45,519,051-95,402,898	GRMZM2G081380 GRMZM2G108527	-	Number of ears per plant,	QTL mapping	(Malosetti <i>et al.</i> 2008)



8	3	6 (110, 111, 112, 113, 114, 115)	86,754,041-118,147,511	-	-	Plant height Low phosphorus tolerance	QTL mapping	(Reiter <i>et al.</i> 1991)
9	3	4 (116, 117, 118, 119)	216,256,506-232,072,064	GRMZM2G114930*+ GRMZM2G098434+	<i>Polygalacturonate 4-alpha-galacturonosyltransferase</i>	Pigment coverage and intensity	QTL mapping	(Lauter, Gustus <i>et al.</i> 2004)
10	3	2 (119, 120)	12,657,6367-168,366,251	-	<i>scmv2 -(resistance to sugarcane mosaic virus)</i>	Resistance to sugarcane mosaic virus disease	Cloning and genetiQ mapping QTL mapping	(Xia <i>et al.</i> 1999) (Leng <i>et al.</i> 2015; Quint <i>et al.</i> 2002)
11	4	9 (129, 130, 131, 132, 133, 134, 135, 136, 137)	17,851,597-118,237,095 33,900,000-59,400,000	GRMZM2G105387*	-	Plant height  Low phosphate tolerance, Root morphology	QTL mapping	(Malosetti <i>et al.</i> 2008) (Azevedo <i>et al.</i> 2015)
12	4	2 (139, 140)	171,670,000-178,110,000			Low phosphate tolerance	QTL meta analysis	(Zhang <i>et al.</i> 2014)
13	4	1 (141)	182,200,000-188,100,000	-	-	Ear length	QTL mapping	(Chen <i>et al.</i> 2014)
14	4	16 (144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159)	204,706,420-235,644,775	GRMZM2G057402*+	-	Grain yield, Days to male flowering	QTL mapping	(Malosetti <i>et al.</i> 2008)
15	5	2 (169, 170)	302,4418-43,661,111 10,207,280-	AC213463.3_FG006	-	Drought tolerance	QTL mapping	(Agrama & Moussa 1996) (Lauter <i>et</i>

			78,362,776				Macrohairs density		<i>al.</i> 2004)
16	5	2 (186, 187)	120,680,000-147,230,000				Low phosphate tolerance	QTL meta analysis	(Zhang <i>et al.</i> 2014)
17	5	1 (194)	175,330,000-180,150,000	GRMZM2G024693*			Low phosphate tolerance	QTL meta analysis	(Zhang <i>et al.</i> 2014)
18	6	16 (195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210)	263,968-81,880,425	GRMZM2G025214** GRMZM2G319104* GRMZM2G062638	<i>scmv1</i> -(resistance to sugarcane mosaic virus1)		Resistance to sugarcane mosaic virus1 disease	Cloning and genetiQ mapping QTL mapping	(Xia <i>et al.</i> 1999) (Quint <i>et al.</i> 2002)
19	6	6 (211, 212, 213, 214, 215, 216)	88,248,022-131,092,297	GRMZM2G171317** GRMZM2G342588+	<i>alm2</i> -(aluminum tolerance2)		Aluminium tolerance	molecular mapping	(Sibov <i>et al.</i> 1999)
20	6	6 (213, 214, 215, 216, 217, 218)	104,450,983-164,802,199	GRMZM2G342588+	-		Anthesis-silking interval Number of ears, Drought tolerance Low phosphate tolerance	QTL mapping QTL meta analysis	(Agrama & Moussa 1996) (Zhang <i>et al.</i> 2014)
21	7	14 (264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277)	143,552,212-163,011,802 150,760,000-156,160,000	GRMZM2G453919 GRMZM2G054048** GRMZM2G316191**	-		Number of ears per plant, Plant height	QTL mapping	(Malosetti <i>et al.</i> 2008)
22	7	32 (232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254,	50,173,540-129,831,204 98,100,000-123,600,000	GRMZM2G453919 GRMZM2G054048**	-		Plant height, Root length, Water logging	QTL mapping	(Qiu <i>et al.</i> 2007) (Azevedo <i>et al.</i> 2015)

		255, 256, 257, 258, 259, 260, 261, 262, 263)					tolerance Low phosphate tolerance, Root morphology		
23	8	4 (279, 280, 281, 282)	0-7,600,000	GRMZM2G081127** GRMZM2G085038**	BP : protein phosphorylation		Tassel branch number	QTL mapping	(Chen <i>et al.</i> 2014)
24	8	18 (283, 284, 285, 286, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300)	2,238,0791-14,195,6453	GRMZM2G115963	-		Grain yield, Drought tolerance, Water deficit stress	QTL mapping	(Agrama & Moussa 1996) (Welcker <i>et al.</i> 2011) (Coque & Gallais 2006)
							Nitrogen use efficiency		
25	9	1 (306)	16,979,346-18,607,113	-	-		Number of ears per plant	QTL mapping	(Malosetti <i>et al.</i> 2008)
26	9	2 (307, 308)	44,400,000-74,900,000	GRMZM2G041732+	<i>Nucleoside-triphosphatase</i>		Tassel branch number	QTL mapping	(Chen <i>et al.</i> 2014)
27	9	9 (312, 313, 314, 315, 316, 317, 318, 319, 320)	126,906,736-130,624,535	GRMZM2G170870** GRMZM2G170812 GRMZM2G448687+ GRMZM2G017460** GRMZM2G007922+	<i>50S ribosomal protein L6-2</i>		Shoot dry weight, Root dry weight, Total dry weight Water logging tolerance	QTL mapping	(Qiu <i>et al.</i> 2007)
28	10	2 (346, 347)	62,100,000-77,500,000	-	-		Low phosphate tolerance,		(Azevedo <i>et al.</i> 2015)

29	10	11 (348, 349, 350, 351, 352, 353, 354, 355, 356, 357, 358)	93,081,741- 127,514,254	GRMZM2G001265**	-	Root morphology Macrohairs density	QTL mapping	(Lauter <i>et al.</i> 2004)
----	----	--	----------------------------	-----------------	---	---	-------------	-----------------------------

<sup>1</sup> Chromosome

<sup>2</sup> Number of SNPs encompassed in the window (SNP ID)

\* Gene in common with Takuno *et al.* (2015)

+ Gene in common with Pyhäjärvi *et al.* (2013)

**Table 4: Candidate adaptive regions with gene annotation. In bold are regions presenting evidence of selection with HapFLK.**

Chro	Window <sup>1</sup>	#SNP <sup>2</sup>	Gene <sup>3</sup>	Trait <sup>4</sup>	Detection methods	Reference
<b>1</b>	<b>114,054,845-114,105,789</b>	<b>20 (0/0/20)</b>	-	<b>Inversion</b>	<b>molecular evolution study</b>	<b>(Fang <i>et al.</i> 2012) (Pyhäjärvi <i>et al.</i> 2013)</b>
1	149,172,778-149,201,068	21(0/0/21)	-	-	-	-
<b>1</b>	<b>168,348,856-168,356,302</b>	<b>14 (14/0/0)</b>	-	<b>Grain yield</b>	<b>QTL mapping</b>	<b>(Malosetti <i>et al.</i> 2008)</b>
2	77,056,232-77,057,195	16 (16/0/0)	-	Phosphorus stress	QTL meta analyses	(Zhang <i>et al.</i> 2014)
2	<b>189,963,637-189,968,647</b>	<b>15 (0/0/15)</b>	<b>GRMZM2G041633<sup>+</sup></b>	<b>MF: nucleic acid binding</b>	-	-
3	<b>6,526,573-6,526,823</b>	<b>16 (16/0/0)</b>	<b>GRMZM2G115834</b>	-	-	-
3	<b>122,595,775-122,599,340</b>	<b>28 (0/0/28)</b>	-	-	-	-
3	<b>122,626,404-122,630,018</b>	<b>61 (2/0/59)</b>	-	-	-	-
3	<b>122,683,118-122,689,852</b>	<b>20 (0/0/20)</b>	-	-	-	-
3	193,060,649-193,066,874	13 (0/0/13)	-	Flowering time Number of ears Drought tolerance Grain yield, Low phosphate tolerance, root morphology	QTL mapping	(Agrama & Moussa 1996)     (Azevedo <i>et al.</i> 2015)
3	193,078,594-193,120,272	38 (1/0/37)	GRMZM2G071441 <sup>+</sup>	Flowering time Number of ears Drought tolerance Grain yield, Low phosphate tolerance, root morphology	QTL mapping	(Agrama & Moussa 1996)     (Azevedo <i>et al.</i> 2015)

3	193,232,722- 193,244,296	11 (0/0/11)	-	Flowering time Number of ears Drought tolerance Grain yield, Low phosphate tolerance, root morphology	QTL mapping	(Agrama & Moussa 1996)     (Azevedo <i>et al.</i> 2015)
4	<b>51,892,191- 51,927,079</b>	<b>31 (0/0/31)</b>	<b>GRMZM2G132371*</b>	<b>kinesin-related protein8 plant height Low phosphate tolerance Root morphology</b>	<b>QTL mapping</b>	<b>(Malosetti <i>et al.</i> 2008) (Azevedo <i>et al.</i> 2015)</b>
5	46,285,309- 46,296,525	11 (0/0/11)	-	Macrohairs density	QTL mapping	(Lauter <i>et al.</i> 2004)
5	<b>148,681,005- 148,701,291</b>	<b>15 (0/0/15)</b>	<b>GRMZM2G424519**+</b>	<b>Water deficit</b>	<b>QTL mapping</b>	<b>(Welcker <i>et al.</i> 2011)</b>
6	<b>51,253,194- 51,273,315</b>	<b>12 (0/0/12)</b>	-	-		
6	<b>55,879,137- 55,892,210</b>	<b>15 (0/0/15)</b>	<b>GRMZM2G104958**+</b>			
6	<b>131,031,608- 131,037,841</b>	<b>19 (0/0/19)</b>	<b>GRMZM2G133631 <i>alm2</i></b>	<b>Aluminium tolerance</b>	<b>molecular mapping</b>	<b>(Sibov <i>et al.</i> 1999)</b>
6	<b>131,063,810- 131,067,504</b>	<b>18 (0/0/18)</b>	<b><i>alm2</i></b>	<b>Aluminium tolerance</b>	<b>molecular mapping</b>	<b>(Sibov <i>et al.</i> 1999)</b>
7	37,252,243- 37,256,811	11 (0/0/11)	GRMZM2G177539**+	Tassel branch number	QTL mapping	(Chen <i>et al.</i> 2014)
7	37,272,195- 37,277,185	21 (0/0/21)	-	Tassel branch number	QTL mapping	(Chen <i>et al.</i> 2014)
7	119,332,561- 119,337,753	18 (0/0/18)	-	water deficit stress Low phosphate tolerance, Root morphology	QTL mapping	(Welcker <i>et al.</i> 2011) (Azevedo <i>et al.</i> 2015)
7	<b>119,633,030- 119,641,785</b>	<b>11 (0/0/11)</b>	<b>GRMZM2G138689</b>	<b>water deficit stress Low phosphate tolerance, Root morphology</b>	<b>QTL mapping</b>	<b>(Welcker <i>et al.</i> 2011) (Azevedo <i>et al.</i> 2015)</b>

7	120,173,488- 120,180,107	16 (0/0/16)	-	water deficit stress Low phosphate tolerance, Root morphology	QTL mapping	(Welcker <i>et al.</i> 2011) (Azevedo <i>et al.</i> 2015)
7	150,188,731- 150,192,479	11 (0/11/0)	-	Number of ears per plant Plant height	QTL mapping	(Malosetti <i>et al.</i> 2008)
8	9,288-21,534	25 (24/0/1)	GRMZM2G122853	Peroxydase BP: response to oxydative stress tassel branch number	QTL mapping	(Chen <i>et al.</i> 2014)
8	78,316-85,270	14 (14/0/0)	-	Tassel branch number	QTL mapping	(Chen <i>et al.</i> 2014)
8	258,666-264,396	12 (12/0/0)	-	Tassel branch number	QTL mapping	(Chen <i>et al.</i> 2014)
8	2,262,952-2,266,809	12 (12/0/0)	-	Tassel branch number	QTL mapping	(Chen <i>et al.</i> 2014)
8	28,694,837- 28,699,799	24 (24/0/0)	GRMZM2G379286	Grain yield drought tolerance	QTL mapping	(Agrama & Moussa 1996)
8	28,760,032- 28,767,192	23 (23/0/0)	GRMZM2G107205	Grain yield Drought tolerance BP: post-translational protein modification regulation of protein metabolic process	QTL mapping	(Agrama & Moussa 1996)
8	28,842,539- 28,855,944	11 (11/0/0)	GRMZM2G122003*	Grain yield drought tolerance	QTL mapping	(Agrama & Moussa 1996)
8	40,947,418- 40,950,270	19 (0/0/19)	GRMZM2G006463+	Grain yield Drought tolerance Nitrogen use efficiency	QTL mapping	(Agrama & Moussa 1996) (Welcker <i>et al.</i> 2011) (Coque & Gallais 2006)
8	127,559,577- 127,579,554	19 (0/0/19)	GRMZM2G085218	Grain yield Drought tolerance Low phosphate tolerance, Root morphology	QTL mapping	(Agrama & Moussa 1996) (Azevedo <i>et al.</i> 2015)
9	105,521,549- 105,529,739	10 (0/0/10)	GRMZM2G085246* <i>macrohairless1</i>	Macrohairs density	gene mapping	(Moose <i>et al.</i> 2004)
10	19,926,647-	14 (0/0/14)	-	-	-	-

	19,962,369					
10	69,976,872- 69,987,170	10 (0/0/10)	-	Low phosphate tolerance, Root morphology	-	(Azevedo <i>et al.</i> 2015)
<b>10</b>	<b>101,033,106- 101,050,457</b>	<b>30 (0/0/30)</b>	<b>GRMZM2G159330*</b>	<b>Macrohairs density</b>	<b>QTL mapping</b>	<b>(Lauter <i>et al.</i> 2004)</b>
<b>10</b>	<b>101,110,424- 101,123,358</b>	<b>12 (0/0/12)</b>	-	<b>Macrohairs density</b>	<b>QTL mapping</b>	<b>(Lauter <i>et al.</i> 2004)</b>
1	122,277,664- 122,302,692	Pool-hmm L	-	-	-	-
6	118,875,048- 119,010,240	Pool-hmm H	<i>alm2</i>	Aluminium tolerance	molecular mapping	(Sibov <i>et al.</i> 1999)
8	51,806,626- 51,834,842	Pool-hmm H	-	Grain yield Drought tolerance	QTL mapping	(Agrama & Moussa 1996) (Welcker <i>et al.</i> 2011)
8	94,622,226- 94,668,629	Pool-hmm L	-	Nitrogen use efficiency Number of branch in lateral inflorescence Grain yield Drought tolerance	QTL mapping	(Coque & Gallais 2006) (Doebley & Stec 1993) (Agrama & Moussa 1996)

<sup>1</sup> Position on Maize reference V2

<sup>2</sup> Total number of SNPs detected significant with (BayeScan/BAYENV2/Fst-scan)

<sup>3</sup> Gene annotation on B73 version 2.

<sup>4</sup> Gene function as reported in MaizeGDB, BP : Biologic Process, MF : molecular function. QTLs reported in literature are also cited with associated references.

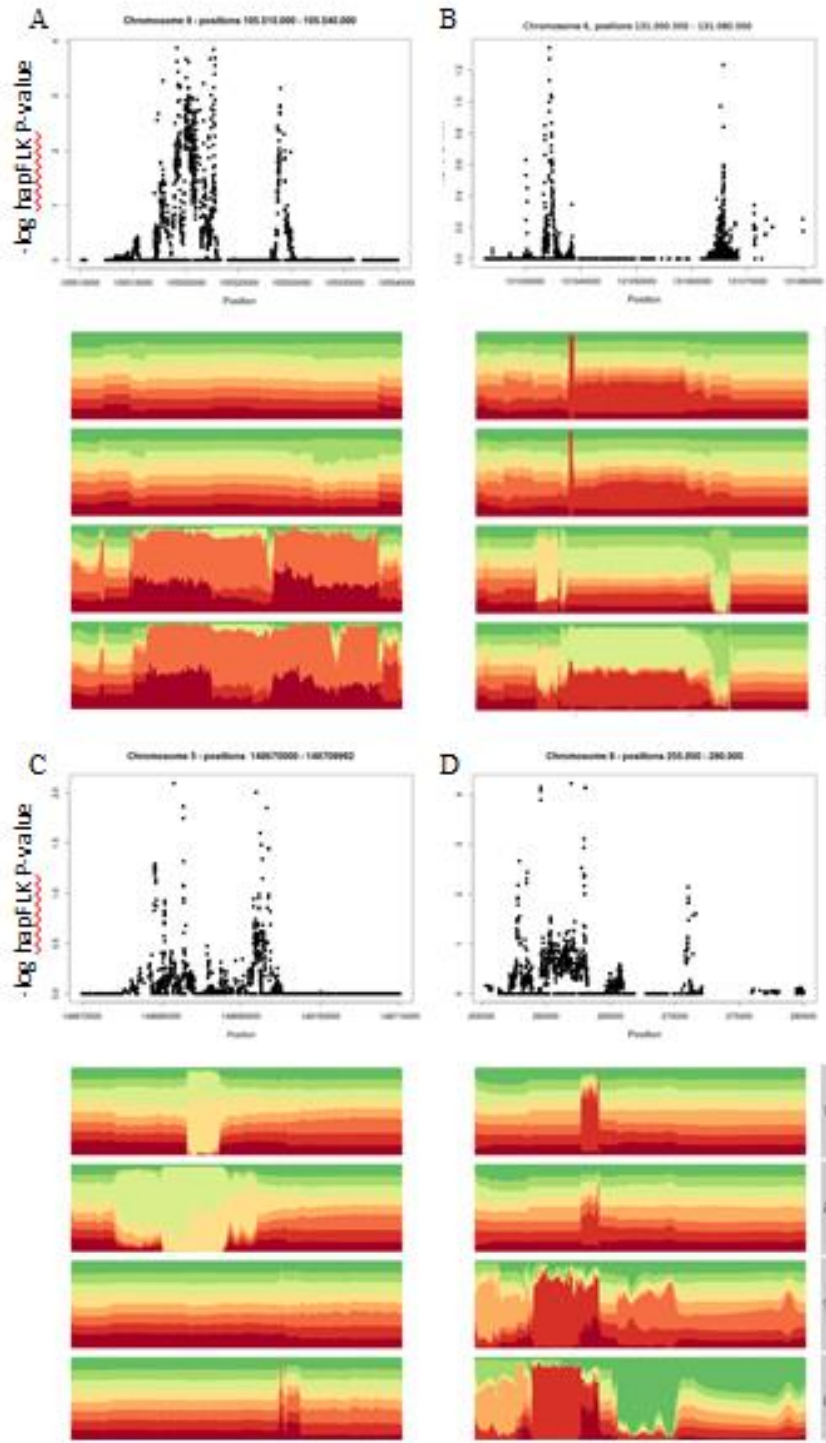
\* Gene in common with (Takuno *et al.* 2015)

+ Gene in common with (Pyhäjärvi *et al.* 2013)



#### **II.4.5 Validating candidate regions with haplotype analyses.**

We performed haplotype analyses with our modified version of hapFLK that accounted for genotyping uncertainties. We restricted our analyses to the candidate regions defined in Table 4 that encompassed clusters of candidate SNPs detected by site-based neutrality tests. Among 43 regions we found evidence for haplotype selection at 25 of them (Table 4). Some regions displayed several significant hapFLK peaks (Figure 4A, 4B), indicating several possible targets of selection. We therefore extended our analyses to 40 peaks representing the 25 regions. Visual inspection of the outputs allowed us to make qualitative assessments. First we determined which population(s) was(were) impacted by selection (Figure 4). Second we evaluated the number of selected haplotypes. Figure 4B indicated that at the *alm2* QTL, a single haplotype was locally selected in *mexicana* populations H1 and H2 (close to position 131,040,000) while two haplotypes increased in frequency (position 131,035,000 and surrounded region) in the *parviglumis* populations. Third, when several populations (lowlands or highlands) underwent selection, we determined whether the same or different haplotypes were selected. Figure 4D displayed both cases: around the first hapFLK peak at position 260,000 the same haplotype was selected in L1 and L2, while around the second peak at position 270,000 two distinct haplotypes were selected. Based on qualitative assessments, we made the following quantitative observations: *parviglumis* populations were targeted by selection more often than *mexicana* populations with 29 and 15 instances, respectively. In most cases selective patterns were consistent across lowlands or highlands, 33 of the 40 significant hapFLK peaks exhibited selection in both highlands and/or both lowlands. Among these, we found 24 cases of convergent signals (same haplotype selected across highlands and/or lowlands) as illustrated in Figure 4 (C, D), 18 in *parviglumis* and 6 in *mexicana*. Note that differential selection between lowlands and highlands were rare (8).



**Figure 4: Haplotype modelling in the extreme populations for 4 candidate regions.** Candidate regions are extracted from Table 4 and located on chromosome 9 (A), 6 (B), 5 (C), and 8 (D). For each region, significance of the hapFLK statistic is plotted along SNP positions (in bp, top panel) with the probability distribution of 8 ancestral haplotypes (bottom panel) in each of the 4 extreme populations, highlands (H1, H2) and lowlands (L1, L2).

## **II.5 DISCUSSION.**

In this paper we present the first genome-wide study of local adaptation in teosintes. We used two gradients as replicates to identify the determinants of local adaptation to altitude in the teosintes subspecies *parviglumis* and *mexicana*, the closest wild relatives of maize. Adaptation to altitude here designates a combination of abiotic (Table S2, Supporting information) and biotic factors that co-vary with elevation. Because ecological niches of both subspecies are defined by altitude (Hufford *et al.* 2012a) genetic differentiation between subspecies is an important confounding factor of altitudinal adaptation. We therefore performed whole genome sequencing of 4 extreme populations (1 lowland and 1 highland from each gradient) as well as 2 intermediate populations with similar altitude and biotic conditions (Figure 1) but belonging to the different subspecies, i.e. a highland *parviglumis* population and a lowland *mexicana* from gradient 1. Note that we found a genetic differentiation between subspecies around 12% in both gradients, which is substantially lower than the previously reported value of 33% (Pyhäjärvi *et al.* 2013). The reason for this pattern is unclear but may be due to differences in SNP sampling.

We employed the increasingly popular Pool-sequencing strategy. While using over 50 distinct individuals per population in Pool-seq is recommended to obtain equally accurate or even better estimation of allele frequency (Schlötterer *et al.* 2014) than individual sequencing, census population sizes often impose constraints on the experimental design. We therefore kept track of individual information in the 4 extreme populations to assess potential biases linked to our limited number of mother plants per population (20). We compared our count-based estimates on >12 individuals to independent estimates based on a much greater number of individuals (on average 130 collected on 20 mother plants per population) genotyped by KASPar PCR-based technology for a subset of 219 SNPs (data not shown). We found high Pearson correlation coefficients ( $r$ ) between the two estimates ( $r = 0.82$  to  $0.87$  in the four extreme populations), indicating that our count-based estimates in the extreme populations accurately measured population allele frequencies.

We expected accuracy to decrease in Pool-seq intermediate populations. Indeed, consistently with previous report by Zhu *et al.* (2012) we observed unequal contribution of individuals to the sequencing effort (Figure S2, Supporting information). Such unequal contribution is likely due to differences in DNA quantity and/or quality among individuals, and generates sampling variance of the number of chromosomes per individual contributing to the pool. As a result when we applied a criterion of 12x depth per site in Pool-seq data,

instead of a criterion of 12 individuals per site (Figure 2 and Figure S5, Supporting information), we observed a deficit of rare variants. Such deficit should have limited impact on the differentiation-based neutrality tests that are weakly sensitive to low frequency variants. But, increased sampling noise due to low coverage depth and resulting imprecise estimations of allele frequencies, likely limit our power to detect selection. By explicitly modelling multinomial sampling of sequencing reads, BAYENV2 offers an efficient way to reduce problems due to pool-sequencing (Günther & Coop 2013). Note however that potential unequal chromosome quantity in the DNA pool is not considered in BAYENV2.

We used a common set of 8,479,581 positions genotyped in all 6 populations to detect patterns of positive selection associated with local adaptation. In order to control for false positives, described as a common drawback of population genomic studies (Biswas & Akey 2006), we employed a number of strategies. First, for *Fst*-scan and Pool-HMM we conditioned our results on detection of convergent signals of selection between gradients. We next tested all *Fst*- and BayeScan- outliers with BAYENV2 and retained as candidate SNPs/regions either sites detected by BAYENV2 or regions exhibiting signal consistency (within or among methods) across neighbouring SNPs. Overall we found little overlap across methods (Table 2 and 4). Two lines of arguments may explain this lack of consistency. Cut-offs values may be too stringent and therefore offered little opportunities for detecting the same targets. Methods also rely on distinct signals of selection, i.e. differentiation between populations, diversity within population, and correlation with environmental variables.

With a cut-off value of 5% in the *Fst* analyses, we expected to observe by chance around 22k common SNPs between the two gradients. That we observed 86,148 (Figure S7, Supporting information) indicated that convergence of selection signals actually existed but was not extensive (Tiffin & Ross-Ibarra 2014) probably because it is more likely to occur at the gene or functional unit level rather than at the SNP level (Tenaillon *et al.* 2012). We nevertheless observed several instances (24 out of 40) of convergence, i.e. same selected haplotype in H1 and H2 and/or L1 and L2 using our modified version of HapFLK as illustrated in Figure 4. Overall the results of HapFLK were consistent with our SNP by SNP detection. Indeed, 58% of tested regions (25 among 43) were significant with HapFLK.

HapFLK offers the opportunity to refine our understanding of the modalities of selection. Impact of selection on surrounding diversity varied substantially across selective events from restricted (Figure 4B, peak located around position 131,040,000 bp) to vast regions covering dozens of kb (Figure 4A). In addition, we discerned hard sweeps from soft sweeps where multiple adaptive alleles associated with multiple haplotypes swept through the

population. We detected several instances of soft sweeps including a region encompassing the *mh11* locus (Figure 4A). Soft sweeps are common and result either from selection of a beneficial mutation associated with multiple haplotypes – standing genetic variation – or from recurrent *de novo* beneficial mutations arising on different haplotypes (for a review Messer & Petrov 2013). Distinguishing between those two models would require evaluating the effects of mutations in the absence of selective pressure, supposedly neutral in the former and deleterious in the latter. However two lines of arguments may favour the recurrent mutation model (Jensen 2014): first, selective pressures are likely ancient with good evidence of niche stability over time (Hufford *et al.* 2012a); second, mutational targets are supposedly large for complex adaptive traits.

Half of our candidate SNPs (Table 2) fell within 10kb of annotated gene(s). Interestingly 28 out of the 54 genes reported on Table 3 and Table 4 were detected in a previous study of local adaptation in teosintes in a much broader sample of 21 populations (Pyhäjärvi *et al.* 2013). Along the same line, we confirmed the adaptive role of *Inv1* with 28 SNPs falling within the inversion (Tables 3 and 4). Interestingly, three of the genes that we detected within the inversion (GRMZM2G137561, GRMZM2G424595, GRMZM2G149903) were previously identified as *Fst* outliers between lowlands and highlands in Mesoamerica (Takuno *et al.* 2015) together with 22 other genes (Table 3 and 4).

We evaluated the positions of our candidate regions and candidate SNPs within annotated genes based on a compilation of over 90 studies including association and QTL mapping in teosintes and/or maize, as well as functional validation studies. We focused on traits that were shown to be related to altitude adaptation in both maize and teosinte (Brewbaker 2015; Lauter *et al.* 2004; Takuno *et al.* 2015) as well as on fitness-related traits. We found a total of 32 candidate regions (Table 4) and 224 candidate SNPs (Table 3) falling within previously characterized QTL regions. These QTLs affect a panel of traits related to plant architecture (height), flowering architecture (number of ears, tassel branch number), yield (kernel weight), flowering time (male flowering time, anthesis-silking interval), leaf morphology (macrohairs, pigmentation). We found a candidate region encompassing the *mh11* locus, known to promote macrohairs initiation on the leaf blade (Moose *et al.* 2004). This locus was also described as a target for adaptive introgression from *mexicana* to the maize highlands (Hufford *et al.* 2013). Macrohairs may act as a barrier against pest (Moose *et al.* 2004), and/or may contribute to limit temperature loss in cold environment (Lauter *et al.* 2004). Moose *et al.* (2004) identified 7 recessive *mh11* mutant alleles in maize capable of reducing or eliminating macrohairs. This finding was consistent with selection of independent

mutations causing such reduction in the lowlands *parviglumis*, a pattern of soft sweep observed in Figure 4A. Furthermore, we detected two genes GRMZM2G054048 and GRMZM2G453919 involved in ear and tassel development (Eveland *et al.* 2014) falling into a QTL for ear number (Table 3).

Many of our candidate regions corresponded to QTL regions or genes involved in soil-responses such as drought tolerance, nitrogen use efficiency, water logging tolerance, aluminium tolerance, and low phosphorus tolerance. We found correspondence between 14 of our regions and QTLs for root morphology (Table 3 and 4). Volcanic soils of Mexican highlands are known to be low-P acid soils as compared with the lowland soils (Dahlgren *et al.* 2004). As a result of soil acidity (pH<5), aluminium is solubilized in toxic ionic forms. Both aluminium and low-P therefore impose strong constraints on plant growth in volcanic soils (Zheng 2010). Sequencing of the Palomero toluqueño landrace distributed along the transvolcanic belt actually revealed signatures of selection in heavy metal- and phosphate-transporter genes (Vielle-Calzada *et al.* 2009). As reviewed in Calderón-Vázquez *et al.* (2011), maize long-term phosphorus starvation provokes responses in a diversity of pathways including nitrogen-related pathways, carbohydrate metabolism, nucleotide synthesis and secondary metabolism. Such starvation also affects root morphology by maintaining the elongation of the primary root and by inhibiting lateral root growth (Mollier & Pellerin 1999). Mycorrhiza contained in the soils also play a role of Pi uptake via the Mycorrhiza Pi Uptake Pathway (PMU).

Among annotated genes in our candidate regions, we found the gene GRMZM2G088413 (Table 3) whose expression is up regulated in a loss-of-function mutant of ZmPT6, a transporter of Pi involved in fungi-maize interactions (Willmann *et al.* 2013). In addition, the expression of GRMZM2G088413 is up-regulated in presence of LCO, a signal molecule of symbiotic interactions secreted by rhizobia (Tanaka *et al.* 2015). This gene is therefore involved in the Pi uptake biotic-dependant pathway. A recent study enlightens the role of co-adaptation between *mexicana* teosinte and their rhizosphere biota in cold environment, revealing both the importance of biotic interactions in cold environment and the impact of soil composition on local adaptation (O'Brien *et al.* 2015). Along the same line, we detected candidate SNPs within two genes implied in drought response: GRMZM2G170870 (Table 3) which is up-regulated in drought conditions (Batlang *et al.* 2014), and GRMZM2G133631 (Table 4), an ABA receptor involved in drought response (Wei & Pan 2014). An additional gene GRMZM2G424519, whose function is currently unknown falling within a water deficit QTL and detected by Pyhäjärvi *et al.* (2013), exhibited footprints of

selection (Table 4). HapFLK patterns were consistent with positive selection acting on *mexicana* populations (Figure 4C). Finally, the *alm2* candidate region exhibited several selective sweeps at independent positions (Figure 4B). Only one of them involved selection of a unique haplotype shared by the two extreme *mexicana* populations (Figure 4B, peak 131,040,000). Because we expected strong selection for aluminium tolerance in *mexicana*, this locus may be a good candidate for the *alm2* QTL.

We detected candidate regions associated with resistance QTLs to sugarcane mosaic virus, namely *scmv1* and *scmv2*. Mosaic virus is particularly prominent on maize fields and causes great loss in grain and forage yields (Fuchs & Grüntzig 1995; Wu *et al.* 2007). There are two major resistance QTLs for this disease, *scmv1* on chromosome 6 and *scmv2* on chromosome 3 (Leng *et al.* 2015; Tao *et al.* 2013; Xia *et al.* 1999). None of the annotated genes encompassing our candidate SNPs corresponded to the *CAS1-like* gene or the *Zmtrx-H* gene, that are currently the best candidates for *scmv1* (Tao *et al.* 2013). However, the *ABP1* gene which is one of the 2 best candidates for *scmv2* falls within our candidate region (Table 3, Leng *et al.* 2015).

Climatic changes in the future decades are predicted to induce greater variation in precipitations, and so frequent drought or water-flood event. As several of the studied populations grow in dry environments, adaptive genes may be interesting in the future. Overall this study points to the importance factors that accompanied climatic changes in the past, such as disease resistance and soil composition in teosinte local adaptation to altitude. Measures of soil acidity, composition in nitrogen, phosphorus, carbon as well as microbial composition would help identifying additional candidate genes that may be relevant for future breeding efforts.

## **ACKNOWLEDGEMENTS.**

We are grateful to Renaud Vitalis for his advices on *Fst* analyses and to Rubén Rellán Álvarez for his precious insights on data interpretation. Daniel Grimanelli discussed the experimental set up and provided precious inputs throughout the study. This work was supported by the Agence Nationale de la Recherche (Project ANR 12-ADAP-0002-01) to MIT and YV, and the ECOS Nord project M12A03 to MIT and LEE. M-AF was funded by the Project ANR 12-ADAP-0002-01.

## **REFERENCES.**

- Agrama HA, Moussa ME (1996) Mapping QTLs in breeding for drought tolerance in maize (*Zea mays* L.). *Euphytica* **91**, 89-97.
- Azevedo GC, Cheavegatti-Gianotto A, Negri BF, *et al.* (2015) Multiple interval QTL mapping and searching for PSTOL1 homologs associated with root morphology, biomass accumulation and phosphorus content in maize seedlings under low-P. *BMC plant biology* **15**, 1-17.
- Batlang U, Ambavaram MM, Krishnan A, Pereira A (2014) Drought responsive genes and their functional terms identified by GS FLX Pyro sequencing in maize. *Maydica* **59**, 306-314.
- Baxter I, Brazelton JN, Yu D, *et al.* (2010) A coastal cline in sodium accumulation in *Arabidopsis thaliana* is driven by natural variation of the sodium transporter AtHKT1; 1. *PLoS Genet* **6**, e1001193
- Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology* **13**, 969-980.
- Biswas S, Akey JM (2006) Genomic insights into positive selection. *Trends in Genetics* **22**, 437-446.
- Boitard S, Kofler R, Françoise P, *et al.* (2013) Pool-hmm: a Python program for estimating the allele frequency spectrum and detecting selective sweeps from next generation sequencing of pooled samples. *Molecular Ecology Resources* **13**, 337-340.
- Bonhomme M, Chevalet C, Servin B, *et al.* (2010) Detecting Selection in Population Trees: The Lewontin and Krakauer Test Extended. *Genetics* **186**, 241-262.
- Brewbaker JL (2015) Diversity and Genetics of Tassel Branch Numbers in Maize. *Crop Science* **55**, 65-78.
- Buckler ES, Goodman MM, Holtsford TP, Doebley JF, Sanchez J (2006) Phylogeography of the wild subspecies of *Zea mays*. *Maydica* **51**, 123-134.
- Calderón-Vázquez C, Sawers RJ, Herrera-Estrella L (2011) Phosphate deprivation in maize: genetics and genomics. *Plant Physiology* **156**, 1067-1077.
- Chen Z, Wang B, Dong X, *et al.* (2014) An ultra-high density bin-map for rapid QTL mapping for tassel and ear architecture in a large F2 maize population. *BMC genomics* **15**, 433-443.
- Conte GL, Arnegard ME, Peichel CL, Schluter D (2012) The probability of genetic parallelism and convergence in natural populations. *Proceedings of the Royal Society of London B: Biological Sciences* **279**, 5039-5047.
- Coop G, Witonsky D, Di Rienzo A, Pritchard JK (2010) Using environmental correlations to identify loci underlying local adaptation. *Genetics* **185**, 1411-1423.
- Coque M, Gallais A (2006) Genomic regions involved in response to grain yield selection at high and low nitrogen fertilization in maize. *Theoretical and Applied Genetics* **112**, 1205-1220.
- Crémieux L, Bischoff A, Šmilauerová M, *et al.* (2008) Potential contribution of natural enemies to patterns of local adaptation in plants. *New Phytologist* **180**, 524-533.
- Dahlgren R, Saigusa M, Ugolini F (2004) The nature, properties and management of volcanic soils. *Advances in Agronomy* **82**, 113-182.
- Doebley J, Stec A (1993) Inheritance of the morphological differences between maize and teosinte: comparison of results for two F2 populations. *Genetics* **134**, 559-570.
- Eveland AL, Goldshmidt A, Pautler M, *et al.* (2014) Regulatory modules controlling maize inflorescence architecture. *Genome Research* **24**, 431-443.
- Fang Z, Pyhäjärvi T, Weber AL, *et al.* (2012) Megabase-scale inversion polymorphism in the wild ancestor of maize. *Genetics* **191**, 883-894.
- Fariello MI, Boitard S, Naya H, SanCristobal M, Servin B (2013) Detecting signatures of selection through haplotype differentiation among hierarchically structured populations. *Genetics* **193**, 929-941.
- Fischer MC, Rellstab C, Tedder A, *et al.* (2013) Population genomic footprints of selection and associations with climate in natural populations of *Arabidopsis halleri* from the Alps. *Molecular Ecology* **22**, 5594-5607.
- Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* **180**, 977-993.
- Fournier-Level A, Korte A, Cooper MD, *et al.* (2011) A map of local adaptation in *Arabidopsis thaliana*. *Science* **334**, 86-89.



- Franks SJ, Weber JJ, Aitken SN (2014) Evolutionary and plastic responses to climate change in terrestrial plant populations. *Evolutionary Applications* **7**, 123-139.
- Fuchs E, Grüntzig M (1995) Effect of sugarcane mosaic virus (SCMV) and maize dwarf mosaic virus (MDMV) on the growth and yield of maize in the GDR. *Zeitschrift für Pflanzenkrankheiten und Pflanzenschutz/Journal of Plant Diseases and Protection* **44**, 44-50.
- Gautier M (2015) Genome-wide scan for adaptive divergence and association with population-specific covariates. *Genetics* **201**, 1555-1579.
- Günther T, Coop G (2013) Robust identification of local adaptation from allele frequencies. *Genetics* **195**, 205-220.
- Hancock AM, Brachi B, Faure N, *et al.* (2011) Adaptation to climate across the *Arabidopsis thaliana* genome. *Science* **334**, 83-86.
- Howden SM, Soussana J-F, Tubiello FN, *et al.* (2007) Adapting agriculture to climate change. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 19691-19696.
- Hufford M, Martinez-Meyer E, Gaut BS, Eguiarte LE, Tenaillon MI (2012a) Inferences from the historical distribution of wild and domesticated maize provide ecological and evolutionary insight. *PLoS ONE* **7**, e47659.
- Hufford MB, Bilinski P, Pyhäjärvi T, Ross-Ibarra J (2012b) Teosinte as a model system for population and ecological genomics. *Trends in Genetics* **28**, 606-615.
- Hufford MB, Lubinsky P, Pyhäjärvi T, *et al.* (2013) The genomic signature of crop-wild introgression in maize. *PLoS ONE* **9**, e1003477.
- Husson F, Lê S, Mazet J (2007) FactoMineR: Factor Analysis and Data Mining with R. R package version 1.05.
- Jensen JD (2014) On the unfounded enthusiasm for soft selective sweeps. *Nature communications* **5**, 1-10.
- Johnson M (2007) Genotype-by-environment interactions leads to variable selection on life-history strategy in common evening Primrose (*Oenothera biennis*). *Journal of Evolutionary Biology* **20**, 190-200.
- Kawecki TJ, Ebert D (2004) Conceptual issues in local adaptation. *Ecology Letters* **7**, 1225-1241.
- Keller SR, Levens N, Olson MS, Tiffin P (2012) Local adaptation in the flowering-time gene network of balsam poplar, *Populus balsamifera* L. *Molecular Biology and Evolution* **29**, 3143-3152.
- Lauter N, Gustus C, Westerbergh A, Doebley J (2004) The inheritance and evolution of leaf pigmentation and pubescence in teosinte. *Genetics* **167**, 1949-1959.
- Le Corre V, Roux F, Reboud X (2002) DNA polymorphism at the FRIGIDA gene in *Arabidopsis thaliana*: extensive nonsynonymous variation is consistent with local selection for flowering time. *Molecular Biology and Evolution* **19**, 1261-1271.
- Leimu R, Fischer M (2008) A meta-analysis of local adaptation in plants. *PLoS ONE* **3**, e4010.
- Leng P, Ji Q, Tao Y, *et al.* (2015) Characterization of sugarcane mosaic virus Scmv1 and Scmv2 resistance regions by regional association analysis in maize. *PLoS ONE* **10**, e0140617.
- Lewontin R, Krakauer J (1975) Testing the heterogeneity of F values. *Genetics* **80**, 397.
- Li H, Handsaker B, Wysoker A, *et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079.
- Lobell DB, Schlenker W, Costa-Roberts J (2011) Climate trends and global crop production since 1980. *Science* **333**, 616-620.
- Lotterhos KE, Whitlock MC (2014) Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. *Molecular Ecology* **23**, 2178-2192.
- Lunter G, Goodson M (2011) Stampy: A statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Research* **21**, 936-939.
- Malosetti M, Ribaut JM, Vargas M, Crossa J, Van Eeuwijk FA (2008) A multi-trait multi-environment QTL mixed model with an application to drought and nitrogen stress trials in maize (*Zea mays* L.). *Euphytica* **161**, 241-257.
- Messer PW, Petrov DA (2013) Population genomics of rapid adaptation by soft selective sweeps. *Trends in Ecology & Evolution* **28**, 659-669.

- 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.
- Moeller DA, Tiffin P (2008) Geographic variation in adaptation at the molecular level: a case study of plant immunity genes. *Evolution* **62**, 3069-3081.
- Mollier A, Pellerin S (1999) Maize root system growth and development as influenced by phosphorus deficiency. *Journal of experimental botany* **50**, 487-497.
- Moose SP, Lauter N, Carlson SR (2004) The maize macrohairless1 locus specifically promotes leaf blade macrohair initiation and responds to factors regulating leaf identity. *Genetics* **166**, 1451-1461.
- Muñoz Diez C, Gaut BS, Meca E, *et al.* (2013) Genome size variation in wild and cultivated maize along altitudinal gradients. *New Phytologist* **199**, 264-276.
- Narum SR, Hess JE (2011) Comparison of FST outlier tests for SNP loci under selection. *Molecular. Ecology. Resources* **11**, 184-194.
- Nei M (1973) Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences of the United States of America* **70**, 3321-3323.
- O'Brien AM, Sawers RJ, Ross-Ibarra J, Strauss SY (2015) Extending the Stress-Gradient hypothesis: increased local adaptation between teosinte and soil biota at the stressful end of a climate gradient. *bioRxiv*, 031195.
- Porter JR, Semenov MA (2005) Crop responses to climatic variation. *Philosophical Transactions of the Royal Society B: Biological Sciences* **360**, 2021-2035.
- Pyhäjärvi T, Hufford MB, Mezouk S, Ross-Ibarra J (2013) Complex patterns of local adaptation in teosinte. *Genome biology and evolution* **5**, 1594-1609.
- Qiu F, Zheng Y, Zhang Z, Xu S (2007) Mapping of QTL associated with waterlogging tolerance during the seedling stage in maize. *Annals of Botany* **99**, 1067-1081.
- Quint M, Mihaljevic R, Dussle C, *et al.* (2002) Development of RGA-CAPS markers and genetic mapping of candidate genes for sugarcane mosaic virus resistance in maize. *Theoretical and Applied Genetics* **105**, 355-363.
- Reiter RS, Coors J, Sussman M, Gabelman W (1991) Genetic analysis of tolerance to low-phosphorus stress in maize using restriction fragment length polymorphisms. *Theoretical and Applied Genetics* **82**, 561-568.
- Rhone B, Remoue C, Galic N, Goldringer I, Bonnin I (2008) Insight into the genetic bases of climatic adaptation in experimentally evolving wheat populations. *Molecular Ecology* **17**, 930-943.
- Ross-Ibarra J, Tenaillon MI, Gaut BS (2009) Historical divergence and gene flow in the genus *Zea*. *Genetics* **181**, 1399-1413.
- Savolainen O, Kujala ST, Sokol C, *et al.* (2011) Adaptive potential of northernmost tree populations to climate change, with emphasis on Scots pine (*Pinus sylvestris* L.). *Journal of Heredity*, esr056.
- Savolainen O, Lascoux M, Merilä J (2013) Ecological genomics of local adaptation. *Nature Reviews Genetics* **14**, 807-820.
- Scheet P, Stephens M (2006) A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *The American Journal of Human Genetics* **78**, 629-644.
- Schlötterer C, Tobler R, Kofler R, Nolte V (2014) Sequencing pools of individuals: mining genome-wide polymorphism data without big funding. *Nature Reviews Genetics* **15**, 749-763.
- Schnable PS, Ware D, Fulton RS, *et al.* (2009) The B73 Maize Genome: Complexity, Diversity, and Dynamics. *Science* **326**, 1112-1115.
- Sibov S, Gaspar M, Silva M, *et al.* (1999) Two genes control aluminum tolerance in maize: genetic and molecular mapping analyses. *Genome* **42**, 475-482.
- Sicard D, Pennings PS, Grandclément C, *et al.* (2007) Specialization and local adaptation of a fungal parasite on two host plant species as revealed by two fitness traits. *Evolution* **61**, 27-41.
- Takuno S, Ralph P, Swarts K, *et al.* (2015) Independent molecular basis of convergent highland adaptation in maize. *bioRxiv*, 013607.
- Tanaka K, Cho S-H, Lee H, *et al.* (2015) Effect of lipo-chitooligosaccharide on early growth of C4 grass seedlings. *Journal of experimental botany*.

- Tao Y, Jiang L, Liu Q, *et al.* (2013) Combined linkage and association mapping reveals candidates for *Scmv1*, a major locus involved in resistance to sugarcane mosaic virus (SCMV) in maize. *BMC plant biology* **13**, 162.
- Tenaillon MI, Sawkins MC, Long AD, *et al.* (2001) Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays ssp mays* L.). *Proceedings of the National Academy of Sciences of the United States of America* **98**, 9161-9166.
- Tenaillon O, Rodríguez-Verdugo A, Gaut RL, *et al.* (2012) The molecular diversity of adaptive convergence. *Science* **335**, 457-461.
- Tiffin P, Ross-Ibarra J (2014) Advances and limits of using population genetics to understand local adaptation. *Trends in Ecology & Evolution* **29**, 673-680.
- Turner TL, Bourne EC, Von Wettberg EJ, Hu TT, Nuzhdin SV (2010) Population resequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. *Nature genetics* **42**, 260-263.
- Vielle-Calzada JP, de la Vega OM, Hernandez-Guzman G, *et al.* (2009) The Palomero genome suggests metal effects on domestication. *Science* **326**, 1078-1078.
- Vitalis R, Dawson K, Boursot P (2001) Interpretation of variation across marker loci as evidence of selection. *Genetics* **158**, 1811-1823.
- Voight BF, Kudaravalli S, Wen X, Pritchard JK (2006) A map of recent positive selection in the human genome. *PLoS Biol* **4**, e72.
- Volis S, Mendlinger S, Ward D (2002) Differentiation in populations of *Hordeum spontaneum* along a gradient of environmental productivity and predictability: life history and local adaptation. *Biological Journal of the Linnean Society* **77**, 479-490.
- Wei K, Pan S (2014) Maize protein phosphatase gene family: identification and molecular characterization. *BMC genomics* **15**, 773-793.
- Weir BS, Hill WG (2002) Estimating F-Statistics. *Annual Review of Genetics* **36**, 721-750.
- Welcker C, Sadok W, Dignat G, *et al.* (2011) A common genetic determinism for sensitivities to soil water deficit and evaporative demand: meta-analysis of quantitative trait loci and introgression lines of maize. *Plant Physiology* **157**, 718-729.
- Willmann M, Gerlach N, Buer B, *et al.* (2013) Mycorrhizal phosphate uptake pathway in maize: vital for growth and cob development on nutrient poor agricultural and greenhouse soils. *Frontiers in plant science* **4**, 433-448.
- Wu J-y, Ding J-q, Du Y-x, Xu Y-b, Zhang X-c (2007) Genetic analysis and molecular mapping of two dominant complementary genes determining resistance to sugarcane mosaic virus in maize. *Euphytica* **156**, 355-364.
- Xia X, Melchinger AE, Kuntze L, Lübberstedt T (1999) Quantitative trait loci mapping of resistance to sugarcane mosaic virus in maize. *Phytopathology* **89**, 660-667.
- Yeaman S, Otto SP (2011) Establishment and maintenance of adaptive genetic divergence under migration, selection, and drift. *Evolution* **65**, 2123-2129.
- Yoder JB, Stanton-Geddes J, Zhou P, *et al.* (2014) Genomic signature of adaptation to climate in *Medicago truncatula*. *Genetics* **196**, 1263-1275.
- Yu J, Li X, Zhu C, *et al.* (2012) Genic and non-genic contributions to natural variation of quantitative traits in maize. *Genome Research*, gr. 140277.140112.
- Zhang H, Uddin MS, Zou C, *et al.* (2014) Meta-analysis and candidate gene mining of low-phosphorus tolerance in maize. *Journal of integrative plant biology* **56**, 262-270.
- Zheng SJ (2010) Crop production on acidic soils: overcoming aluminium toxicity and phosphorus deficiency. *Annals of Botany* **106**, 183-184.
- Zhu Y, Bergland AO, González J, Petrov DA (2012) Empirical validation of pooled whole genome population re-sequencing in *Drosophila melanogaster*. *PLoS ONE* **7**, e41901.

## **AUTHOR CONTRIBUTIONS.**

MIT, DM, LEE and YV designed the study; MIT generated the data; M-AF performed the analyses under the supervision of DM and MIT; J-TB helped with scripts writing and analyses; SB and JL developed the mathematical framework for modifying the hapFLK method and ran the corresponding analyses. DM, LEE, YV, SB and MIT assisted M-AF with data interpretation. M-AF and MIT wrote the manuscript. All authors provided comments on the manuscript.

## **SUPPORTING INFORMATION.**

SupInfoTableS1-S3-FigS1-S7.pdf

Table S1. Location and sampling information of teosinte populations.

Supplementary Table S2. Climatic variables for all sampled teosinte populations.

Table S3. SNPs detected using BAYENV2.

Fig. S1. Map of Mexico.

Fig. S2. Relative contribution of individuals to sequencing effort in the 4 extreme populations

Fig. S3. Depth per site per population.

Fig. S4. Distribution of the number of individuals that cover sites comprise between 12 and 50x per population.

Fig. S5. SFS for  $\text{SNP}^{12}$  and  $\text{SNP}^{12x}$  with count-based estimation of allele frequency.

Fig. S6. Homemade versus BCFtools SFS estimation on  $\text{SNP}^{12}$  for the 4 extreme populations.

Fig. S7. Top 5%  $F_{st}$  for each gradient.

Supplementary Table S1: Location and sampling information of teosinte populations.

Population	Nomenclature	Subspecies <sup>1</sup>	Latitude	Longitude	Altitude	Place	State	Coldate
<i>Gradient1</i>								
L1	CM04	<i>parviglumis</i>	16.9811	-99.2855	581	Tecoanapa	Guerrero	2010
2	CM03	<i>parviglumis</i>	17.1719	-99.5415	343	Tierra Colorada	Guerrero	2010
3	CM05	<i>parviglumis</i>	17.3918	-99.4776	1201	Chilpancingo	Guerrero	2010
4	CM06	<i>parviglumis</i>	17.4219	-99.4507	1251	Mazatlán	Guerrero	2010
5	CM07	<i>parviglumis</i>	17.4596	-99.3683	1107	Mochitlán	Guerrero	2010
6	CM08	<i>parviglumis</i>	18.2377	-99.218	1101	Paso Morelos	Guerrero	2010
7	CM01	<i>parviglumis</i>	18.3498	-99.8411	1649	Teloloapan	Guerrero	2010
IP1	CM02	<i>parviglumis</i>	18.411	-99.9084	1439	Alcholoa	Guerrero	2010
IM1	CM09	<i>mexicana</i>	18.9741	-99.0703	1669	Huilotepec	Morelos	2010
10	SMH581	<i>parviglumis</i>	19.3274	-100.4214	1383	Enandio	Michoacán	2011
H1	CM10	<i>mexicana</i>	19.4075	-99.6271	2581	VillaSeca	Mexico	2010
<i>Gradient2</i>								
12	SMH577	<i>parviglumis</i>	19.7325	-104.8719	572	El Llanito	Jalisco	2009
L2	CM13	<i>parviglumis</i>	19.7154	-104.8035	504	Telpitita	Jalisco	2010
14	CM12	<i>parviglumis</i>	20.6273	-104.4079	1426	Guachinango	Jalisco	2010
15	SMH578	<i>parviglumis</i>	19.5354	-104.0583	1369	El Rodeo	Jalisco	2009
16	SMH564	<i>parviglumis</i>	19.91	-104.1726	1407	Ejutla	Jalisco	2009
17	SMH565	<i>parviglumis</i>	19.8961	-104.1768	1317	Ejutla	Jalisco	2009
18	SMH566	<i>parviglumis</i>	19.9027	-104.1572	1140	El Estanco	Jalisco	2009
19	SMH567	<i>parviglumis</i>	19.913	-104.1169	976	El Estanco	Jalisco	2009
20	MIT15	<i>parviglumis</i>	19.9129	-104.1123	976	Ejutla	Jalisco	2010
21	SMH569	<i>parviglumis</i>	19.9218	-104.0977	963	Los Naranjos	Jalisco	2009
22	SMH570	<i>parviglumis</i>	19.9339	-104.0077	944	La Labor	Jalisco	2009
23	MIT14	<i>parviglumis</i>	19.9458	-103.9958	958	San Lorenzo	Jalisco	2010
24	CM11	<i>mexicana</i>	20.1394	-102.0684	1846	Churintzio	Michoacán	2010
25	SMH579	<i>mexicana</i>	20.1338	-101.4342	2002	Armadillo	Michoacán	2010
26	SMH576	<i>mexicana</i>	20.1607	-101.3731	1856	Tejocote de Calera	Guanajuato	2009

27	SMH580	<i>mexicana</i>	19.9646	-101.2354	1844	Capacho	Michoacán	2010
28	SMH575	<i>mexicana</i>	20.0532	-101.0881	1849	San Rafael	Michoacán	2009
29	SMH573	<i>mexicana</i>	19.9828	-100.9599	1878	Puerto de Cabras	Guanajuato	2009
H2	SMH571	<i>mexicana</i>	19.9917	-100.8849	1861	San José de las Pilas	Guanajuato	2009
31	SMH572	<i>mexicana</i>	19.9582	-100.8497	1854	Andocutin	Guanajuato	2009

<sup>1</sup> Subspecies status was based on distribution maps of the two subspecies (Hufford *et al.* 2012)

**Supplementary Table S2: Climatic variables for teosinte populations**

Pop	T1 <sup>1</sup>	T2 <sup>2</sup>	T3 <sup>3</sup>	T4 <sup>4</sup>	T5 <sup>5</sup>	T6 <sup>6</sup>	T7 <sup>7</sup>	T8 <sup>8</sup>	T9 <sup>9</sup>	T10 <sup>10</sup>	T11 <sup>11</sup>	P12 <sup>12</sup>	P13 <sup>13</sup>	P14 <sup>14</sup>	P15 <sup>15</sup>	P16 <sup>16</sup>	P17 <sup>17</sup>	P18 <sup>18</sup>	P19 <sup>9</sup>
<i>Gradient1</i>																			
L1	25.8	12.2	0.76	0.25	33.5	17.4	16.2	26.1	25.2	26.7	24.8	1595	107	0	118	1046	0	488	15
2	27.5	13.4	0.71	0.3	36.9	18.2	18.8	27.6	27.4	28.8	26.5	1283	80	0	117	831	0	306	16
3	22.3	12.4	0.71	0.39	31	13.5	17.5	22.7	21.9	23.7	20.8	1107	66	0	113	732	0	311	23
4	22	12.5	0.71	0.4	30.9	13.1	17.7	22.4	21.6	23.5	20.5	1072	64	0	112	703	0	255	24
5	23.5	13.4	0.72	0.4	32.7	14	18.7	23.9	23.2	25	21.9	1030	59	0	110	664	0	245	24
6	25.3	16.9	0.64	0.72	38.3	11.8	26.5	25.9	22.5	28.2	22.5	895	50	0	110	564	0	181	12
7	21	11.9	0.63	0.54	31.1	12.1	19	20.9	21.6	23.4	19.3	1096	62	0	110	699	0	143	23
IP1	21.9	12.2	0.65	0.53	32.2	13.2	19	21.5	22.6	24.2	20.3	1122	62	0	112	717	0	135	18
IM1	19.7	14.6	0.71	0.47	30.2	9.5	20.7	20.1	18.2	21.5	17.9	931	51	0	106	572	0	183	20
10	21.7	13.7	0.67	0.5	32.6	12	20.5	21.9	21.7	23.9	20	879	47	0	101	529	0	200	38
H1	13.1	15.9	0.68	0.68	24	0.5	23.5	14.5	10.4	15.1	10.3	898	49	0	93	530	38	296	38
<i>Gradient2</i>																			
12	25.1	15.2	0.66	0.63	35.9	13	22.9	26.6	24.9	27.1	22.5	1498	96	0	117	1051	0	665	53
L2	25.1	15.1	0.66	0.63	35.9	13	22.9	26.6	25	27.1	22.5	1517	97	0	117	1063	0	682	52
14	20	17.2	0.64	0.86	32.9	6.2	26.7	22.2	20.1	22.6	16.4	924	59	0	109	610	0	371	51
15	22.4	15.6	0.66	0.61	33.9	10.2	23.7	23.6	22	24.3	19.8	1042	61	0	106	672	0	345	49
16	21.6	17	0.64	0.77	34.6	8	26.6	23.4	21.3	24.1	18.5	893	52	0	100	547	0	330	55
17	21.8	16.9	0.64	0.77	34.8	8.3	26.6	23.6	21.5	24.3	18.7	902	52	0	101	554	0	331	54
18	22.1	16.9	0.63	0.78	35.1	8.4	26.7	23.8	21.8	24.6	18.8	853	49	0	100	520	0	317	52
19	23	16.9	0.62	0.81	36.1	9.1	27	24.9	22.7	25.6	19.6	800	46	0	101	490	0	300	49

20	23	16.9	0.62	0.81	36.1	9.1	27	24.9	22.7	25.6	19.6	800	46	0	101	490	0	300	49
21	23.2	16.9	0.62	0.82	36.3	9.2	27.1	25.1	22.9	25.8	19.7	792	45	0	101	486	0	299	49
22	22.3	17.3	0.63	0.82	35.7	8.2	27.5	24.2	22.1	24.9	18.9	766	42	0	98	461	0	295	49
23	22.7	17.3	0.63	0.83	36.1	8.5	27.6	24.6	22.5	25.3	19.2	768	43	0	99	467	0	296	49
24	18.4	16.7	0.63	0.82	31.4	5	26.4	19.8	17.5	21.3	15.1	913	61	0	114	618	0	257	26
25	17.9	16	0.63	0.82	30.4	5	25.5	19.3	17.1	20.7	14.5	790	50	0	107	514	0	234	33
26	18.4	16.1	0.63	0.84	31	5.5	25.5	19.9	17.7	21.3	15	767	49	0	106	496	0	228	32
27	18.2	16.5	0.64	0.8	31.1	5.3	25.8	19.6	17.7	21	15	722	44	0	104	461	0	215	29
28	18.4	16.2	0.64	0.82	31	5.5	25.5	19.9	17.6	21.1	15	708	43	0	105	456	0	208	27
29	18.2	16.2	0.64	0.8	30.7	5.2	25.5	19.7	17.4	20.9	14.9	708	43	0	106	458	0	209	26
H2	17.5	16.1	0.64	0.79	29.7	4.6	25.1	19	16.6	20.1	14.3	750	46	0	107	488	0	222	29
31	18	16.2	0.64	0.8	30.4	5	25.4	19.5	17.2	20.7	14.7	744	46	0	107	484	0	220	28

<sup>1</sup> Annual Mean Temperature

<sup>2</sup> Mean Temperature Diurnal Range (Mean of monthly (max temp - min temp))

<sup>3</sup> Isothermality (BIO2/BIO7) (\* 100)

<sup>4</sup> Temperature Seasonality (standard deviation \*100)

<sup>5</sup> Max Temperature of Warmest Month

<sup>6</sup> Min Temperature of Coldest Month

<sup>7</sup> Temperature Annual Range (BIO5-BIO6)

<sup>8</sup> Mean Temperature of Wettest Quarter

<sup>9</sup> Mean Temperature of Driest Quarter

<sup>10</sup> Mean Temperature of Warmest Quarter

<sup>11</sup> Mean Temperature of Coldest Quarter

<sup>12</sup> Annual Precipitation

<sup>13</sup> Precipitation of Wettest Month

<sup>14</sup> Precipitation of Driest Month

<sup>15</sup> Precipitation Seasonality (Coefficient of Variation)

<sup>16</sup> Precipitation of Wettest Quarter

<sup>17</sup> Precipitation of Driest Quarter

<sup>18</sup> Precipitation of Warmest Quarter

<sup>19</sup> Precipitation of Coldest Quarter Annual Mean Temperature

**Supplementary Table S3: SNPs detected using BAYENV2.**

SNP ID	Chromosome	Position	Method <sup>1</sup>	Alpha
1	1	12,134	P	0.500000
2	1	45,142	P	0.500000
3	1	46,587	P	0.499985
4	1	228,841	P	0.499995
5	1	262,410	P	0.499995
6	1	1,096,356	F	0.499985
7	1	2,936,016	P	0.499985
8	1	21,247,570	P	0.500000
9	1	21,247,981	P	0.499985
10	1	25,432,485	P	0.499995
11	1	45,207,355		0.499995
12	1	50,343,819		0.500000
13	1	52,464,433		0.499990
14	1	54,814,954		0.500000
15	1	77,809,785		0.499985
16	1	77,924,182		0.499995
17	1	77,924,190		0.499995
18	1	81,430,526		0.500000
19	1	86,587,166		0.500000
20	1	89,255,353		0.500000
21	1	110,081,625		0.499990
22	1	110,085,275		0.499985
23	1	116,022,109	P	0.500000
24	1	118,894,397	P	0.499990
25	1	136,331,699		0.499995
26	1	151,255,775		0.499990
27	1	151,289,387		0.499990
28	1	151,358,504		0.500000
29	1	155,911,962	P	0.499990
30	1	159,699,356		0.499990
31	1	163,266,738	P	0.499985
32	1	163,361,370		0.500000
33	1	168,262,581	F	0.499985
34	1	171,425,283		0.500000
35	1	177,549,865	F	0.499995
36	1	177,550,780	F	0.499985
37	1	198,394,275		0.500000
38	1	198,398,505		0.499995
39	1	215,082,407		0.499985
40	1	225,253,352	F	0.499985
41	1	225,284,905	F	0.500000
42	1	227,754,317		0.499990



43	1	227,807,998		0.499990
44	1	230,122,878		0.500000
45	1	230,173,056		0.500000
46	1	254,420,125	P	0.499990
47	1	291,122,374		0.500000
48	2	112,106	P	0.499990
49	2	51,939,797	B	0.499985
50	2	52,800,529	B	0.500000
51	2	53,612,290	P	0.499995
52	2	54,371,246	P	0.500000
53	2	54,404,785	P	0.500000
54	2	54,592,860	P	0.499995
55	2	54,762,048	P	0.500000
56	2	55,683,874	P	0.500000
57	2	55,692,076	P	0.499985
58	2	57,846,609	F	0.499990
59	2	58,098,145	P	0.499985
60	2	62,759,571	P	0.500000
61	2	62,761,339	P	0.500000
62	2	63,856,700	P	0.499985
63	2	66,774,064	P	0.499985
64	2	72,452,064	P	0.500000
65	2	72,643,501	P	0.500000
66	2	73,697,835	P	0.500000
67	2	831,28,463		0.500000
68	2	91,642,561	F	0.499990
69	2	91,648,561	F	0.499985
70	2	95,736,189	B	0.499995
71	2	99,847,185	F	0.499985
72	2	99,927,829	F	0.500000
73	2	100,496,382	F	0.499990
74	2	103,823,073	B	0.500000
75	2	114,759,083		0.500000
76	2	116,669,476		0.500000
77	2	122,228,632	B	0.499995
78	2	123,406,276		0.499990
79	2	128,070,497		0.499985
80	2	130,278,695		0.500000
81	2	138,476,899		0.500000
82	2	139,710,576		0.500000
83	2	141,787,782		0.499985
84	2	142,103,728		0.499985
85	2	1423,31,852		0.499985
86	2	142,959,191		0.499985
87	2	154,242,256	P	0.500000

88	2	154,244,179	P	0.499995
89	2	154,256,917	P	0.500000
90	2	154,256,989	P	0.500000
91	2	154,291,913	P	0.500000
92	2	172,114,057	P	0.500000
93	2	172,114,208	P	0.500000
94	3	205,844	P	0.500000
95	3	888,296	P	0.500000
96	3	1,021,506	P	0.499990
97	3	1,035,963	P	0.499995
98	3	1,283,959	P	0.500000
99	3	13,07,021	P	0.500000
100	3	62,335,914		0.499986
101	3	72,509,411	P	0.499995
102	3	79,448,823	P	0.500000
103	3	80,658,513	P	0.499985
104	3	80,935,880	P	0.499985
105	3	80,971,807	P	0.499995
106	3	80,971,815	P	0.500000
107	3	82,825,027	P	0.499985
108	3	83,062,818	P	0.499995
109	3	83,689,848	P	0.500000
110	3	97,876,421		0.499999
111	3	103,335,800	P	0.500000
112	3	105,894,422	P	0.500000
113	3	107,133,087	P	0.499995
114	3	107,136,100	P	0.499995
115	3	107,203,280	P	0.500000
116	3	121,865,009	P	0.500000
117	3	122,059,667	P	0.500000
118	3	124,636,853	P	0.500000
119	3	127,006,571	P	0.499990
120	3	127,007,175	P	0.499990
121	4	73,598	F	0.499990
122	4	747,240	P	0.499985
123	4	902,129	F	0.499995
124	4	1,179,838	F	0.500000
125	4	1,260,940	F	0.499995
126	4	1,264,277	B	0.499995
127	4	1,463,659	P	0.500000
128	4	1,882,079	P	0.500000
129	4	30,284,843	F	0.499985
130	4	38,758,749		0.499995
131	4	98,137,460		0.499991
132	4	111,672,135	F	0.499990

133	4	111,673,535	F	0.500000
134	4	117,523,056	F	0.499990
135	4	117,541,890	B	0.499985
136	4	117,545,193	B	0.499995
137	4	117,562,672	P	0.499995
138	4	125,423,633		0.500000
139	4	176,040,083	P	0.499995
140	4	176,045,147	P	0.499985
141	4	183,224,855	P	0.499985
142	4	190,624,327		0.500000
143	4	194,669,154	P	0.499990
144	4	210,659,244	P	0.500000
145	4	218,718,213	P	0.499990
146	4	218,721,672	P	0.499985
147	4	218,731,915	P	0.499995
148	4	218,743,314	P	0.499995
149	4	218,831,763	P	0.499995
150	4	218,896,314	P	0.500000
151	4	220,635,806	P	0.499985
152	4	220,662,439	P	0.499990
153	4	220,663,737	P	0.499995
154	4	225,079,347	P	0.499990
155	4	225,087,593	F	0.500000
156	4	225,087,600	F	0.500000
157	4	225,087,602	F	0.500000
158	4	225,087,669	F	0.500000
159	4	225,087,721	F	0.500000
160	5	24,355	P	0.500000
161	5	57,766	P	0.499990
162	5	105,013	P	0.499990
163	5	862,260	P	0.500000
164	5	863,518	P	0.500000
165	5	864,062	P	0.500000
166	5	886,565	P	0.500000
167	5	905,366	P	0.499990
168	5	906,782	P	0.500000
169	5	13,044,690	P	0.499990
170	5	64,475,297	P	0.499985
171	5	93,790,528	P	0.499995
172	5	93,830,870	P	0.499995
173	5	96,913,957	P	0.500000
174	5	98,355,693	P	0.500000
175	5	99,505,680	P	0.499995
176	5	99,510,435	P	0.499985
177	5	99,556,495	P	0.499995

178	5	100,234,762	F	0.499995
179	5	104,407,375	F	0.499995
180	5	104,708,732	F	0.499990
181	5	105,412,335	F	0.500000
182	5	105,418,461	F	0.500000
183	5	105,619,861	F	0.500000
184	5	107,250,983	F	0.499990
185	5	107,507,247		0.500000
186	5	123,622,532		0.499987
187	5	123,845,060	F	0.500000
188	5	155,121,656	F	0.499985
189	5	166,497,758	F	0.500000
190	5	166,497,906	F	0.499995
191	5	166,500,698	B	0.500000
192	5	166,536,582	P	0.499995
193	5	168,665,990		0.499995
194	5	178,800,657	P	0.499990
195	6	890,254	P	0.499985
196	6	901,961	P	0.500000
197	6	912,738	P	0.499990
198	6	4,013,815		0.499993
199	6	30,689,831	F	0.499985
200	6	30,707,217	F	0.499985
201	6	30,976,198	F	0.499995
202	6	43,551,893	P	0.500000
203	6	43,626,257	P	0.499995
204	6	46,350,399	P	0.499995
205	6	46,371,290	P	0.499985
206	6	73,181,447	P	0.499990
207	6	73,279,821	P	0.499985
208	6	73,365,536	P	0.500000
209	6	73,365,980	P	0.499995
210	6	73,517,006	P	0.499990
211	6	95,937,143	P	0.499990
212	6	104,237,725	P	0.500000
213	6	117,819,529	P	0.500000
214	6	122,072,082	P	0.499995
215	6	122,096,668	P	0.499985
216	6	122,102,873	P	0.499990
217	6	147,686,869		0.499985
218	6	154,779,127		0.499990
219	6	169,172,592	P	0.499995
220	7	369,663	P	0.499990
221	7	375,955	P	0.499985
222	7	452,589	P	0.500000

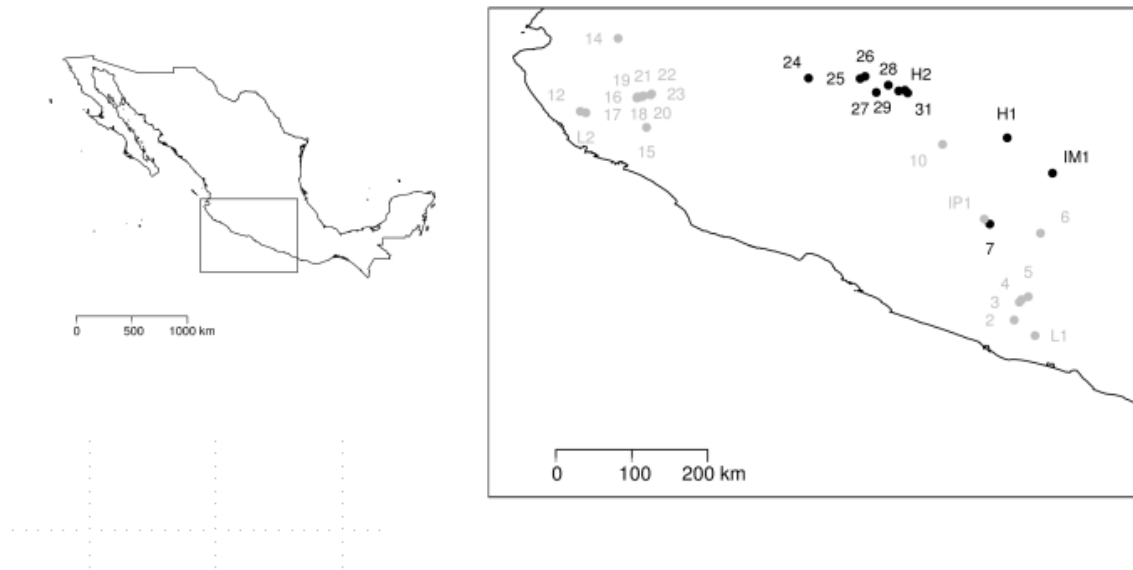
223	7	12,505,861	P	0.499995
224	7	40,849,196	B	0.500000
225	7	44,959,355	P	0.500000
226	7	47,008,899	P	0.499985
227	7	47,112,103	P	0.499990
228	7	48,482,447	P	0.499985
229	7	48,510,146	P	0.499990
230	7	48,574,377	P	0.499985
231	7	49,688,077	P	0.499985
232	7	54,353,310	P	0.499990
233	7	54,754,538	P	0.500000
234	7	55,033,464	P	0.499990
235	7	55,438,565	P	0.499995
236	7	57,064,857	P	0.499985
237	7	59,565,153	P	0.500000
238	7	60,636,011	P	0.499985
239	7	60,983,056	P	0.500000
240	7	61,117,463	P	0.499985
241	7	61,531,063	P	0.500000
242	7	61,757,560	P	0.499990
243	7	63,329,464	P	0.499985
244	7	63,374,097	P	0.499995
245	7	63,397,973	P	0.499990
246	7	63,556,731	P	0.499995
247	7	64,481,206	P	0.500000
248	7	64,545,392	P	0.499995
249	7	64,980,153	P	0.500000
250	7	65,311,080	P	0.499995
251	7	65,315,826	P	0.499990
252	7	66,598,015	P	0.500000
253	7	71,892,833	P	0.499995
254	7	72,521,260	P	0.500000
255	7	73,160,949	P	0.499990
256	7	73,256,740	P	0.499995
257	7	73,391,536	P	0.500000
258	7	74,315,698	P	0.500000
259	7	74,368,578	P	0.500000
260	7	80,800,283	P	0.499995
261	7	90,542,276		0.499985
262	7	99,062,111	P	0.499990
263	7	99,072,654	P	0.499985
264	7	150,171,759	P	0.499995
265	7	150,188,731	P	0.500000
266	7	150,188,747	P	0.500000
267	7	150,188,909	P	0.500000

268	7	150,188,922	P	0.500000
269	7	150,192,260	P	0.500000
270	7	150,192,297	P	0.500000
271	7	150,192,308	P	0.500000
272	7	150,192,311	P	0.500000
273	7	150,192,365	P	0.500000
274	7	150,192,389	P	0.500000
275	7	150,192,479	P	0.500000
276	7	150,210,991	P	0.500000
277	7	150,211,140	P	0.500000
278	7	172,294,887	P	0.500000
279	8	168,614	P	0.500000
280	8	331,700	P	0.499995
281	8	353,663	P	0.500000
282	8	354,942	P	0.499990
283	8	24,503,982		0.499990
284	8	44,343,382		0.499985
285	8	44,875,468	B	0.500000
286	8	54,023,855	P	0.499995
287	8	54,187,852	F	0.499995
288	8	54,189,740	F	0.499990
289	8	54,826,949	F	0.499995
290	8	55,154,034	F	0.499995
291	8	55,168,154	F	0.499995
292	8	57,793,758	F	0.500000
293	8	58,581,082	B	0.499990
294	8	62,206,107	B	0.500000
295	8	70,044,326	P	0.499985
296	8	80,634,168	P	0.499985
297	8	96,486,833	P	0.499990
298	8	97,542,327	P	0.500000
299	8	98,308,835	P	0.500000
300	8	100,599,028		0.500000
301	8	158,357,292		0.500000
302	9	68,504	F	0.500000
303	9	68,526	F	0.500000
304	9	180,357	F	0.499990
305	9	404,330	F	0.499990
306	9	17,947,270	P	0.499985
307	9	66,903,239	P	0.500000
308	9	66,905,346	P	0.500000
309	9	83,966,167	P	0.499990
310	9	93,802,804	P	0.499995
311	9	94,348,525	P	0.499995
312	9	128,660,333	P	0.500000

313	9	128,663,547	P	0.500000
314	9	128,667,900	P	0.499985
315	9	128,767,558	P	0.499995
316	9	128,778,135	F	0.500000
317	9	128,788,408	F	0.499985
318	9	128,850,602	F	0.500000
319	9	128,961,927	F	0.499985
320	9	128,976,314	F	0.500000
321	9	135,144,815	P	0.499997
322	9	135,798,339		0.499985
323	9	140,776,510	B	0.500000
324	9	140,781,231	B	0.500000
325	9	141,325,862	P	0.500000
326	9	141,327,827	P	0.500000
327	9	141,329,773	P	0.500000
328	9	141,388,843	P	0.499995
329	9	141,849,577	P	0.499990
330	9	141,932,855	P	0.500000
331	9	141,933,515	P	0.499990
332	9	142,178,380	P	0.500000
333	10	624,828	P	0.499995
334	10	687,365	P	0.499985
335	10	760,519	P	0.499990
336	10	1,019,407	P	0.499990
337	10	32,521,386	P	0.499985
338	10	44,842,409	P	0.500000
339	10	44,842,411	P	0.500000
340	10	44,842,485	P	0.499990
341	10	45,955,425	P	0.499990
342	10	46,180,893	P	0.499985
343	10	47,895,433	P	0.499985
344	10	48,301,385	P	0.499990
345	10	59,367,334		0.499985
346	10	75,966,585	P	0.499985
347	10	76,038,694	F	0.499990
348	10	103,214,725	B	0.499990
349	10	121,003,639	B	0.500000
350	10	121,003,662	P	0.500000
351	10	121,003,810	P	0.500000
352	10	121,003,812	P	0.500000
353	10	121,003,846	P	0.500000
354	10	121,004,803	P	0.500000
355	10	121,004,921	P	0.500000
356	10	123,878,028	P	0.499985
357	10	123,878,713	P	0.499995

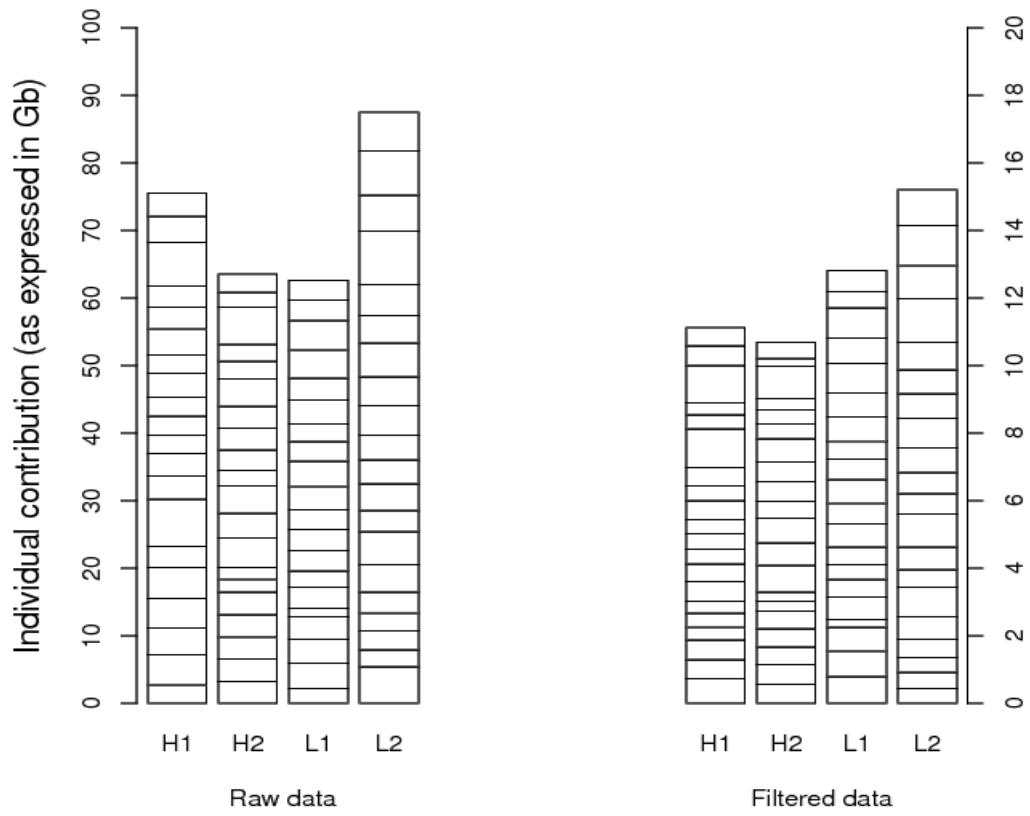
358	10	123,884,759	P	0.500000
359	10	146,786,301	P	0.500000

<sup>1</sup> Methods used to detect SNPs prior to BAYENV2. P :Pool-hmm, F :Fst outlier, B :Bayescan.

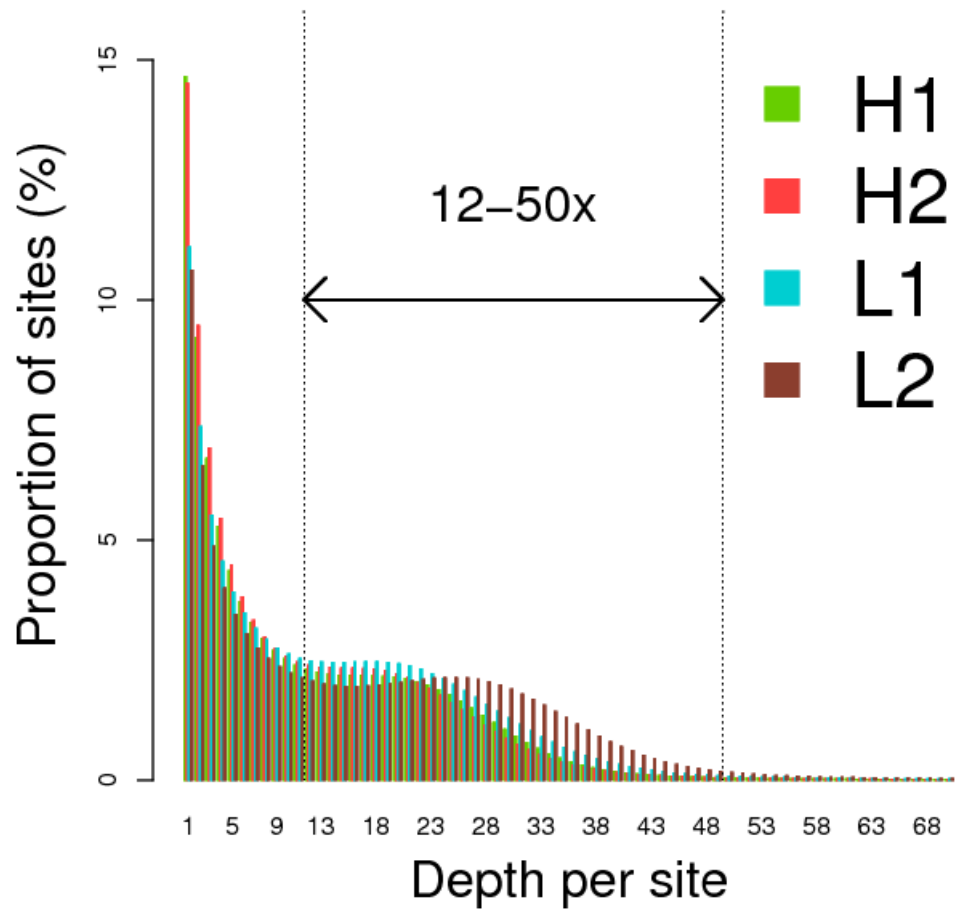


**Supplementary Figure S1: Geographic location of populations.** *mexicana* populations are represented in black, *parviglumis* populations in grey.

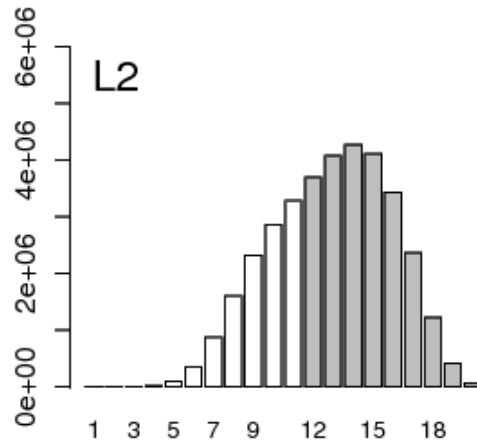
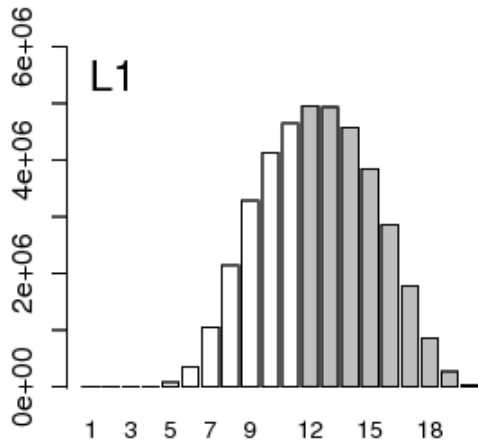
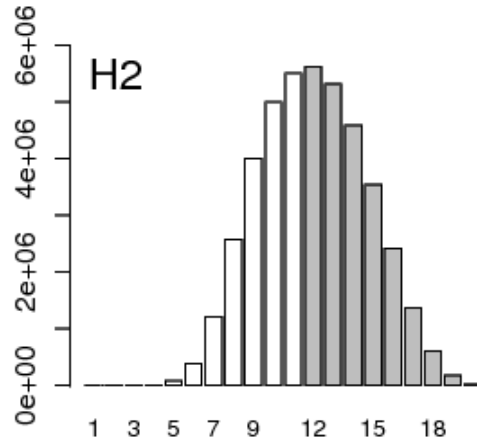
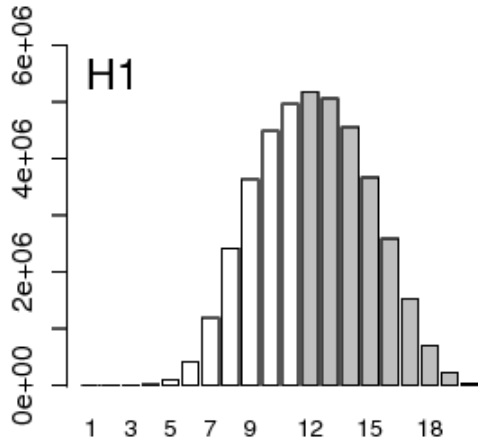




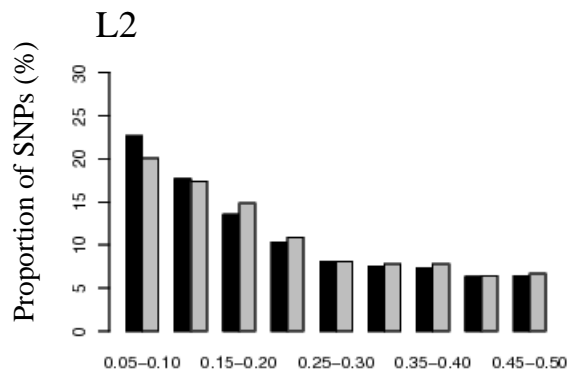
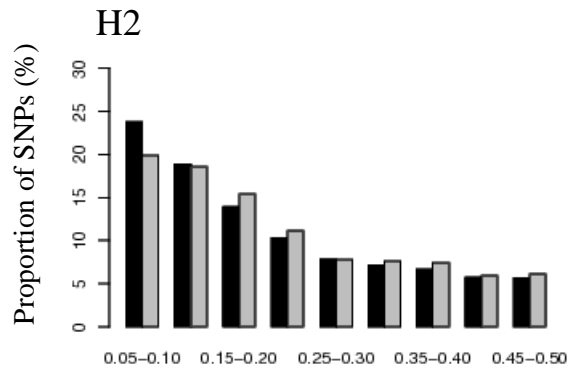
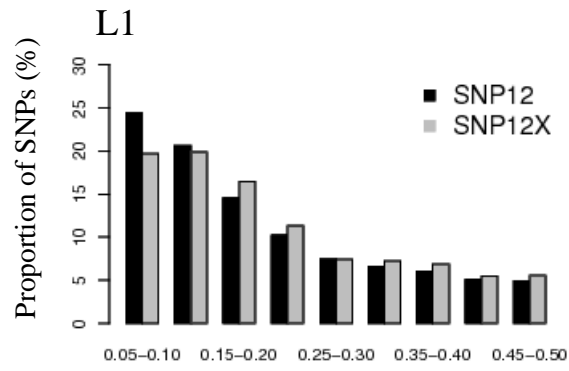
**Supplementary Figure S2: Relative contribution of individuals to sequencing effort in the 4 extreme populations before and after filtering (left : raw data, right : filtered data for base mapping quality and depth 12x-50x)**



Supplementary Figure S3: Depth per-site per-population.

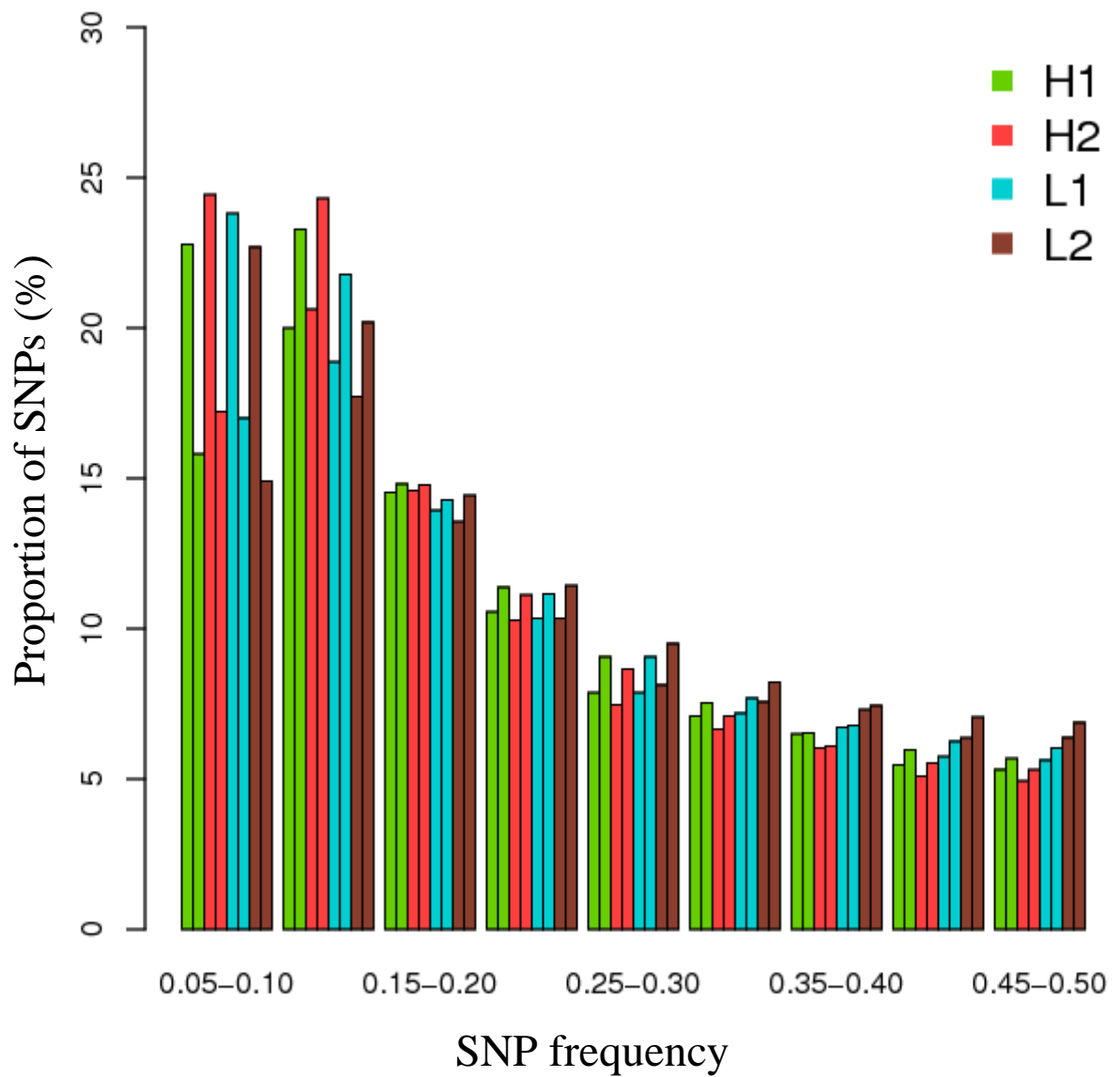


**Supplementary Figure S4: Distribution of the number of individuals that cover sites with a depth comprised between 12x and 50x per population.** The grey bars represent the SNP<sup>12</sup> set.



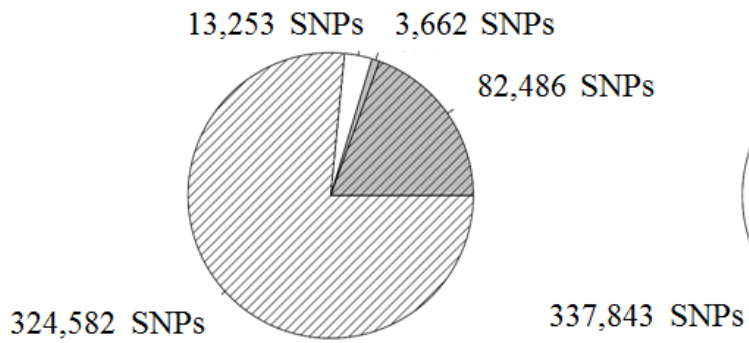
SNP frequency

**Supplementary Figure S5: SFS for SNP<sup>12</sup> and SNP<sup>12x</sup> with count-based estimation of allele frequencies.** SNPs with MAF < 0.05 are discarded.

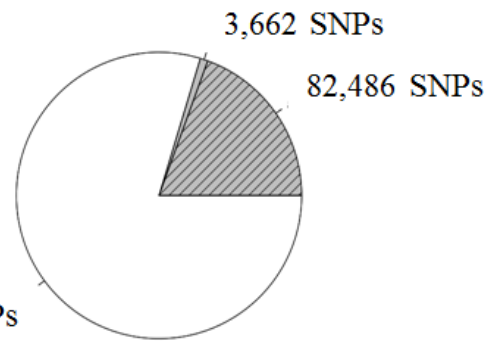


**Supplementary Figure S6: Home-made (left) versus BCFtools (right) SFS estimation on SNP<sup>12</sup> for the 4 extreme populations.**

## Gradient 1



## Gradient 2



**Supplementary Figure S7: Top 5% *F<sub>st</sub>* for each gradient.** In white SNPs detected specifically in gradient 1 or 2, in grey SNPs detected in both gradient 1 and 2, shaded are the SNPs involved in subspecies differentiation and excluded from our outliers set.



### **III. CHAPITRE 2 : VALIDATION DES SNP CANDIDATS PAR L'ETUDE DE CLINES DE FRÉQUENCES**





### III.1 INTRODUCTION

Les populations adaptées localement présentent, dans leurs conditions « natives », des phénotypes plus performants que d'autres populations. L'adaptation locale est donc le résultat d'un processus de sélection naturelle spatialisée. Ce processus peut produire pour certains caractères des patrons d'adaptation le long de clines présentant des changements graduels, tels que des clines d'altitude ou de latitude par exemple. Chez *Lythrum salicaria*, Olsson & Ågren (2002) ont montré que des caractères reliés à la phénologie et morphologie tels que la date de floraison, la hauteur et la production de bourgeons varient le long d'un transect de latitude du nord vers le sud de la Suède. Ainsi les plantes issues de populations du nord fleurissent plus tôt, sont plus petites, produisent moins de bourgeons que les celles des populations du sud. Ces caractères sont des réponses à la durée de la saison de croissance, maintenues par la sélection. Des clines peuvent aussi être observés au niveau de fréquences alléliques et résultent alors d'un équilibre entre la sélection, qui maintient la présence de formes alléliques différentes de part et d'autre du cline, et la migration responsable d'une zone transitoire où les populations présentent des fréquences alléliques intermédiaires (Haldane 1948). La dérive est cependant un effet confondant important qu'il faut prendre en compte. L'étude de ces clines permet de mieux appréhender le rôle relatif de ces forces évolutives et la forme de la sélection (localisation, intensité) dans un système donné.

En plus des clines résultant de processus d'adaptation locale, on peut observer des clines génétiques ou phénotypiques dans des zones hybrides, c'est-à-dire des zones de contact entre deux sous-espèces ou espèces proches (Derryberry *et al.* 2014 ; Gay *et al.* 2008). Selon les modèles, ils peuvent alors résulter d'une contre-sélection des hybrides (on s'attend alors à une distribution bimodale voire trimodale des phénotypes) ou au contraire d'une supériorité des hybrides par rapport aux formes/espèces parentales dans une région géographique restreinte qui constitue une niche à laquelle les hybrides sont adaptés (Delmore *et al.* 2013). Dans ces systèmes, le génome est divisé en plusieurs compartiments : un qui « diffuse » librement, c'est-à-dire où les fréquences alléliques sont régies par l'équilibre entre sélection, migration et dérive, et qui présente (si existence) des clines larges, et un autre compartiment qui porte la présence de facteurs isolants au sein des populations d'origine (comme les réarrangements chromosomiques) et qui présente des clines plus abrupts. Pour les téosintes *parviglumis* et *mexicana*, il existe peu d'hybrides entre les deux sous-espèces, (Fukunaga *et al.* 2005 ; Matsuoka *et al.* 2002). Comme il existe au sein de chaque sous-espèce une variation d'altitude parallèle à la variation d'un grand nombre de facteurs environnementaux biotiques

et abiotiques, on s'attend à observer plutôt un patron de variation clinale liée à l'altitude. N'ayant pas de relevés phénotypiques le long de ces clines, nous nous sommes focalisés ici uniquement sur les fréquences alléliques.

L'étude de clines nécessite un grand nombre de populations (De Mita *et al.* 2013). L'obtention de jeux de données moléculaires pour l'ensemble des populations représente un coût important, même avec les nouvelles technologies de séquençage. Afin de limiter ces coûts, un nombre réduit de marqueurs peut être utilisé en se concentrant sur des locus prometteurs d'un point de vue adaptatif. Nous avons basé le choix de ces locus « candidats » sur trois stratégies. Deux de ces stratégies proviennent de l'analyse des données de séquençage de six populations de téosinte décrites dans le chapitre 1 qui a permis de mettre en évidence des SNP et des régions du génome candidates à l'adaptation à l'altitude. La première stratégie a consisté à employer une batterie de tests SNP à SNP afin de détecter des traces de sélection et à s'appuyer sur la concordance entre certains des tests pour choisir les candidats. Nous avons notamment conditionné notre choix sur une méthode corrélative (BAYENV) qui teste la corrélation entre les fréquences alléliques des 6 populations et un jeu de variables environnementales résumé par l'axe 1 d'une ACP (cf. section II.4.1). Dans la seconde stratégie, une méthode de Fst-scan le long du génome a été utilisée. Cette fois, nous nous sommes uniquement basés sur la congruence d'un signal de différenciation le long du génome pour identifier des régions candidates. Ces deux approches de « reverse-ecology » permettent d'identifier des cibles potentielles de l'adaptation à l'altitude sans *a priori* sur les caractères phénotypiques et gènes sous-jacents impliqués dans l'adaptation locale. Une troisième stratégie a consisté à tirer profit des connaissances sur l'adaptation à l'altitude ou sur les bases génétiques de caractères adaptatifs chez le maïs et la téosinte. Une synthèse bibliographique a été réalisée. Des locus provenant de cette synthèse ont également été sélectionnés.

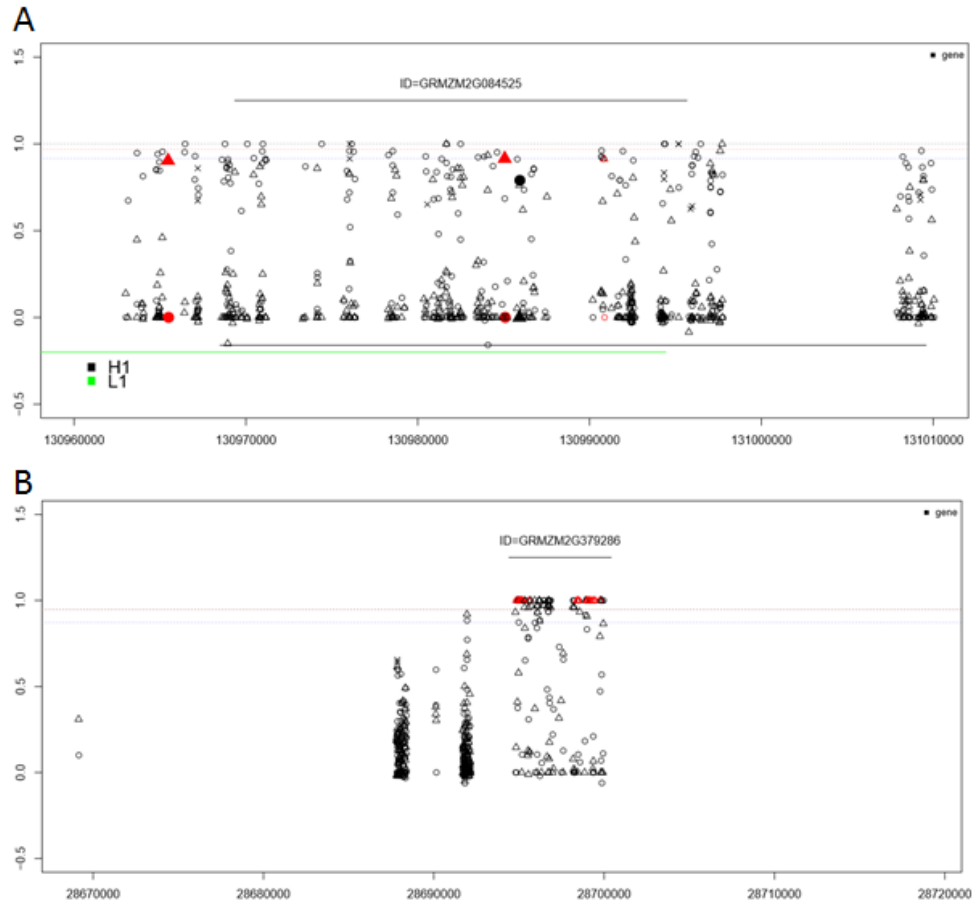
Au total, 270 SNP ont été sélectionnés pour être génotypés dans 37 populations de téosintes. La première partie de ce chapitre décrit le choix des cibles candidates par les trois stratégies employées ainsi que le travail de validation technique des SNP choisis par la plateforme de génotypage KASPar. La seconde partie montre l'analyse des données brutes de génotypage. La troisième et dernière partie est consacrée à l'analyse des clines de fréquence de ces SNP candidats.

## **III.2 OBTENTION DES 270 SNP CANDIDATS**

### **III.2.1 Cohérence des signaux de détection de la sélection**

L'analyse des données de séquençage de six populations présentées dans le chapitre 1 a permis de mettre en évidence des régions et des SNP candidats à l'adaptation. Pour cela, plusieurs méthodes de détection de la sélection ont été employées, des tests de différenciation entre populations (*F<sub>st</sub>*, Weir & Hill 2002 ; Bayescan, Foll & Gaggiotti 2008), un test de sélection intra-populationnel basé sur le niveau de diversité le long du génome (Pool-HMM, Boitard *et al.* 2012), et un test de corrélation entre les fréquences alléliques et une variable environnementale (BAYENV2, Günther & Coop 2013). L'emploi de *F<sub>st</sub>*, de Bayescan et de Pool-HMM a permis de mettre en évidence 166 061 SNP outliers sur un total de 8 479 581 SNP communs aux six populations. BAYENV2 a ensuite été utilisé sur ces SNP outliers pour détecter les corrélations avec une variable environnementale synthétique liée à l'altitude (section II.4.1).

Des régions génomiques (appelées cibles) ont été choisies lorsqu'un SNP outlier détecté par BAYENV2 était présent (il était alors le plus souvent détecté par une autre méthode), où lorsque la zone présentait une forte densité de SNP outliers (5 SNP à moins de 10 000 bases détectés par une ou plusieurs méthodes), où bien lorsque deux SNP identifiés par BAYENV2 ou Bayescan étaient situés à moins de 1000 bases. Les résultats de Pool-HMM ont également servi de support au choix des cibles, notamment pour l'exclusion de régions présentant des indices de balayages sélectifs congruents pour les trois populations de la même sous-espèce, i.e. pouvant indiquer une divergence entre sous-espèces plutôt qu'un patron d'adaptation. Au total 605 cibles ont été analysées et 93 retenues (Tableau 2). Deux exemples sont présentés dans la Figure 22.



**Figure 22 : Cibles identifiées avec les méthodes de détection de sélection.** Chaque SNP est visualisé par ses valeurs de  $F_{st}$  entre les populations extrêmes du gradient 1 (rond) et du gradient 2 (triangle). Les SNP outliers pour BayeScan sont représentés en rouge, ceux outliers pour BAYENV2 sont représentés en plein. A. Cible située sur le chromosome 9 regroupant des SNP outliers pour BAYENV2, BayeScan et  $F_{st}$ . Le SNP 130990858, outlier pour BayeScan a été sélectionné pour le génotypage KASPar. B. Cible située sur le chromosome 8. La cible regroupe des SNP outliers BayeScan et  $F_{st}$  condensés au sein d'un même gène. Le SNP 28699032, significatif à la fois pour BayeScan et  $F_{st}$  a été sélectionné pour le génotypage KASPar.

A l'intérieur de ces zones, les SNP ont été classés en trois catégories d'importance décroissante : SNP identifiés par différentes méthodes, SNP identifiés par une méthode, SNP non identifiés. Le choix des SNP à l'intérieur des cibles a été basé en partie sur ces catégories (cf section III.2.4).

### III.2.2 $F_{st}$ -scan

Les régions génomiques montrant des différenciations importantes entre populations poussant dans des environnements contrastés ont été considérées comme des régions candidates à l'adaptation. Afin de détecter ces régions, les valeurs de  $F_{st}$  des SNP communs aux six populations (SNP<sup>6</sup>, section II.3.3) ont été moyennées sur 20 SNP consécutifs avec un pas de 5 SNP. Ces moyennes ont été calculées pour les  $F_{st}$  entre populations extrêmes des deux

gradients (H1-L1 et H2-L2), entre populations de la même sous-espèce poussant à des altitudes contrastées pour le gradient 1 (*mexicana* H1-IM1, *parviglumis* L1-IP1), ainsi qu'entre les deux populations intermédiaires appartenant à des sous-espèces différentes (IM1-IP1) (Figure 23). En moyenne, les *Fst* entre populations extrêmes des gradients sont plus élevés que les *Fst* intra-sous-espèce ( $F_{st_{H1-L1}}$  0.119 ;  $F_{st_{H2-L2}}$  0.122 ;  $F_{st_{H1-IM1}}$  0.048 ;  $F_{st_{L1-IP1}}$  0.058).

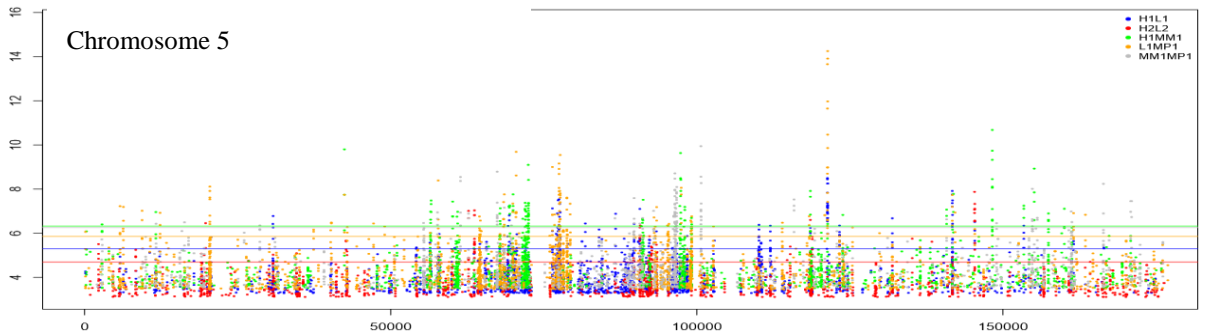
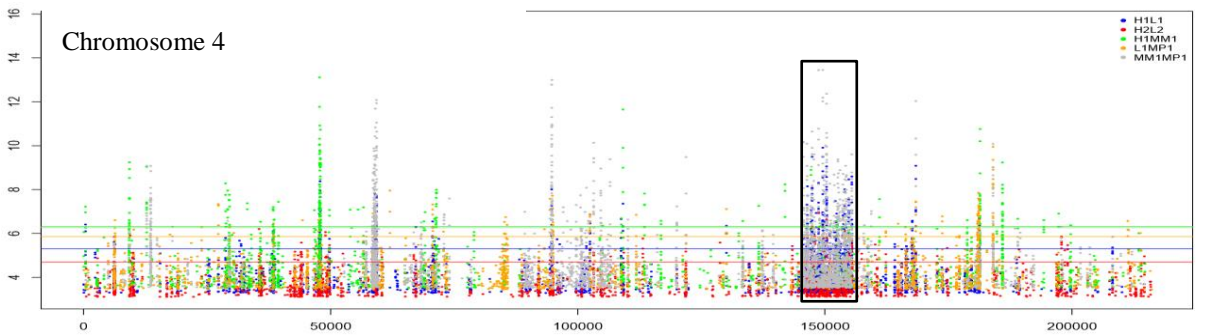
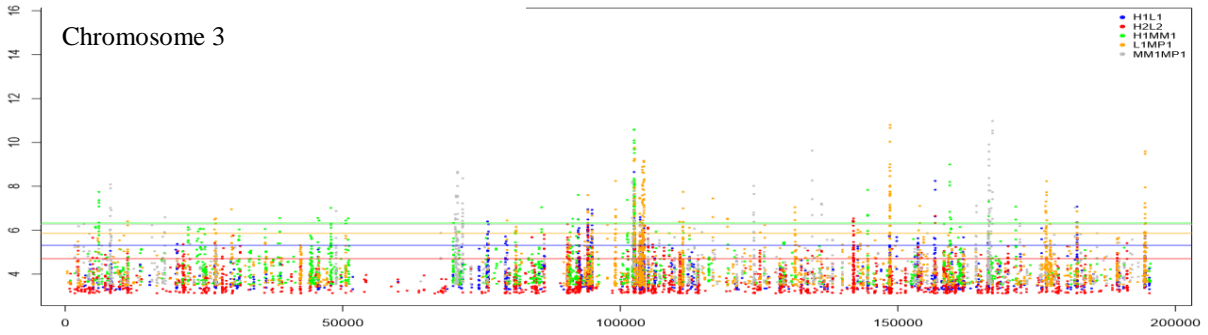
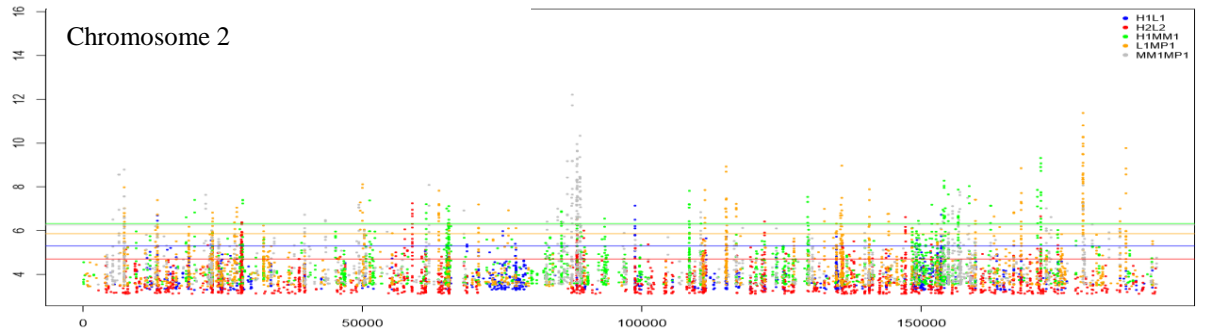
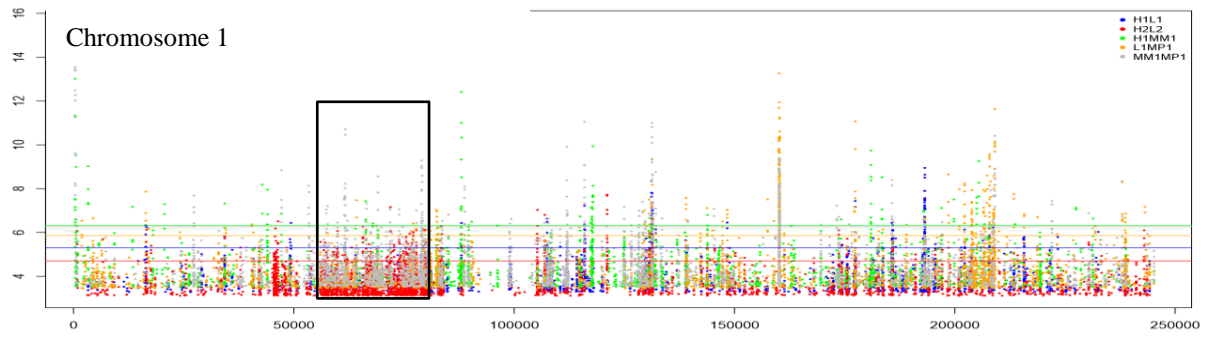
Afin d'identifier les régions candidates à l'adaptation liée à l'altitude, nous avons recherché les régions présentant des *Fst* élevés au sein des gradients 1 et/ou 2, ainsi qu'une différenciation importante au sein d'une ou des deux sous-espèces mais pas entre sous-espèces. Pour cela, deux seuils ont été définis : un seuil stringent de 1‰ permettant d'identifier les régions les plus différenciées entre deux populations, puis un seuil plus bas de 5‰ afin de vérifier si les autres couples de populations suivaient le même patron. L'indice de diversité de Nei (1973), qui représente l'hétérozygotie moyenne des populations, a été calculé sur *L* locus bialléliques de fréquence allélique  $p_i$  :

$$Div = \frac{\sum_{i=1}^L p_i - p_i^2}{L}$$

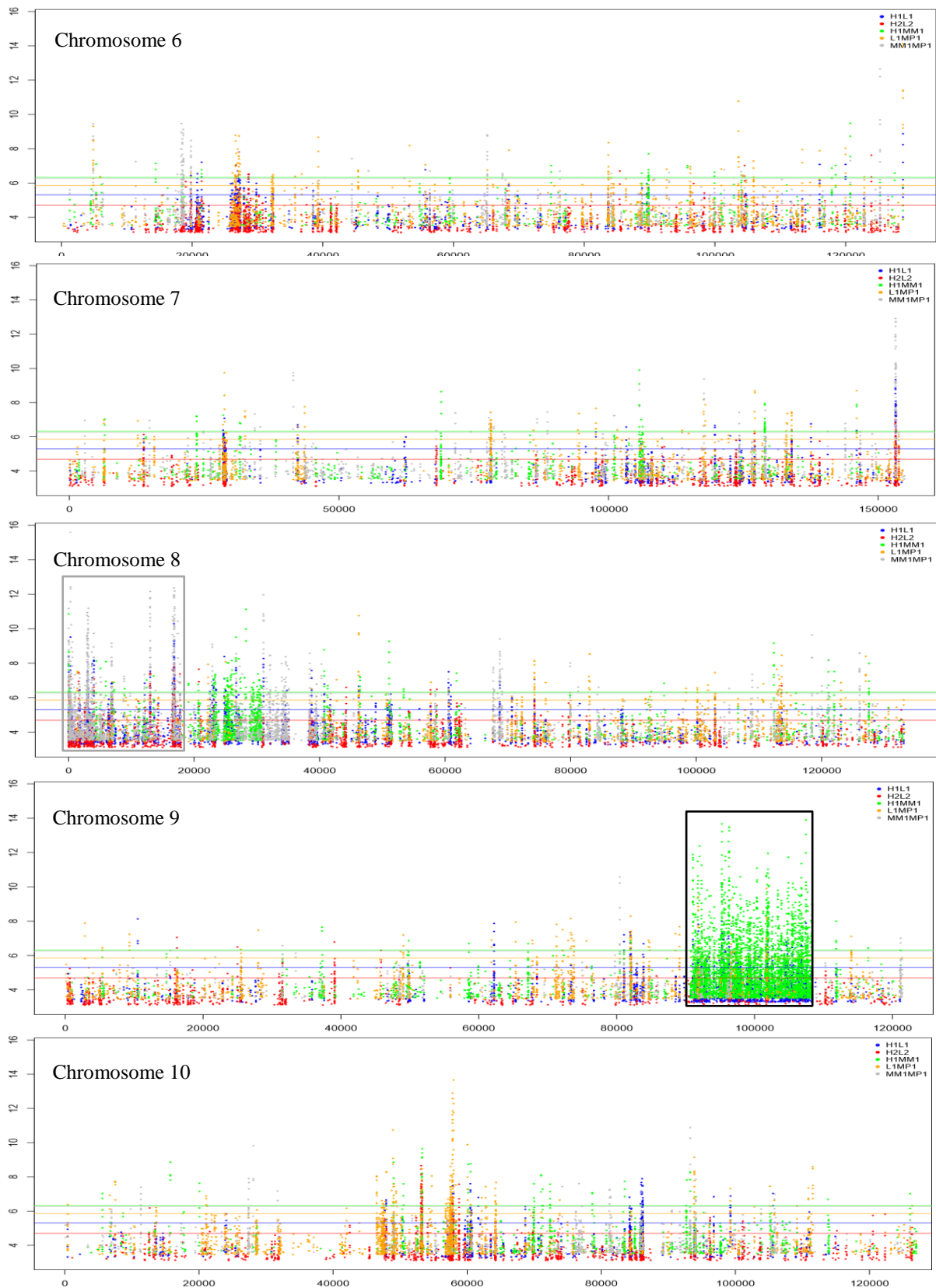
La variation de la valeur de cet indice le long des chromosomes a été évaluée pour chaque population de la même manière que pour les *Fst*, c'est-à-dire par une valeur moyenne sur 20 SNP consécutifs avec un pas de 5 SNP. Cette diversité moyenne est comprise entre 0.10 et 0.15 pour les six populations (H1 0.14 ; L1 0.13 ; H2 0.15 ; L2 0.10 ; IM1 0.10 ; IP1 0.11). Lorsqu'une région présentait une différenciation entre populations extrêmes, nous avons vérifié qu'au moins une des populations intermédiaires était polymorphe. Nous avons cherché en priorité les régions différenciées entre populations extrêmes du gradient 1 (pics bleus, Figure 23) et/ou du gradient 2 (pics rouges, Figure 23) en excluant celles fortement différenciées entre les deux populations intermédiaires (pics gris, Figure 23). Nous avons également choisi pour le gradient 1 des régions qui étaient fortement différenciées au sein d'une ou des deux sous-espèces, que ce soit *mexicana* (pics verts, Figure 23) ou *parviglumis* (pics orange, Figure 23).

Certains fragments de chromosomes présentent des patrons très homogènes (Figure 23, chromosomes 1, 4, 8 et 9) typiques d'événements génomiques importants tels que des inversions, au sein desquelles le déséquilibre de liaison est important. Trois de ces régions, nommées *Inv1n*, *Inv4m* et *Inv9e* et localisées sur les chromosomes 1, 4 et 9 respectivement, correspondent à des inversions précédemment décrites par Fang *et al.* (2012)

et Pyhäjärvi *et al.* (2013). Elles sont signalées par un rectangle noir dans la Figure 23. *Inv1n* et *Inv4m* montrent un patron de différenciation entre les sous-espèces alors que *Inv9e* montre une différenciation importante au sein de *mexicana* également reportée par Pyhäjärvi *et al.* (2013). Nous avons donc choisi plusieurs SNP au sein de cette inversion. La région localisée sur le chromosome 8 a déjà été identifiée (Fang *et al.* 2012 ; Pyhäjärvi *et al.* 2013) mais n'a été pas considérée comme une inversion mais plutôt comme une erreur d'assemblage sur le génome de référence du maïs. Cette région est représentée par un rectangle gris sur la Figure 23.

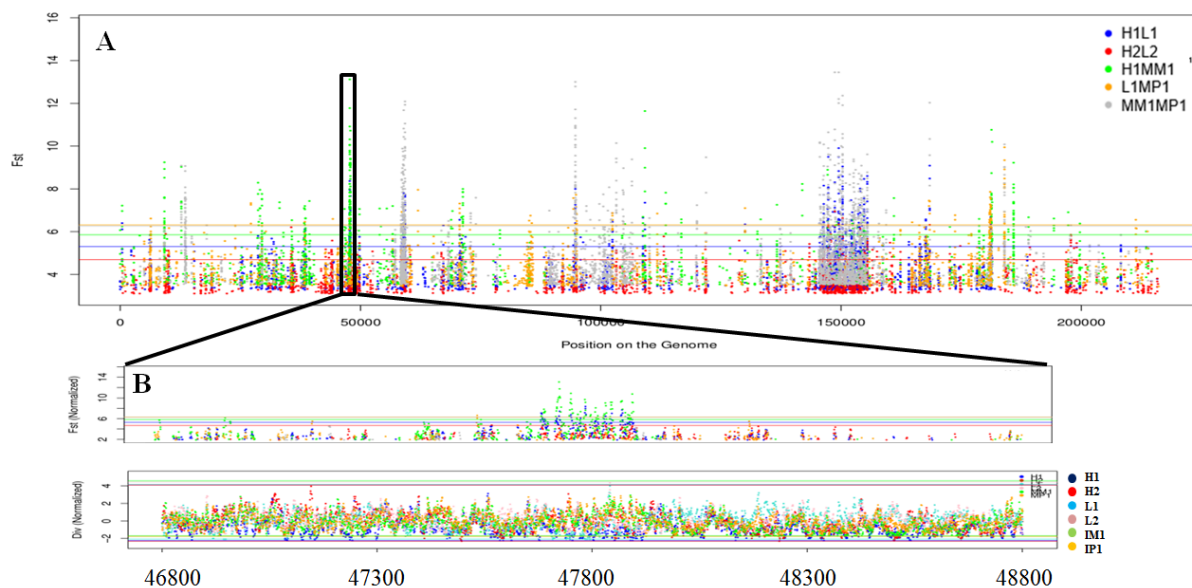






**Figure 23 : Fst par Chromosome.** Chaque point correspond au Fst moyen sur 20 SNP consécutifs. L'abscisse donne la position moyenne des 20 SNP sur le chromosome. Seules les 5% valeurs de Fst les plus élevées sont représentées. Les barres horizontales représentent les seuils à 1%. Les cadres noirs indiquent la position des inversions, le cadre gris la région reportée comme une erreur d'assemblage. Pour une meilleure visualisation, les valeurs de Fst ont été normalisées pour chaque couple de population.

Nous avons analysé plus en détail un total de 178 régions afin de confirmer la présence ou non d'un pic de  $F_{st}$  localisé (Figure 24).



**Figure 24 : Exemple de cible sur le chromosome 4.** Chaque point correspond à une moyenne de  $F_{st}$  sur 20 SNP consécutifs. L'abscisse indique la position moyenne des 20 SNP. Seules les valeurs dans le top 1% sont représentées. Les barres horizontales représentent les seuils à 1%. Pour une meilleure visualisation, les valeurs de  $F_{st}$  ont été normalisées pour chaque couple de population. Une région repérée sur le chromosome 4 (A) a ensuite été visualisée à une échelle plus fine afin de confirmer un pic localisé de  $F_{st}$  (B). Ici, le SNP 5 190 230, faisant parti des points moyennés aux alentours du pic situé à 47 800 pas a été choisi pour être génotypé avec la méthode KASPar.

### III.2.3 Synthèse Bibliographique

Afin de tirer profit des études précédentes sur l'adaptation des téosintes et du maïs, et sur les bases génétiques des caractères impliqués dans l'adaptation, une synthèse bibliographique a été réalisée (Annexe III.1). Cette synthèse concerne en majorité des études « Top-down », du phénotype vers le génotype (telles que les études de détection de QTL et de génétique d'association), qui ont mis en évidence des locus ou gènes codant des caractères potentiellement adaptatifs chez les téosintes, ou des locus/gènes connus pour être impliqués dans les processus de domestication et/ou d'amélioration du maïs. Dans l'Annexe III.1, nous avons distingué les études d'association (association mapping), de celles de QTL (QTL mapping), et de celles qui couplent les deux dans le dispositif de Nested Association Mapping (NAM, (Yu *et al.* 2008). Nous avons signalé les études qui sont allées jusqu'à la validation fonctionnelle des gènes (functional) et celles dans lesquelles les gènes candidats ont été validés par une étude d'expression (expression). Enfin, nous avons mentionné les cas où des gènes candidats de fonctions connues ont fait l'objet d'études d'évolution moléculaire

(molecular evolution) qui ont permis *a posteriori* de proposer leur implication dans la variation de caractères considérés.

Nous nous sommes concentrés plus particulièrement sur des caractères potentiellement liés à l'adaptation à l'altitude. Les téosintes poussant en altitude sont confrontées à des conditions spécifiques : étés courts, saisonnalité importante, précipitations rares et rayonnements importants. On s'attend donc à un cycle végétatif court et une forte précocité de floraison avec des morphologies caractérisées par une taille réduite, une pilosité abondante propice au maintien de l'humidité et offrant une meilleure protection au froid, et une coloration vive des organes reproducteurs conférant une protection vis-à-vis des rayonnements. Certains de ces caractères ont été répertoriés comme distinguant les deux sous-espèces. La sous-espèce *mexicana* possède par exemple de nombreux poils sur les tiges et structures foliaires (Lauter *et al.* 2004; Moose *et al.* 2004), des grains et des épillets plus gros et des panicules moins ramifiée que *parviglumis* (Hufford *et al.* 2012). Il est difficile de conclure si ces différences phénotypiques sont adaptatives ou causées par la dérive. Cependant, certains caractères, comme la pigmentation et les poils sont également trouvés chez les maïs de haute altitude, présumant un caractère adaptatif obtenu par convergence ou introgression (Takuno *et al.* 2015).

Les téosintes présentent une variabilité morphologique importante : certaines peuvent avoir une morphologie buissonnante et une taille modérée, d'autres peuvent atteindre des tailles supérieures à 4 mètres et présenter plusieurs talles, ou encore présenter une talle unique et une taille moyenne. La morphologie des feuilles est également variables, certaines téosintes présentant des feuilles allongées, d'autres plus courtes et larges. Bien qu'encore non reliés directement à l'adaptation à l'altitude, les locus régissant l'architecture végétative de la plante chez le maïs et/ou la téosinte ont été incorporés à la synthèse bibliographique (Briggs *et al.* 2007 ; Chuck *et al.* 2007 ; Clark *et al.* 2006 ; Doebley *et al.* 1995 ; Malosetti *et al.* 2008 ; Peiffer *et al.* 2014 ; Studer & Doebley 2011 ; Tian *et al.* 2011 ; Weber *et al.* 2007 ; Weber *et al.* 2008 ; Whipple *et al.* 2011 ; Yu *et al.* 2012 ; Annexe III.1). L'architecture de l'inflorescence des téosintes est également grandement variable et présente de grandes différences avec le maïs cultivé. Les études mettant en évidence les locus impliqués dans l'architecture de l'inflorescence chez la téosinte et/ou le maïs ont également été incorporées (Agrama & Moussa 1996 ; Briggs *et al.* 2007 ; Chen *et al.* 2014 ; Chuck *et al.* 2007 ; Doebley & Stec 1993 ; Gallavotti *et al.* 2004 ; Malosetti *et al.* 2008 ; Sigmon & Vollbrecht 2010 ; Wang *et al.* 2005 ; Weber *et al.* 2008 ; Yu *et al.* 2012 ; Annexe III.1). Enfin nous avons inclus des études concernant le rendement (Malosetti *et al.* 2008) et la composition des grains

(Bhave *et al.* 1990 ; Cook *et al.* 2012 ; Hirsch *et al.* 2014 ; Lipka *et al.* 2013 ; Manicacci *et al.* 2009 ; Pouvreau *et al.* 2011 ; Tracy *et al.* 2006 ; Vicente-Carbajosa *et al.* 1997 ; Weber *et al.* 2008 ; Whitt *et al.* 2002 ; Wilson *et al.* 2004).

Deux caractéristiques abiotiques liées à l'altitude sont le froid et la sécheresse. Nous avons donc inclus dans l'Annexe III.1 des locus de tolérance au froid (Anderson *et al.* 1994 ; Berberich & Kusano 1997 ; Kusano *et al.* 1995 ; Steward *et al.* 2000), ainsi que des locus de tolérance à la sécheresse (Agrama & Moussa 1996 ; Setter *et al.* 2010 ; Welcker *et al.* 2011).

Récemment, des études ont suggéré un rôle important des variables édaphiques dans l'adaptation locale (O'Brien *et al.* 2015 ; Takuno *et al.* 2015). En particulier, les téosintes de haute altitude poussent sur des sols volcaniques acides, pauvres en phosphate bio-disponible et riches en aluminium. Chez le maïs, il existe des QTL de tolérance à l'aluminium (Maron *et al.* 2013 ; Ninamango-Cárdenas *et al.* 2003 ; Sibov *et al.* 1999), ainsi que des QTL de tolérance à de faibles niveaux de phosphate (Azevedo *et al.* 2015 ; Baek *et al.* 2013 ; Reiter *et al.* 1991 ; Shin *et al.* 2004 ; Svistoonoff *et al.* 2007) et d'azote (Bertin & Gallais 2001 ; Coque & Gallais 2006 ; Gamuyao *et al.* 2012 ; Malosetti *et al.* 2008) que nous avons incorporés à l'Annexe III.1. Des locus contrôlant la morphologie racinaire ont aussi été pris en compte (Azevedo *et al.* 2015 ; Zhang *et al.* 2014). L'adaptation aux conditions de basse altitude nécessite une tolérance à la chaleur (Nieto-Sotelo *et al.* 2002) et aux sols détremés (Carlson *et al.* 2002 ; Dordas *et al.* 2003 ; Qiu *et al.* 2007 ; Subbaiah & Sachs 2001) afin d'éviter l'asphyxie de la plante et la verse.

L'adaptation locale peut également avoir lieu en réponse à une variation d'environnement biotique, par exemple via des gènes de défense contre l'herbivorie ou les parasites (Crémieux *et al.* 2008 ; Kawecki & Ebert 2004). Les locus de résistance aux insectes et aux parasites ainsi que des gènes d'immunité ont également été incorporés à l'Annexe III.1 (Butrón *et al.* 2010 ; Chung *et al.* 2010 ; Frey *et al.* 2003 ; Jiang *et al.* 1995 ; Kyetere *et al.* 1999 ; Leng *et al.* 2015 ; McMullen & Simcox 1995 ; McMullen & Louie 1989 ; McMullen *et al.* 1994 ; Meihls *et al.* 2013 ; Meyer *et al.* 2007 ; Ming *et al.* 1997 ; Moeller & Tiffin 2008 ; Nomura *et al.* 2003 ; Ogliari *et al.* 2005 ; Pechan *et al.* 1999 ; Poland *et al.* 2011 ; Quint *et al.* 2002 ; Simcox & Bennetzen 1993 ; Simmons *et al.* 2001 ; Szalma *et al.* 2005 ; Tao *et al.* 2013 ; Tzin *et al.* 2015 ; van der Linde *et al.* 2012 ; Welz *et al.* 1999 ; Xia *et al.* 1999 ; Yan *et al.* 2014 ; Yin *et al.* 2003 ; Zhang *et al.* 2002 ; Zila *et al.* 2014).

Finalement, les gènes de domestication pouvant être des cibles de la sélection naturelle chez les téosintes (Ross-Ibarra *et al.* 2007), ils ont également été incorporés à l'Annexe III.1 (Hirsch *et al.* 2014 ; Hung *et al.* 2012 ; Wright *et al.* 2005).

Nous souhaitons génotyper environ 90 SNP identifiés dans la littérature. Parmi ceux-ci, nous avons intégré des SNP provenant de locus sous-jacents à des phénotypes potentiellement impliqués dans l'adaptation à l'altitude. Nous y avons ajouté des SNP impliqués dans la domestication et l'amélioration du maïs détectés par des approches moléculaires, ainsi que des SNP candidats détectés dans des études antérieures ou en cours sur l'adaptation locale de la téosinte. Nous avons notamment intégré des SNP provenant d'une inversion située sur le chromosome 1 qui est potentiellement adaptative (Fang *et al.* 2012 ; Pyhäjärvi *et al.* 2013), ainsi que des SNP provenant de régions candidates détectées par Jonas Aguirre *et al.* (soumis). Le nombre de cibles potentielles excédant 90, nous les avons analysées à l'aide des méthodes de détection de la sélection, afin de choisir dans ces cibles des SNP présentant des signaux de sélection au niveau des six populations séquencées dans le chapitre 1 pour le génotypage KASPar. Nous avons récupéré les cibles présentant des signaux de sélection, puis classé les SNP en différentes catégories selon qu'ils présentaient un fort signal de sélection (catégorie 1) ou un signal de sélection moyen (catégorie 2), ou qu'ils étaient situés proches d'un SNP sous sélection (catégorie 3).

### **III.2.4 Génotypage par la méthode KASPar**

La technologie KASPar permet de génotyper des individus à des positions connues pour être polymorphes et bialléliques. Le génotypage s'effectue au moyen d'une PCR compétitive entre les deux formes alléliques existantes (cf section I.1.3.a). Des amorces complémentaires des deux allèles sont construites en amont de la PCR compétitive, nécessitant de connaître les séquences flanquantes des SNP ciblés. Le LGC Group s'engage à génotyper des SNP dont les 50 positions de part et d'autre du SNP cible ne comportent aucun INDEL et au maximum un SNP. Or, de tels critères sont très rarement remplis chez la téosinte qui présente un niveau de polymorphisme très élevé. En regard du fort taux d'échec de génotypage des SNP par KASPar lorsque ces critères ne sont pas remplis (communications personnelles, Matthieu Falque et Yves Vigouroux), le choix des 270 SNP a été effectué en étroite collaboration avec le LGC Group. De nouveaux critères ont été définis (pas d'INDEL dans les 20 bases de part et d'autres du SNP cible), et des prédictions de génotypage (probabilité de succès) ont été fournies. Au total, 6 864 SNP ont été testés par ordre décroissant d'intérêt : les SNP de catégorie 1 ont été testés en premier, en cas d'échec de ces SNP, les SNP de la catégorie 2 ont été testés en remplacement. En cas d'échec des SNP de ces deux catégories, la cible a été abandonnée. 270 SNP ont été retenus en combinant l'intérêt de la cible et les prédictions de qualité déterminées par LGC.

Sur les 270 SNP à génotyper, 93 proviennent des SNP identifiés sous sélection suite aux analyses effectuées sur les six populations (SNP1), 97 proviennent des scans génomiques de *Fst* (SNP2), et 80 proviennent de la synthèse bibliographique (SNP3). Les SNP identifiés par les méthodes de détection de la sélection et par les scans génomiques de *Fst*, et qui tombent dans des locus identifiés dans l'Annexe III.1, ont été indiqués dans le tableau de cette annexe. Les notes des prédictions de génotypage LGC varient entre 24,7 et 431,1 (Tableau 2). Les SNP présentant des prédictions de génotypage supérieures à 300 ont été conservés uniquement si la région cible présentait un fort intérêt et s'il n'y avait pas d'autres SNP à plus faible valeur de prédiction. La couverture du génome par nos SNP n'est pas homogène : un chromosome est couvert par 5 SNP (chromosome 10) à 38 SNP (chromosome 5). Les SNP candidats consécutifs sont distants de 7,24 Mb en moyenne, les plus proches (Fst\_93 et SB\_76) étant distants de 27 pb et les plus éloignés (Fst\_81 et SB\_69) de 68,25 Mb (Figure 25). Il est intéressant de noter que les SNP1 sont plus clusterisés que les SNP2 et SNP3, nous avons conditionnés dans certains cas SNP1 sur le fait qu'il y ait un regroupement de SNP candidats et pris des candidats dans des cibles très proches les unes de autres (Figure 25).

**Tableau 2 : Origine de 270 SNP candidats génotypés par KASPar avec leur Position Chromosomique et leur Annotation dans l'Annexe III.1 pour les SNP3. L'identité des SNP (SNPID) est définie par la méthode de détection et le numéro du SNP.**

SNPID	Position	Chro <sup>1</sup>	Prédiction	Annotation <sup>2</sup>
MS_1	168286704	1	82,4	-
MS_2	168348856	1	57,7	-
MS_3	179343553	1	105	-
MS_4	234782949	1	48,8	-
MS_5	236806700	1	87,6	-
MS_6	90318615	2	317,6	-
MS_7	165616363	2	131,2	-
MS_8	229255903	2	374,4	-
MS_9	9584508	3	38,4	-
MS_10	9708557	3	75,4	-
MS_11	62335914	3	312,7	-
MS_12	62488387	3	403,1	-
MS_13	64509459	3	110,2	-
MS_14	109095362	3	62,5	-
MS_15	111561600	3	179,9	-

5				
MS_1 6	47564793	4	328,6	-
MS_1 7	170388650	4	136,1	-
MS_1 8	170711055	4	78,5	-
MS_1 9	170783306	4	77	-
MS_2 0	171111841	4	288,9	-
MS_2 1	173995149	4	72,6	-
MS_2 2	177881170	4	162,6	-
MS_2 3	180504078	4	58,1	-
MS_2 4	181223878	4	58,8	-
MS_2 5	225087669	4	336,8	-
MS_2 6	101940822	5	91,6	-
MS_2 7	102660991	5	89	-
MS_2 8	103126178	5	56	-
MS_2 9	103151251	5	72,8	-
MS_3 0	103575632	5	151	-
MS_3 1	103652454	5	170,3	-
MS_3 2	104653139	5	155,6	-
MS_3 3	105397238	5	59,8	-
MS_3 4	107146543	5	116,5	-
MS_3 5	112610568	5	71,2	-
MS_3 6	112693361	5	106,2	-
MS_3 7	148344999	5	67,9	-
MS_3 8	45626361	6	73,6	-
MS_3 9	150171759	7	240,2	-

MS_4 0	18204	8	89,4	-
MS_4 1	21240	8	93,2	-
MS_4 2	79301	8	388,4	-
MS_4 3	228284	8	53,8	-
MS_4 4	260479	8	431,1	-
MS_4 5	356976	8	53,6	-
MS_4 6	400983	8	287,9	-
MS_4 7	559686	8	165,8	-
MS_4 8	618589	8	276,9	-
MS_4 9	825774	8	70,5	-
MS_5 0	828512	8	234,8	-
MS_5 1	939485	8	154,2	-
MS_5 2	1343730	8	45	-
MS_5 3	1372690	8	84,7	-
MS_5 4	2265471	8	289,4	-
MS_5 5	2289581	8	74,8	-
MS_5 6	2315211	8	69,2	-
MS_5 7	2350251	8	251,6	-
MS_5 8	3576402	8	104,8	-
MS_5 9	3979145	8	173,6	-
MS_6 0	4515370	8	57,6	-
MS_6 1	4598922	8	109,9	-
MS_6 2	4616707	8	85,3	-
MS_6 3	4801331	8	64,1	-
MS_6	4843324	8	163	-



4				
MS_6	5316402	8	101,8	-
5				
MS_6	5915456	8	108,9	-
6				
MS_6	6809333	8	326,5	-
7				
MS_6	8452261	8	251	-
8				
MS_6	13716313	8	153,7	-
9				
MS_7	17372541	8	326,8	-
0				
MS_7	28068889	8	87,3	-
1				
MS_7	28124172	8	82,2	-
2				
MS_7	28479127	8	48,5	-
3				
MS_7	28525840	8	54,3	-
4				
MS_7	28699032	8	56,2	-
5				
MS_7	28763790	8	56,5	-
6				
MS_7	28852684	8	51,7	-
7				
MS_7	29016489	8	81	-
8				
MS_7	29232368	8	47,2	-
9				
MS_8	30010033	8	292,8	-
0				
MS_8	32570959	8	87,5	-
1				
MS_8	43099181	8	65,8	-
2				
MS_8	52086769	8	170	-
3				
MS_8	55936851	8	84,4	-
4				
MS_8	56361271	8	55,2	-
5				
MS_8	56361366	8	228,5	-
6				
MS_8	67729118	8	259	-
7				
MS_8	130990858	9	140,5	-
8				

MS_8 9	135144815	9	100,6	-
MS_9 0	145592593	9	97,8	-
MS_9 1	145664206	9	49	-
MS_9 2	121003812	10	172,1	-
MS_9 3	123371760	10	68,6	-
SB_1	50033864	1	59,6	phyB1 flowering time
SB_2	51320026	1	31,7	p1 (pericarp color1) insect resistance, flavonoid content, leave coloration
SB_3	70582465	1	36,1	<i>Inv1n</i>
SB_4	103827827	1	85,4	<i>Inv1n</i>
SB_5	165413025	1	185,9	candidate SNP <i>parviglumis</i> (Aguirre <i>et al</i> , soumis)
SB_6	168412574	1	276,4	bhlh158 flowering time
SB_7	168414611	1	124,6	bhlh158 flowering time
SB_8	180944170	1	55,9	csu61b grain yield
SB_9	195761492	1	60	umc1035 Nitrogen use efficiency
SB_10	197386628	1	165	umc1335 Nitrogen use efficiency
SB_11	198735122	1	86,1	seed size,
SB_12	249318279	1	66,7	brd1 (brassinosteroid-deficient dwarf1) peiffer 2014
SB_13	252393890	1	102,6	bhlh43 flowering time
SB_14	262457397	1	60,3	seed size 7
SB_15	266029571	1	108	candidate SNP <i>mexicana</i> (Aguirre <i>et al</i> , soumis)
SB_16	41485594	2	210,6	cry2 -(crystal proteinIA(b)2) insect resistance
SB_17	143673112	2	67,3	seed size
SB_18	152169574	2	130,4	candidate SNP <i>parviglumis</i> (Aguirre <i>et al</i> , soumis)
SB_19	193191211	2	44,7	pebp18 (phosphatidylethanolamine-binding protein18)
SB_20	21318315	3	106,6	floury2 kernel composition
SB_21	84903125	3	37,3	candidate SNP <i>parviglumis</i> (Aguirre <i>et al</i> , soumis)
SB_22	111866508	3	51,1	candidate SNP <i>parviglumis</i> (Aguirre <i>et al</i> , soumis)
SB_23	112197934	3	76,3	candidate SNP <i>parviglumis</i> (Aguirre <i>et al</i> , soumis)
SB_24	198395013	3	247,2	root length, water logging tolerance
SB_25	201798588	3	148	flowering time/photoperiod sensitivity
SB_26	216419176	3	82	sh2 -shrunken2 starch content
SB_27	3150731	4	134,3	bx5 benzoxazinone synthesis5 insect resistance
SB_28	18573737	4	267,5	bx7 benzoxazinone synthesis7 insect resistance
SB_29	33118370	4	294,1	seed size

SB_30	41373553	4	66	su1 kernel content
SB_31	44539727	4	338,5	tga1 inflorescence trait
SB_32	45663714	4	106,7	candidate SNP <i>parviglumis</i> (Aguirre <i>et al</i> , soumis)
SB_33	46065852	4	285,3	seed size
SB_34	233810694	4	237,3	candidate SNP <i>parviglumis</i> (Aguirre <i>et al</i> , soumis)
SB_35	1358397	5	276,4	carotenoid isomerase pigmentation
SB_36	5103643	5	219,2	photoperiod response
SB_37	6381066	5	291,3	photoperiod response
SB_38	39893932	5	61,1	photoperiod response
SB_39	42301877	5	103,4	umc1315 water deficit stress
SB_40	49517639	5	227,1	amino acid biosynthesis
SB_41	64116218	5	64,5	ncsu4 (cca1) circadian clock associated1 (QTL) leaf number
SB_42	64427153	5	88,5	plant growth
SB_43	64775301	5	64,3	fusarium ear resistance
SB_44	65125729	5	73,5	plant growth
SB_45	92182318	5	92,6	cmu2 (chorismate mutase2)
SB_46	100073232	5	148,2	candidate SNP <i>mexicana</i> (Aguirre <i>et al</i> , soumis)
SB_47	148413102	5	49,4	water deficit stress
SB_48	215830190	5	56,3	lut5 carotenoid
SB_49	51228486	6	77,2	photoperiod response
SB_50	87120463	6	112,9	eif4A flowering time
SB_51	88593564	6	250	photoperiod response
SB_52	104503351	6	53,7	elfg1 anthesis-silking interval
SB_53	13846752	7	49	carotenoid
SB_54	143113709	7	83,5	photoperiod response
SB_55	174983719	7	44,4	photoperiod response
SB_56	4957045	8	113	photoperiod response
SB_57	40263865	8	84	nitrogen paucity, drought tolerance
SB_58	44447408	8	63,2	nitrogen paucity, drought tolerance
SB_59	45263545	8	42,8	candidate SNP <i>mexicana</i> (Aguirre <i>et al</i> , soumis)
SB_60	47311640	8	39,5	nitrogen paucity, drought tolerance
SB_61	51741986	8	61,3	photoperiod response
SB_62	56453845	8	60,5	nitrogen paucity, drought tolerance
SB_63	68313700	8	37,5	candidate SNP <i>parviglumis</i> (Aguirre <i>et al</i> , soumis)
SB_64	72300837	8	43,6	candidate SNP <i>parviglumis</i> (Aguirre <i>et al</i> , soumis)
SB_65	124654224	8	111,7	photoperiod response
SB_66	132878592	8	71,8	atdof1.8 (QTL) flowering time/photoperiod sensitivity

SB_67	136050565	8	39,2	sdw1 water logging tolerance
SB_68	161862471	8	40,4	elongated mesocotyl1 elm1 flowering time
SB_69	119304019	9	93,5	candidate SNP <i>mexicana</i> (Aguirre <i>et al</i> , soumis)
SB_70	124233766	9	33,3	umc38c number of ears per plant
SB_71	126017135	9	45,8	candidate SNP <i>mexicana</i> (Aguirre <i>et al</i> , soumis)
SB_72	126940455	9	62,9	drough tolerance (Malosetti)
SB_73	136302086	9	63,2	candidate SNP <i>mexicana</i> (Aguirre <i>et al</i> , soumis)
SB_74	136573380	9	59,4	csu634 water deficit stress
SB_75	138613085	9	45,3	photoperiod response
SB_76	139382337	9	47,7	candidate SNP <i>mexicana</i> (Aguirre <i>et al</i> , soumis)
SB_77	140781294	9	87	candidate SNP <i>mexicana</i> (Aguirre <i>et al</i> , soumis)
SB_78	141470887	9	78	candidate SNP <i>mexicana</i> (Aguirre <i>et al</i> , soumis)
SB_79	145528034	9	109,5	hb1 hemoglobin1 water flood
SB_80	145765201	9	100,1	candidate SNP <i>mexicana</i> (Aguirre <i>et al</i> , soumis)
Fst_1	15514733	1	403,7	-
Fst_2	35452915	1	126,6	-
Fst_3	49815938	1	38,8	-
Fst_4	54922664	1	259,5	-
Fst_5	137979331	1	116,7	-
Fst_6	168348943	1	27,8	-
Fst_7	227989956	1	227,4	-
Fst_8	228005584	1	61,7	-
Fst_9	236818610	1	56,8	-
Fst_10	243023543	1	61,1	-
Fst_11	258688386	1	319,7	-
Fst_12	268560405	1	72,5	-
Fst_13	12442449	2	97,2	-
Fst_14	31448551	2	52,8	-
Fst_15	72456999	2	231,5	-
Fst_16	74084625	2	97,7	-
Fst_17	83127028	2	171,3	-
Fst_18	133085160	2	146,6	-
Fst_19	146125013	2	84,5	-
Fst_20	163413504	2	95,9	-
Fst_21	189968292	2	38,5	-
Fst_22	217634370	2	60,1	-
Fst_23	230520562	2	33,2	-
Fst_24	8606769	3	46,8	-
Fst_25	14375497	3	94,1	-
Fst_26	40930358	3	88,3	-
Fst_27	110514686	3	140,9	-
Fst_28	112158024	3	204,5	-

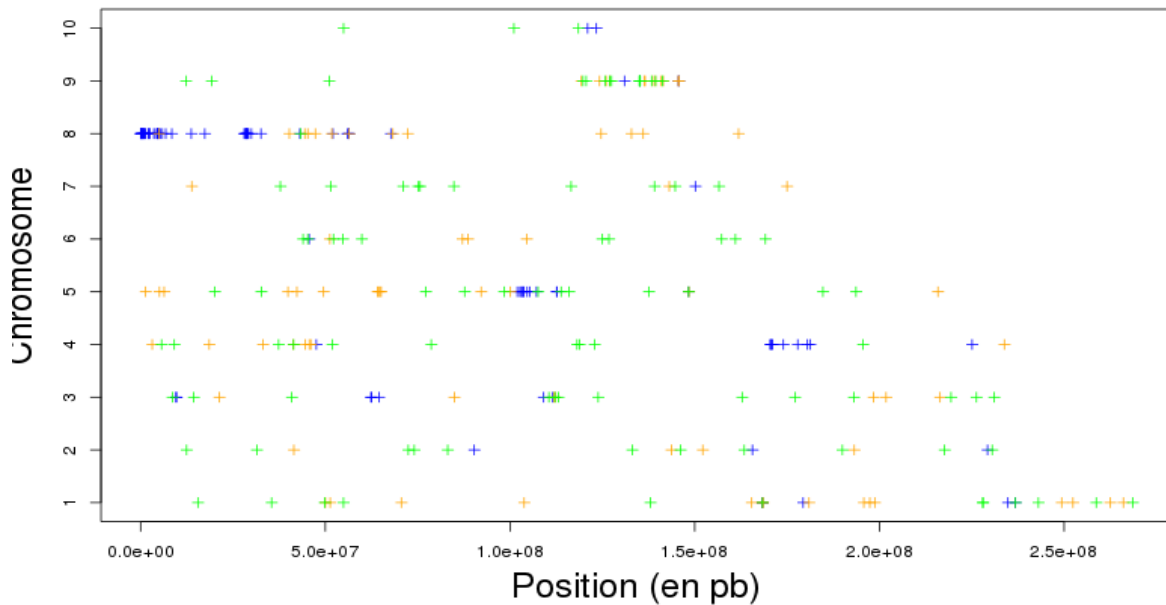
Fst_29	113076088	3	57,2	-
Fst_30	123867416	3	51,7	-
Fst_31	162828569	3	77	-
Fst_32	177124534	3	25,1	-
Fst_33	193062132	3	90,7	-
Fst_34	219383329	3	71,3	-
Fst_35	226255690	3	98,6	-
Fst_36	231056268	3	82,2	-
Fst_37	5666978	4	66,7	-
Fst_38	9126461	4	55,6	-
Fst_39	37352326	4	103,2	-
Fst_40	41405883	4	51,9	-
Fst_41	51902320	4	96,5	-
Fst_42	78668023	4	85,5	-
Fst_43	118141492	4	72,3	-
Fst_44	118825345	4	144,9	-
Fst_45	122884528	4	33,5	-
Fst_46	195551119	4	90,1	-
Fst_47	20080442	5	90,2	-
Fst_48	32661209	5	73,1	-
Fst_49	77196738	5	60,9	-
Fst_50	87715294	5	88,4	-
Fst_51	98391122	5	54,9	-
Fst_52	107601473	5	62,1	-
Fst_53	113828969	5	33,3	-
Fst_54	115978816	5	46,5	-
Fst_55	137582803	5	152,9	-
Fst_56	148348395	5	83,9	-
Fst_57	184701046	5	57,4	-
Fst_58	193653062	5	62,3	-
Fst_59	43975209	6	158,4	-
Fst_60	45242004	6	84,4	-
Fst_61	52199461	6	45,9	-
Fst_62	54776449	6	82,7	-
Fst_63	59881283	6	67,2	-
Fst_64	124971371	6	59,6	-
Fst_65	126780736	6	198,3	-
Fst_66	157274662	6	110	-
Fst_67	161009080	6	288,8	-
Fst_68	169128855	6	145,1	-
Fst_69	37836093	7	65,8	-
Fst_70	51505486	7	115,9	-
Fst_71	71077552	7	99,6	-
Fst_72	75099491	7	271,6	-

Fst_73	75676145	7	275,1	-
Fst_74	84801067	7	27,4	-
Fst_75	116544684	7	45,2	-
Fst_76	139116154	7	83	-
Fst_77	144729861	7	51,6	-
Fst_78	156573041	7	33,3	-
Fst_79	43297237	8	37,3	-
Fst_80	12269764	9	54	-
Fst_81	19232016	9	89,3	-
Fst_82	51050862	9	104,1	-
Fst_83	119766875	9	160,8	-
Fst_84	120626342	9	163,2	-
Fst_85	125690268	9	114,3	-
Fst_86	127001582	9	67,2	-
Fst_87	127497303	9	143,6	-
Fst_88	134996351	9	66,9	-
Fst_89	135082028	9	24,7	-
Fst_90	135153323	9	58,8	-
Fst_91	135258226	9	37,7	-
Fst_92	138324711	9	103,5	-
Fst_93	139382364	9	89,9	-
Fst_94	141331982	9	56,4	-
Fst_95	54940038	10	45,3	-
Fst_96	101049847	10	75,8	-
Fst_97	118507744	10	87,8	-

---

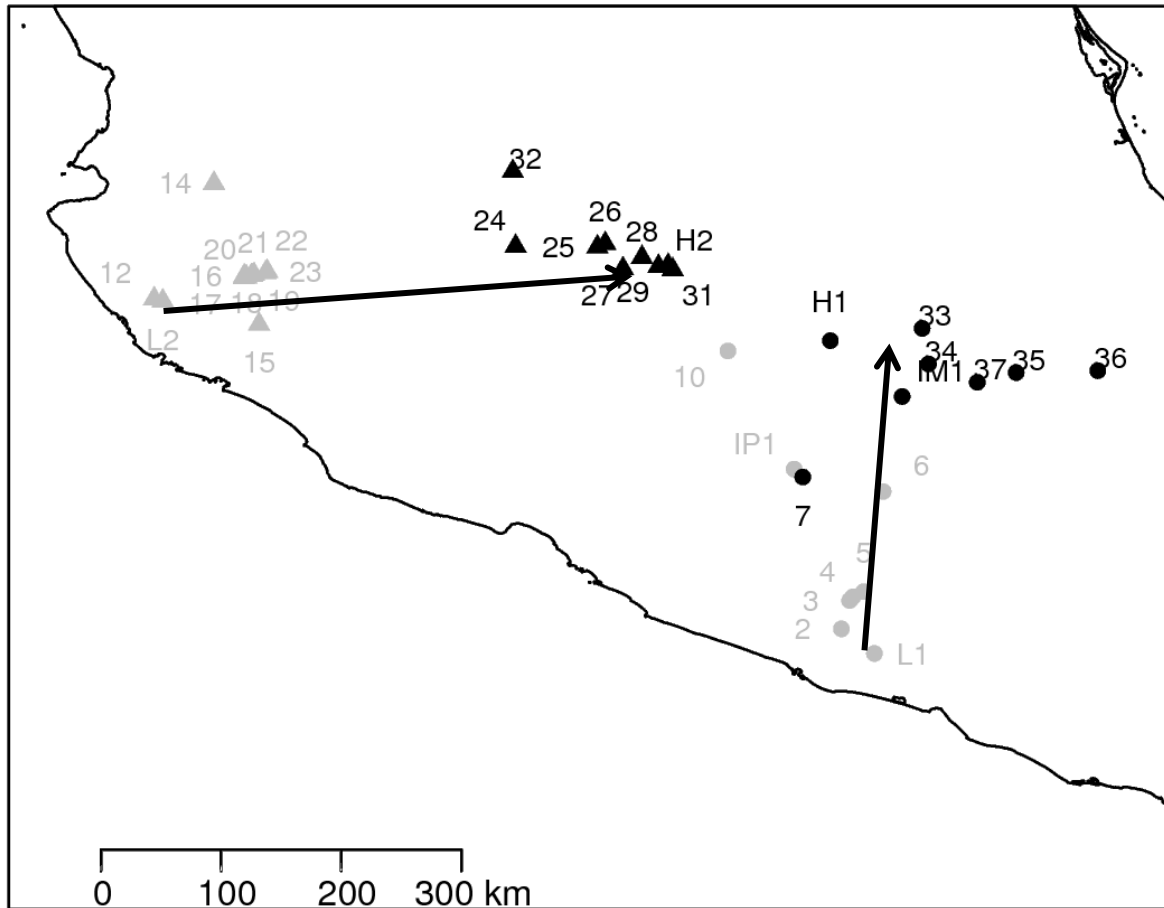
<sup>1</sup> Chromosome

<sup>2</sup> Les SNP provenant de la synthèse bibliographique sont annotés. Ceux provenant des études moléculaires, non présentées dans l'Annexe III.1, sont indiqués avec les références de l'étude d'origine.



**Figure 25 : Positions des SNP candidats le long des 10 chromosomes.** En bleu, les SNP issus des analyses du Chapitre 1, en orange, les SNP issus de la synthèse bibliographique, et vert les SNP issus des scans Fst.

Les 270 SNP ont été génotypés sur 2 037 individus provenant de 37 populations échantillonnées le long des deux gradients (Annexe I.2). Les 31 populations disponibles pour le chapitre 1 et six autres populations fournies par Jonas Aguirre et Luis Eguiarte de l'UNAM sont réparties sur deux gradients altitudinaux, le gradient 1 comportant 16 populations et le gradient 2 comprenant 21 populations (Figure 26). Les populations sont divisées en 2 catégories : celles utilisées également dans le dispositif de génétique d'association (populations L1, L2, 3, 4, 6, 7, 15, 17, 25, H1 et H2, cf section IV.2.2) pour lesquelles nous disposons d'un grand nombre d'individus provenant d'un nombre restreint de plantes-mères par population (entre 122 et 152 individus provenant de 18 à 21 plantes-mères), et celles exclusivement impliquées dans les études de clines pour lesquelles chaque individu provient d'une plante-mère différente (entre 10 et 21 individus et plantes-mères, à l'exception de la population 11 pour laquelle une plante-mère est répétée deux fois). Sur les 2 037 individus génotypés, 42 ont été enlevés des analyses du fait d'une mauvaise attribution de l'identité de l'individu, cf. section IV.3).



**Figure 26 : Carte du centre du Mexique.** Les 37 populations sont réparties le long de deux gradients. Les 16 populations du gradient 1 sont représentées par des ronds, les 21 populations du gradient 2 sont représentées par des triangles. Les populations *parviglumis* sont représentées en gris, les populations *mexicana* sont représentées en noir. Les deux gradients sont représentés symboliquement par des flèches noires allant de basse vers haute altitude.

Parmi les 270 SNP candidats, le génotypage KASPar a fonctionné (au moins partiellement) sur 218 SNP et a échoué totalement sur 52 SNP (Annexe III.2). Lorsque le SNP a été génotypé, le résultat fourni par LGC est présenté : (1) soit sous la forme du génotype de l'individu en question, (2) soit, lorsque le génotypage n'est pas de la meilleure qualité, comme « bad » ou « ? » (un retour au signal lumineux des données permet dans certains cas de déterminer le génotype manuellement), (3) soit avec l'indication « uncallable » lorsque le signal lumineux ne permet aucune identification de génotype.

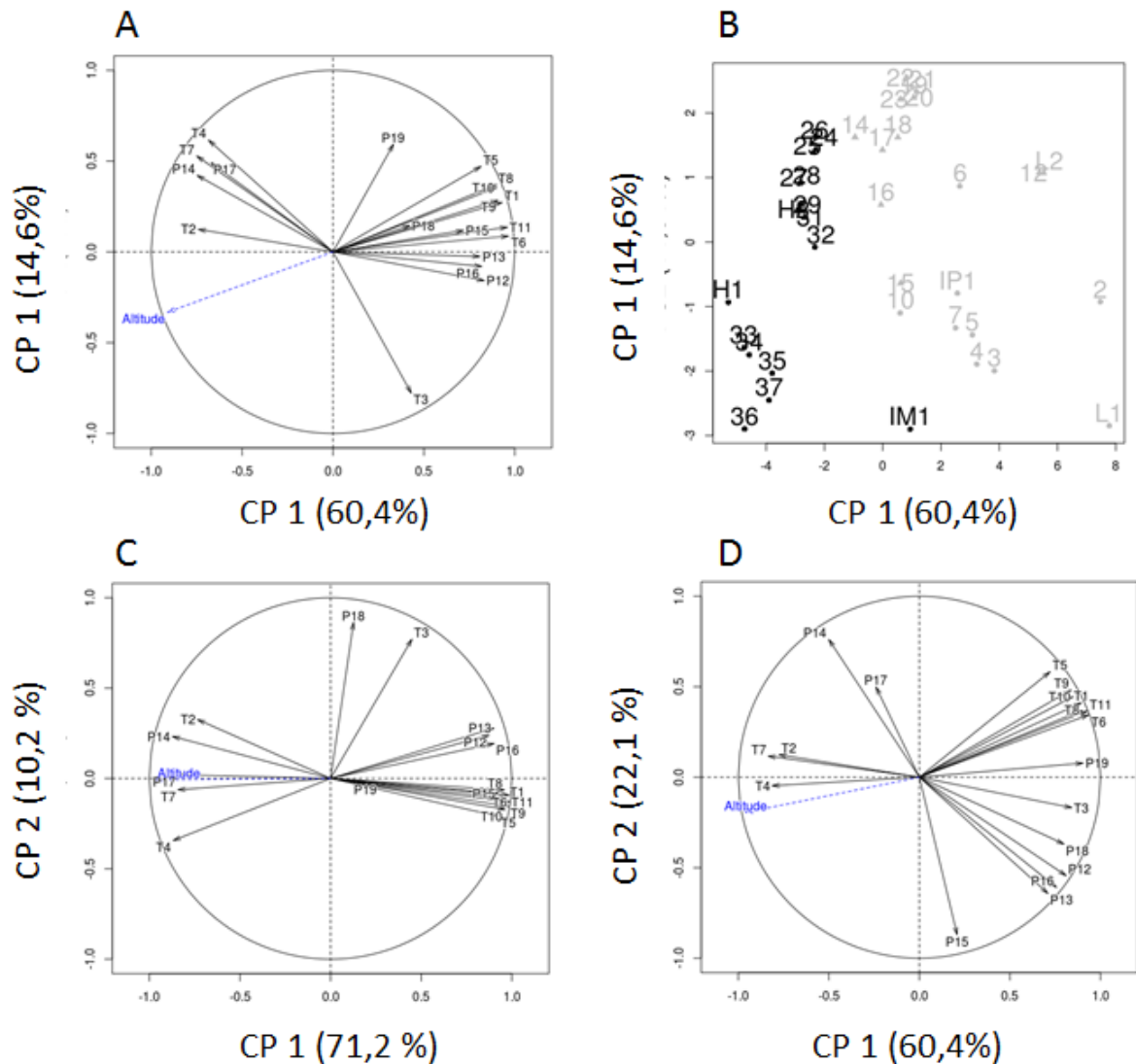
Dans un premier temps, les fréquences alléliques ont été calculées par population (Annexe III.2). Comme nous ne disposons pas d'un groupe externe, ces SNP n'ont pas été orientés et les fréquences alléliques ont été calculées sur l'un des deux allèles pris au hasard dans la population. Ces fréquences alléliques ont été utilisées dans le chapitre 1 (cf section II.5) afin de valider notre approche d'estimation des fréquences alléliques des populations séquencées en mélange (H1, L1, H2, L2).



### **III.3 CLINES DE FREQUENCES**

Dans cette partie nous cherchons à établir des clines de fréquences pour valider les 270 SNP candidats. Des clines peuvent être modélisés selon leur largeur, la localisation du point d'inflexion, et la forme de la queue de distribution de part et d'autre du cline. Pour faire ceci, nous avons utilisé HZAR, un package de R permettant d'estimer et de représenter graphiquement des clines pour des traits génétiques ou phénotypiques (Derryberry *et al.* 2014). Le package HZAR permet de tester des modèles de complexité variée et laisse une grande liberté de paramétrage des différents composants des clines. HZAR a été développé afin de caractériser le changement des fréquences alléliques dans un cadre géographique. L'environnement HZAR nécessite un échantillonnage le long d'un transect géographique en supposant le minimum de variation perpendiculairement au gradient (Derryberry *et al.* 2014). Les deux gradients ont donc été testés séparément, et ont été considérés comme des répliquats biologiques. Les distances géographiques entre populations ont été calculées relativement à la population la plus extrême du gradient (Populations L1 et L2 pour les gradients 1 et 2, respectivement) pour et en faisant l'hypothèse qu'elles sont alignées sur une droite grâce à la fonction *hzar.map.greatCircleDistance* (Derryberry *et al.* 2014).

Dans cette étude nous cherchons à mettre en évidence des polymorphismes dont les fréquences alléliques varient en fonction de l'altitude et/ou de variables environnementales. Les clines ont donc été testés de deux manières différentes (Derryberry *et al.* 2014). D'une part, l'altitude a été directement utilisée pour confronter les fréquences alléliques. D'autre part, une variable environnementale synthétique a été construite à partir d'une ACP des 37 populations sur les 19 variables environnementales. Le premier axe de l'ACP (60,04% de variance, Figure 27) a été retenu et les distances entre populations ont été calculées à partir des coordonnées sur l'axe 1. Que ce soit en utilisant l'altitude ou la variable environnementale, les gradients ont été testés soit ensemble, soit séparément. Lorsque les gradients ont été analysés séparément, la variable environnementale a été définie pour chaque gradient à partir du premier axe d'une ACP portant sur les 16 populations du gradient 1 (Axe 1 : 59,42%) d'une part et les 21 populations du gradient 2 (Axe 1 : 61,75) d'autre part (Figure 27).



**Figure 27 : Analyses en Composantes Principales basées sur 19 variables climatiques et les 37 populations échantillonnées sur les deux gradients (A et B), sur les 16 populations de téosintes échantillonnées sur le gradient 1 (C) et sur 21 populations de téosintes échantillonnées sur le gradient 2 (D).** A, C et D : Cercles des corrélations des variables sur le premier plan de l'ACP. T1 :Température annuelle moyenne, T2 : Gamme de Température journalière moyenne (Moyenne des mois (max temp-min temp), T3 : Isothermalité (T2/T7)\*100, T4 : Saisonnalité de la température (écart-type\*100), T5 : Température maximale du mois le plus chaud, T6 : Température minimale du mois le plus froid, T7 : Gamme de variation annuelle des température (T5-T6), T8 : Température moyenne du trimestre le plus humide, T9 : Température moyenne du trimestre le plus sec, T10 : Température moyenne du trimestre le plus chaud, T11 : Température moyenne du trimestre le plus froid, P12 : Précipitations annuelles, P13 : Précipitation du mois le plus humide, P14 : Précipitation du mois le plus sec, P15 : Saisonnalité de la précipitation (coefficient de variation), P16 : Précipitation du trimestre le plus humide, P17 : Précipitation du trimestre le plus sec, P18 : Précipitation du trimestre le plus chaud, P19 : Précipitation du trimestre le plus froid. B : Projection des populations sur le premier plan de l'ACP commune. La variable Altitude a été ajoutée à posteriori. B : Projection des populations sur le premier plan de l'ACP commune. Les populations *mexicana* sont représentées en noir, les populations *parviglumis* en gris.

Deux modèles ont été testés et croisés avec les 3 méthodologies mentionnées ci-dessus: le modèle nul qui ne suppose pas de variation particulière des fréquences alléliques entre populations, et un modèle de cline qui laisse une liberté des fréquences alléliques totale entre 0 et 1 et ne comporte pas de queue de distribution exponentielle (*scaling= 'fixed'*,

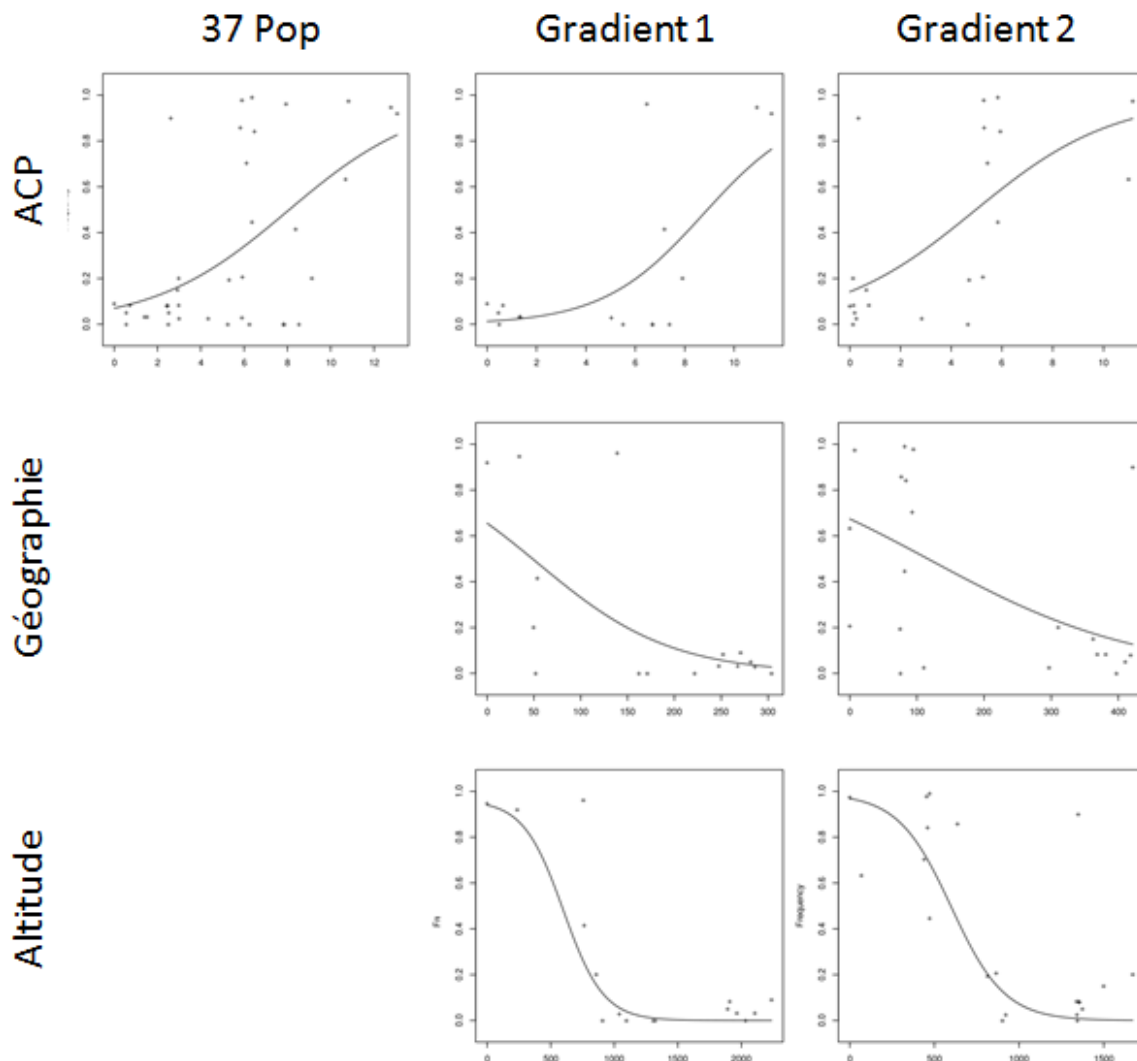
*tails= 'none'*). Le critère d'AIC a permis de sélectionner, pour chaque analyse, le modèle le plus vraisemblable.

Dans le cadre de clines par rapport aux distances géographiques, le modèle de cline est supporté par 102 SNP pour le gradient 1 et 45 SNP pour le gradient 2 (Tableau 3), dont 26 SNP communs aux deux gradients. Lorsque les clines sont modélisés en fonction de la variable environnementale synthétique, le modèle de cline est soutenu par 167 SNP lorsque toutes les populations sont analysées conjointement, par 166 SNP pour le gradient 1, et 153 SNP pour le gradient 2, avec 135 SNP communs aux deux gradients (Tableau 3). Au contraire, lorsque l'altitude est utilisée pour établir les clines de fréquence, très peu de SNP supportent le modèle de cline que ce soit lorsque les deux gradients sont analysés ensemble (1 SNP) ou séparément (8 SNP communs pour les gradients 1 et 2) (Tableau 3). L'altitude et le premier axe de l'ACP sont corrélés négativement, les clines sont donc inversés entre ces deux variables (Figures 28 et 29).

**Tableau 3 : Nombre de SNP présentant un cline dans les trois contextes testés**

Populations analysées	Distance géographique	ACP	Altitude
37 Populations	Non analysé	167	1
Gradient 1	102	166	8
Gradient 2	45	153	8
Gradient 1 et 2	26	135	8

Au total, 21 SNP présentent des clines significatifs avec l'ensemble des conditions géographiques et environnementales (Tableau 4). Parmi ceux-ci, les SNP SB\_65 et Fst\_30 sont également significatifs avec l'altitude pour les gradients 1 et 2 (Figure 28, Tableau 4).

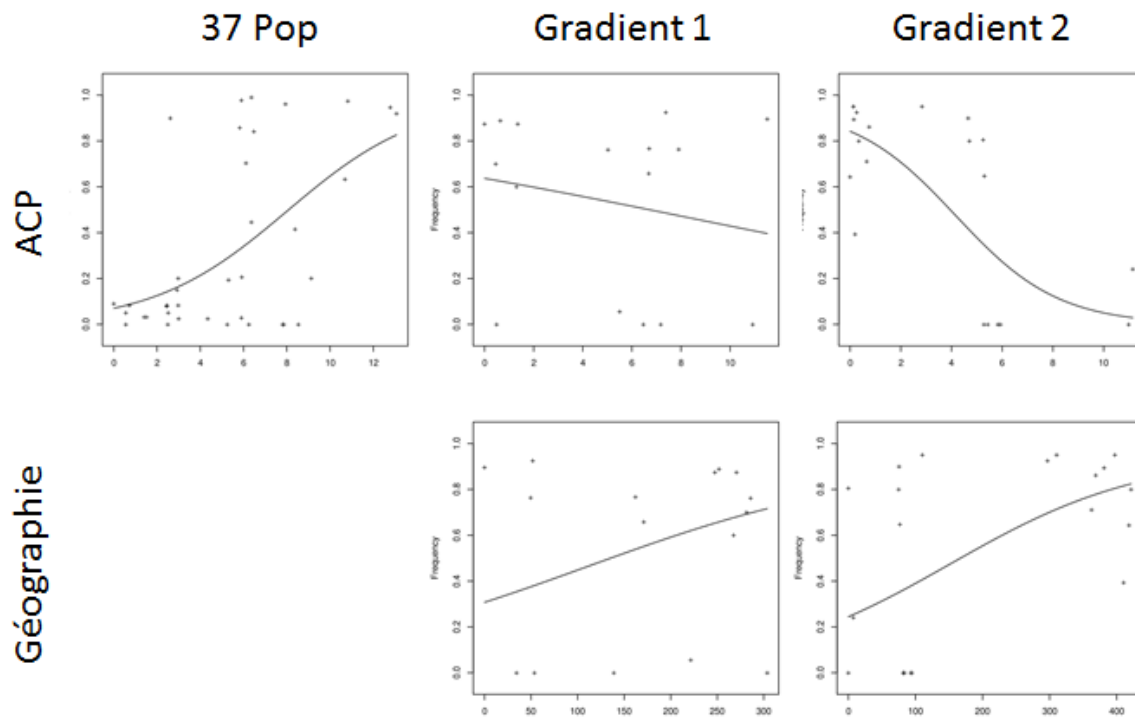


**Figure 28 : Clines de fréquence du SNP SB\_65 par rapport aux conditions environnementales (Haut), aux distances géographiques (Milieu) et à l'altitude (Bas).** A gauche, le cline pour l'ensemble des populations ; au centre, le cline pour les populations du gradient 1 ; à droite, le cline pour les populations du gradient 2. Chaque point représente une population.

Les SNP candidats ont été définis selon trois méthodologies différentes. De forts signaux (clines selon toutes les approches ou/et top 5% AIC, Tableau 4) sont observés pour les SNP des trois origines (Fst-scan : Fst\_9, Fst\_30, Fst\_49, Fst\_87, Fst\_91 ; concordance des signaux de méthodes de détection de la sélection : MS\_15, MS\_61 ; Synthèse bibliographique SB\_1, SB\_2, SB\_34, SB\_39, SB\_50, SB\_52, SB\_65, SB\_66, SB\_69, SB\_71, SB\_75).

Parmi les SNP provenant de la synthèse bibliographique, six SNP (SB\_1, SB\_50, SB\_52, SB\_65, SB\_66, SB\_75) montrant un fort signal clinal sont impliqués dans la sensibilité à la photopériode et/ou la date de floraison (Tableau 4, Figure 28). Ceci est cohérent avec le caractère adaptatif de la date de floraison le long des gradients altitudinaux de téosinte. La précocité de la floraison est un caractère qui a été très étudié chez le maïs (Camus-Kulandaivelu *et al.* 2006 ; Hung *et al.* 2012). Il est intéressant de noter que parmi les

20 SNP provenant de gènes de domestication ou d'amélioration du maïs, tous sauf le SNP 14 présentent un cline (l'exemple du SB\_65 est donné dans la Figure 28). Parmi les 10 SNP originaires de la synthèse bibliographique et ressortant le plus fortement dans les analyses de clines, trois (SB\_34, SB\_69 et SB\_71, Figure 29) proviennent d'analyses de SNP de la puce 50K détectés avec BayeScan (analyses effectuées dans le cadre du doctorat de Jonas Aguirre).



**Figure 29 : Clines de fréquence du SNP SB\_69 par rapport aux conditions environnementales (Haut) et aux distances géographiques (Bas).** A gauche, le cline pour l'ensemble des populations, au centre, le cline pour les populations du gradient 1, à droite, le cline pour les populations du gradient 2. Chaque point représente une population.

**Tableau 4 : Valeur d’AIC entre le modèle avec cline et le modèle nul. Seules les AIC en faveur du modèle avec cline sont montrés. Les valeurs en gras correspondent au top 5% AIC dans les modèles avec les distances géographiques et les conditions environnementales.**

SNPID	Geo_G1	Geo_G2	ACP_FP	ACP_G1	ACP_G2	Alt_FP	Alt_G1	Alt_G2
Fst_3	-26,01	-	-100,17	<b>-218,51</b>	-	-	-	-
Fst_4	-20,28	-	-38,72	-39,28	-1,28	-	-	-
Fst_5	-	-	-10,96	-5,05	-15,21	-	-	-
Fst_7	-	-	-27,26	-29,12	-	-	-	-
Fst_8	-	-	-20,68	-30,03	-	-	-	-
Fst_9	-	<b>-116,86</b>	<b>-227,16</b>	-77,30	<b>-154,71</b>	-	-	-
Fst_12	-	-	-23,81	-25,53	-5,17	-	-	-
Fst_13	-	-	-32,48	-15,71	-22,11	-	-	-
Fst_14	-	-	-	-	-4,61	-	-	-
Fst_16	-54,92	-	-65,21	-41,56	-15,33	-	-	-
Fst_18	-4,06	-	-7,72	-	-37,36	-	-	-
Fst_19	-	-	-23,41	-45,99	-	-	-	-
Fst_20	-6,50	-12,62	-	-13,41	-30,43	-	-	-
Fst_21	-54,50	-	-80,58	-100,83	-4,61	-	-	-
Fst_22	-	-	-43,87	-13,44	-26,05	-	-	-
Fst_23	-32,71	-	-29,24	-29,06	-	-	-	-
Fst_24	-	-	-8,19	-3,32	-	-	-	-
Fst_25	-5,91	-	-18,70	-8,11	-8,52	-	-	-
Fst_27	-	-	-	-	-	-	-	-
Fst_28	-	-	-66,81	-39,22	-35,86	-	-	-
Fst_29	-	-	-26,34	-14,75	-38,13	-	-	-
Fst_30	<b>-105,66</b>	<b>-72,40</b>	<b>-174,88</b>	<b>-221,02</b>	-57,73	-	-45,49	-45,49
Fst_31	-11,26	-	-	-7,37	-8,49	-	-	-
Fst_32	-47,61	-	-17,87	-15,81	-	-	-	-
Fst_34	-	-	-68,78	-58,16	-8,75	-	-	-
Fst_35	-26,99	-6,43	-112,27	-54,34	-68,71	-	-	-
Fst_36	-	-	-11,09	-18,28	-	-	-	-
Fst_37	-3,95	-	-1,12	-3,91	-	-	-	-
Fst_38	-	-	-62,41	-63,69	-9,15	-	-	-
Fst_40	-	-	-55,99	-50,04	-15,30	-	-	-
Fst_41	-	-	-29,44	-30,61	-4,99	-	-	-
Fst_45	-10,30	-	-2,30	-13,52	-	-	-	-
Fst_46	-	-	-22,59	-24,89	-	-	-	-
Fst_47	-	-	-40,15	-39,39	-3,16	-	-	-
Fst_49	<b>-125,59</b>	-	<b>-266,66</b>	<b>-265,65</b>	-45,57	-	-	NA
Fst_50	-12,92	-	-62,87	-41,34	-8,79	-	-	-
Fst_51	-69,94	-	-13,49	-61,88	-0,91	-	-	-
Fst_53	-	-	-0,19	-3,12	-	-	-	-
Fst_54	-15,90	-	-21,94	-43,84	-	-	-	-
Fst_55	-44,49	-	-92,93	-108,21	-5,50	-	-	-
Fst_56	-	-	-17,81	-28,86	-	-	-	-
Fst_57	-	-	-45,85	-5,58	-35,21	-	-	-
Fst_58	-	-10,00	-19,54	-	-76,47	-	-	-

Fst_59	-	-	-10,38	-1,60	-20,83	-	-	-
Fst_60	-8,96	-	-9,30	-10,13	-1,29	-	-	-
Fst_61	-80,33	-	-34,89	-142,65	-12,47	-	-	-
Fst_62	-	-	-46,06	-21,60	-8,88	-	-	-
Fst_66	-4,80	-	-	-	-	-	NA	NA
Fst_68	-	-	-0,73	-7,66	-	-	-	-
Fst_69	-34,33	-	-4,54	-36,11	-	-	-	-
Fst_70	-	-	-12,32	-4,88	-25,43	-	-	-
Fst_71	-	-	-	-	-	-	-	-
Fst_73	-	-6,22	-	-	-4,41	-	-	-
Fst_74	-	-	-79,37	-53,99	-15,27	-	-	-
Fst_75	-	-	-57,36	-6,82	-40,70	-	-	-
Fst_76	-	-	-26,55	-24,27	-7,87	-	-	-
Fst_77	-12,76	-54,01	-32,22	-57,57	-62,17	-	-	-
Fst_78	-	-	-	-9,74	-7,83	-	-	-
Fst_80	-	-	-18,33	-8,56	-12,35	-	-	-
Fst_82	-	-	-5,13	-4,73	-0,74	-	-	-
Fst_83	-26,39	-3,49	-71,19	-86,93	-4,44	-	-	-
Fst_84	-1,21	-6,16	-21,47	-28,59	-2,05	-	-	-
Fst_85	-2,48	-47,73	-33,72	-14,26	-46,03	-	-	-
Fst_87	-11,75	<b>-72,82</b>	-26,75	-21,01	-39,81	-	-	-
Fst_88	-	-	-38,84	-60,72	-10,25	-	-	-
Fst_89	-3,86	-72,20	-	-9,97	-37,25	-	-	-
Fst_90	-50,70	-52,88	-156,10	-125,79	-69,95	-	-	-
Fst_91	-17,62	<b>-74,49</b>	-40,48	-20,52	<b>-89,67</b>	-	-	-
Fst_92	-	-	-59,42	-64,02	-30,43	-	-	-
Fst_94	-24,55	-	-47,97	-63,71	-	-	-	-
Fst_95	-27,85	-	-38,46	-29,52	-14,32	-	-	-
Fst_96	-31,78	-	-0,17	-56,15	-22,70	-	-	-
Fst_97	<b>-123,61</b>	-	-134,08	-153,78	-23,76	-	-38,90	-38,90
MS_1	-	-	-78,79	-27,12	-66,79	-	-	-
MS_2	-0,95	-	-2,39	-1,65	-	-	-	-
MS_3	-20,22	-	-9,51	-9,87	-5,56	-	-	-
MS_9	-	-	-	-7,61	-1,53	-	-	-
MS_11	-	-	-2,86	-	-3,29	-	-	-
MS_13	-	<b>-79,72</b>	-7,23	-	-56,86	NA	-	-
MS_14	-	-	-2,34	-	-3,70	-	-	-
MS_15	-7,61	<b>-85,50</b>	-5,63	-40,22	<b>-179,00</b>	-	-	-
MS_16	<b>-133,67</b>	-	-90,30	-82,56	-8,75	-	-	-
MS_17	-	-	-47,57	-45,84	-11,24	-	-	-
MS_18	-20,82	-	-7,04	-61,01	-	-	-23,50	-23,50
MS_19	-0,39	-	-20,24	-32,38	-	-	-	-
MS_20	-	-	-	-11,57	-31,34	-	-	-
MS_21	-23,22	-	-8,11	-61,59	-	-	-30,30	-30,30
MS_22	-21,41	-	-18,87	-36,07	-2,48	-	-	-
MS_23	-	-	-1,29	-6,93	-	-	-	-

MS_24	-19,45	-	-5,97	-63,50	-	-	-35,96	-35,96
MS_25	-	-	-	-	-	-	-	-
MS_26	-	-2,25	-	-	-0,92	-	-	-
MS_27	-34,02	-47,20	-119,31	-38,31	<b>-115,79</b>	-	-	-
MS_28	-	-	-	-	-	-	-	-
MS_29	-	-	-	-	-	-	-	-
MS_30	-44,57	-	-44,49	-140,79	-7,31	-	-	-
MS_31	-18,82	-	-55,26	-51,51	-9,42	-	-	-
MS_32	-66,87	-	-27,56	-130,07	-21,88	-	-	-
MS_34	-69,11	-	-41,41	-144,90	-13,74	-	-	-
MS_35	-	-	-	-9,30	-38,76	-	-	-
MS_36	-	-5,81	-10,95	-0,70	-71,03	-	-	-
MS_37	-14,28	-	-59,51	-8,59	-62,36	-	-	-
MS_39	-	-11,49	-	-5,29	-14,80	-	-	-
MS_40	-	-	-1,92	-	-7,13	-	-	-
MS_44	-	-7,32	-	-0,04	-7,69	-	-	-
MS_45	-	-29,87	-10,88	-	-53,93	-	-	-
MS_47	-	-	-6,23	-	-7,90	-	-	-
MS_48	-7,32	-	-20,22	-7,38	-17,69	-	-	-
MS_49	-52,32	-	-144,02	<b>-163,26</b>	-10,16	-	-	-
MS_50	-	-	-63,21	-58,49	-17,71	-	-	-
MS_51	-	-	-21,90	-13,65	-17,72	-	-	-
MS_55	-	-	-68,77	-41,32	-21,01	-	-	-
MS_56	-	-	-3,36	-5,20	-2,55	-	-	-
MS_57	-	-	-86,43	-80,50	-12,41	-	-	-
MS_58	-39,07	-	-142,55	-137,66	-14,67	-	-	-
MS_59	<b>-105,63</b>	-	<b>-193,07</b>	<b>-173,26</b>	-67,97	-	-	-
MS_60	-	-	-73,80	-55,11	-23,34	-	-	-
MS_61	-91,61	-40,94	<b>-203,05</b>	-133,40	<b>-114,13</b>	-	-	-
MS_62	-18,16	-	-	-	-	-	-	-
MS_63	-18,70	-62,82	-148,76	-63,53	-88,26	-	-	-
MS_64	-42,41	-	-125,47	-74,84	-67,62	-	-	-
MS_66	-23,91	-	-111,53	-100,59	-17,81	-	-	-
MS_67	-6,11	-	-90,29	-49,43	-54,19	-	-	-
MS_68	-	-	-2,60	-12,84	-	-	-	-
MS_69	-	-	-2,21	-0,17	-	-	-	-
MS_70	-61,95	-10,33	-59,09	-45,08	-22,61	-	-	-
MS_74	-84,86	-	<b>-179,41</b>	-124,15	-77,00	-	-	-
MS_75	-	-58,58	-23,76	-	-103,08	-	-	-
MS_76	-51,42	-	-147,95	-160,31	-10,27	-	-	-
MS_77	-	-59,81	-23,15	-	-96,46	-	-	-
MS_78	-51,01	-	-145,30	-123,26	-28,29	-	-	-
MS_81	-	-	-13,15	-	-48,43	-	-	-
MS_82	-40,64	-	-65,61	-42,63	-2,21	-	-	-
MS_83	-19,16	-	-21,91	-3,29	-14,11	-	-	-
MS_87	-	-	-2,37	-12,93	-76,14	-	-	-



MS_88	-	-	-	-	-	-	-	-
MS_89	-63,36	-2,07	-4,35	-97,10	-48,60	-	-	-
SB_1	-	<b>-144,75</b>	-1,63	-23,27	<b>-137,73</b>	-	-	-
SB_2	-18,18	-	-83,63	-34,05	-39,00	-2,50	-40,38	-40,38
SB_3	-	-	-23,77	-25,90	-	-	-	-
SB_4	-13,98	-	-0,08	-19,58	-8,21	-	-	-
SB_7	-10,38	-	-33,92	-61,69	-	-	-	-
SB_8	-	-	-0,88	-21,79	-10,94	-	-	-
SB_9	-19,73	-	-17,36	-14,88	-0,18	-	-	-
SB_10	-	-	-0,38	-	-	-	-	-
SB_12	-	-11,13	-10,44	-7,58	-1,69	-	-	-
SB_13	-	-	-	-	-	-	-	-
SB_14	-	-	-	-	-	-	-	-
SB_15	-	-	-53,81	-55,32	-4,22	-	-	-
SB_16	-6,01	-	-21,85	-24,43	-1,81	-	-	-
SB_20	-	-4,66	-	-	-3,36	-	-	-
SB_22	-23,83	-	-4,93	-20,16	-61,71	-	-	-
SB_25	-	-	-33,61	-7,61	-38,48	-	-	-
SB_26	-	-	-27,49	-40,25	-	-	-	-
SB_27	-	-52,62	-115,70	-41,11	-100,41	-	-	-
SB_28	-	-	-1,75	-	-	-	-	-
SB_29	-61,33	-	-33,94	-48,25	-	-	-	-
SB_30	-18,66	-	-91,78	-47,85	-62,98	-	-	-
SB_32	-	-	-51,28	-20,47	-52,88	-	-	-
SB_33	-	-7,55	-	-	-6,12	-	-	-
SB_34	-47,63	<b>-85,36</b>	-152,65	<b>-181,28</b>	-25,78	-	-	-
SB_35	-40,29	-	-146,85	-87,48	-50,17	-	-	-
SB_38	-30,82	-	-77,43	-44,07	-31,47	-	-	-
SB_39	<b>-224,66</b>	-1,78	<b>-374,99</b>	<b>-416,61</b>	-27,08	-	-	-
SB_41	-	-	-12,20	-30,90	-11,36	-	-	-
SB_43	-	-	-34,53	-38,72	-19,59	-	-	-
SB_44	-	-	-	-	-5,26	-	-	-
SB_45	-46,42	-	-41,57	-67,54	-32,86	-	-	-
SB_46	-88,72	-	-83,14	-126,09	-1,86	-	-	-
SB_47	-39,00	-	-92,53	-16,84	-77,02	-	-	-
SB_48	-	-	-	-	-1,31	-	-	-
SB_49	-31,10	-	-	-26,38	-28,67	-	-	-
SB_50	-83,85	-	<b>-203,14</b>	<b>-268,73</b>	-4,51	-	-	-
SB_51	-	-	-0,86	-0,71	-	-	-	-
SB_52	-34,26	-62,30	-140,64	-21,20	<b>-152,06</b>	-	-	-
SB_54	-	-	-0,23	-1,65	-	-	-	-
SB_55	-40,31	-	-51,56	-59,04	-10,85	-	-	-
SB_56	-6,07	-	-87,90	-63,70	-26,50	-	-	-
SB_57	-	-11,48	-63,71	-5,26	-97,06	-	-	-
SB_58	<b>-117,37</b>	-	-139,62	-124,86	-27,95	-	-	-
SB_59	-39,05	-	-81,88	-105,87	-	-	-	-

SB_60	-36,99	-	-47,42	-6,40	-49,28	-	-	-
SB_61	-	-0,48	-	-	-0,48	-	-	-
SB_63	<b>-247,29</b>	-	-135,46	-209,82	-18,45	-	-	-
SB_64	-0,38	-	-8,25	-54,98	-19,41	-	-	-
SB_65	<b>-134,67</b>	<b>-100,28</b>	<b>-243,15</b>	-147,99	<b>-190,78</b>	-	-160,23	-160,23
SB_66	<b>-292,21</b>	-	<b>-277,24</b>	<b>-269,65</b>	-32,04	-	-	-
SB_69	-	<b>-146,72</b>	-128,74	-11,48	<b>-255,83</b>	-	-	-
SB_70	-27,74	-66,46	-7,32	-59,73	-34,66	-	-	-
SB_71	-18,64	-9,95	-50,42	-58,98	-7,62	-	-	-
SB_72	-1,88	-	-	-27,07	-43,38	-	-	-
SB_73	-6,30	-39,24	-	-10,26	-26,22	-	-	-
SB_74	-51,96	-	-53,09	-34,34	-50,46	-	-	-
SB_75	-8,40	-42,46	-21,37	-17,23	-42,35	-	-	-
SB_76	-1,29	-15,30	-	-34,28	-59,37	-	-	-
SB_77	-64,63	-	-67,11	-67,02	-	-	-11,42	-11,42
SB_79	-	-	-84,13	-16,35	-76,71	-	-	-
SB_80	-	-	-3,19	-	-3,44	-	-	-

### **III.4 DISCUSSION**

Ce chapitre est consacré au choix et à la validation moléculaire de SNP candidats à l'adaptation à l'altitude. 270 SNP provenant aussi bien d'approches « Top-down » que d'approches « Bottom-up » ont été retenus. Le génotypage de certains SNP reste un défi chez la téosinte au vu du fort niveau de polymorphisme. Malgré un effort substantiel pour choisir des SNP compatibles avec le génotypage par KASPar, le taux d'échec total ou de SNP monomorphes reste relativement élevé (presque 20%). Ce taux reste cependant relativement faible comparé à d'autres études (près de 44% d'échec total ou de données monomorphes chez le mil, Yves Vigouroux, communication personnelle). Pour certaines combinaisons individus/SNP, l'interprétation du signal lumineux par les logiciels de détection n'a pas abouti à un génotypage. La lecture manuelle permet dans certains cas d'identifier le génotype de SNP supplémentaires et d'augmenter ainsi la taille du jeu de données. A ce stade de la rédaction, ces relectures chronophages n'ont pas encore été effectuées mais seront réalisées prochainement. Elles permettront d'affiner les fréquences alléliques dans certains cas, mais surtout d'obtenir le génotypage des SNP pour lesquels le signal nécessite un calibrage manuel pour l'ensemble des données le concernant.

Les scans de *Fst* le long du génome permettent de visualiser des régions présentant un fort déséquilibre de liaison. Trois de ces régions ont été reportées comme des inversions dans la littérature (Fang *et al.* 2012 ; Pyhäjärvi *et al.* 2013). Une autre inversion localisée sur le

chromosome 9 et détectée par (Pyhäjärvi *et al.* 2013) n'est pas visualisée ici : elle ne présente pas de patron de différenciation entre sous-espèces ni de signal d'implication dans l'adaptation locale. Nous obtenons un signal en faveur de l'implication de *Inv1n* dans la différenciation sub-spécifique alors que Fang *et al.* (2012) et Pyhäjärvi *et al.* (2013) soutiennent le caractère adaptatif de cette inversion. L'étude de Pyhäjärvi *et al.* (2013), basée sur 21 populations de téosintes, montre que cette inversion a des fréquences très différentes selon les populations de *parviglumis*. L'absence de signal adaptatif obtenu dans notre étude est probablement dû à notre échantillonnage et met en avant une limite des analyses basées sur seulement six populations. Les neuf SNP géotypés avec succès dans l'inversion *Inv9e* présentent des clines de fréquences pour les deux gradients 1 et 2 en fonction des variables environnementales. Ce résultat renforce l'hypothèse du rôle adaptatif des inversions. En effet, la présence d'une inversion conduit à la suppression de la recombinaison chez les individus hétérozygotes, conduisant à la protection d'allèles adaptés localement (Kirkpatrick & Barton 2006).

Plusieurs études ont montré des clines de fréquences alléliques (Hoffmann & Weeks 2007 ; Labbé *et al.* 2009 ; Mullen & Hoekstra 2008 ; Mullen *et al.* 2009). Deux phénomènes peuvent être à l'origine de tels clines: une pression sélective le long d'un gradient, ou un contact secondaire après spéciation allopatrique. Dans certain cas, notamment lorsqu'une pression de sélection coïncide avec la distribution des taxons, il est difficile de distinguer ces deux scénarios. Le cas de *Peromyscus polionotus*, la souris des sables, a été très étudié. Il existe deux principaux morphes de coloration, un clair et un foncé, dont la fréquence varie le long d'un gradient de coloration du sol. L'analyse conjointe de données morphologiques et environnementales, ainsi que de données microsatellites considérées comme neutres et de deux gènes candidats pour la coloration de la robe, *Mclr* et *Agouti*, a permis de trancher en faveur d'une pression sélective, probablement liée à la prédation, à l'origine de cette différenciation (Mullen & Hoekstra 2008 ; Mullen *et al.* 2009). Les téosintes *parviglumis* et *mexicana* sont également distribuées dans des conditions environnementales différentes. Dans ce chapitre, les clines de fréquence ont été recherchés par rapport à différentes variables : les distances géographiques entre populations, une variable environnementale synthétique et l'altitude. La variable environnementale synthétique est celle qui permet d'établir un plus grand nombre de clines de fréquences et montre des valeurs d'AIC plus élevées. Certains de ces SNP montrent aussi un cline avec la variable de distance géographique mais aucun SNP ne présente un cline exclusivement dans un contexte géographique. Il est toutefois surprenant que très peu de SNP présentent des clines en relation avec l'altitude, alors que cette variable

est très corrélée avec le premier axe d'ACP sur les variables environnementales, ce qui met en question la pertinence de prendre l'altitude comme proxy de l'environnement.

Il y a un plus grand nombre de clines dans le gradient 1 que dans le gradient 2 (Tableau 3). Il est probable que l'échantillonnage plus régulier dans le gradient 1 facilite la détection de clines, malgré le fait qu'il y ait moins de populations (16 populations) que dans le gradient 2 (21 populations). D'une part nous avons échantillonné peu de populations d'altitude intermédiaire dans le gradient 2. D'autre part, les populations de basse altitude du gradient 2 présentent une structuration importante qui peut également brouiller la détection de clines, notamment dans le cadre des clines avec les distances géographiques (Moeller *et al.* 2007).

Le modèle utilisé pour ajuster les données aux clines est un modèle simple qui permet une première approche exploratoire des données dans un laps de temps réduit. Des modèles plus complexes peuvent être employés qui modélisent finement la largeur du cline, le point d'inflexion et les longueurs des queues de distribution de part et d'autre du cline. Ce travail est réalisable avec HZAR en utilisant des modèles plus complets et en testant les différents paramètres possibles. Ce genre de modèle nécessite un temps de calcul important et sera employé en priorité pour les SNP montrant des clines dans ce chapitre. L'analyse de la distribution des paramètres de ces nouveaux clines, en particulier le point d'inflexion et la largeur du cline, permettra d'affiner la vision de la force sélective agissant sur les populations. La position du point d'inflexion donne des indications sur la localisation géographique de la pression de sélection, tandis que la largeur du cline indique l'importance relative de la sélection par rapport à la migration (Gay *et al.* 2008).

Nos résultats montrent des signaux de cline pour la plupart des SNPs provenant de locus de domestication ou d'amélioration du maïs. En particulier, les SNP Fst\_97, MS\_59, SB\_2, SB\_39, SB\_65 et SB\_66 présentent un signal clinal important et appartiennent à des gènes différenciés entre maïs de haute et de basse altitude (Takuno *et al.* 2015). Un certain nombre de SNP appartenant à des locus impliqués dans la sensibilité à la photopériode, caractère clé de l'adaptation des maïs aux zones tempérées, a également été détecté. Ceci est curieux car il n'y a pas réellement de variation de photopériode dans nos populations qui proviennent de latitudes similaires. Il est possible que ceci reflète la sélection sur une variable très liée à la photopériode, comme par exemple la précocité de floraison. La sélection de certains génotypes chez le maïs au cours de la domestication puis de l'amélioration a pu porter sur des gènes également sous sélection chez la téosinte dans des conditions naturelles. Puisque la domestication et l'amélioration variétale du maïs se sont accompagnées de goulots

d'étranglement (Yamasaki *et al.* 2005), des cibles de sélection existantes chez la téosinte ont pu être perdues chez le maïs. La découverte de nouveaux polymorphismes adaptatifs chez la téosinte pourrait aboutir à l'identification d'allèles d'intérêt pour les sélectionneurs. En effet, l'introgession d'un génotype « téosinte » vers le maïs pourrait contribuer à l'amélioration de ces derniers à des conditions extrêmes. Ceci a déjà été observé chez les maïs traditionnels de haute altitude qui se sont adaptés après la domestication aux conditions plus froides et sèches que celles de son aire de domestication (Takuno *et al.* 2015 ; van Heerwaarden *et al.* 2011). Cette pratique se retrouve aussi chez d'autres couples « espèces cultivée-sauvages », comme chez la tomate où des allèles sauvages ont contribué à la pigmentation des tomates et l'amélioration de la masse (Tanksley & McCouch 1997).

## **REFERENCES BIBLIOGRAPHIQUES**

- Agrama HA, Moussa ME (1996) Mapping QTLs in breeding for drought tolerance in maize (*Zea mays* L.). *Euphytica* **91**, 89-97.
- Anderson MD, Prasad TK, Martin BA, Stewart CR (1994) Differential gene expression in chilling-acclimated maize seedlings and evidence for the involvement of abscisic acid in chilling tolerance. *Plant Physiology* **105**, 331-339.
- Azevedo GC, Cheavegatti-Gianotto A, Negri BF, *et al.* (2015) Multiple interval QTL mapping and searching for PSTOL1 homologs associated with root morphology, biomass accumulation and phosphorus content in maize seedlings under low-P. *BMC plant biology* **15**, 1.
- Baek D, Kim MC, Chun HJ, *et al.* (2013) Regulation of miR399f transcription by AtMYB2 affects phosphate starvation responses in Arabidopsis. *Plant Physiology* **161**, 362-373.
- Berberich T, Kusano T (1997) Cycloheximide induces a subset of low temperature-inducible genes in maize. *Molecular and General Genetics MGG* **254**, 275-283.
- Bertin P, Gallais A (2001) Genetic variation for nitrogen use efficiency in a set of recombinant inbred lines II-QTL detection and coincidences. *Maydica* **46**, 53-68.
- Bhave MR, Lawrence S, Barton C, Hannah LC (1990) Identification and molecular characterization of Shrunken-2 cDNA of maize. *Plant Cell* **2**, 581-588.
- Boitard S, Schlötterer C, Nolte V, Pandey RV, Futschik A (2012) Detecting selective sweeps from pooled next-generation sequencing samples. *Molecular Biology and Evolution* **29**, 2177-2186.
- Briggs WH, McMullen MD, Gaut BS, Doebley J (2007) Linkage mapping of domestication loci in a large maize-teosinte backcross resource. *Genetics* **177**, 1915-1928.
- Butrón A, Chen Y, Rottinghaus G, McMullen MD (2010) Genetic variation at bx1 controls DIMBOA content in maize. *Theoretical and Applied Genetics* **120**, 721-734.
- Camus-Kulandaivelu L, Veyrieras J-B, Madur D, *et al.* (2006) Maize adaptation to temperate climate: relationship between population structure and polymorphism of *Dwarf8* gene. *Genetics* **172**, 2449-2463.
- Carlson SJ, Chourey PS, Helentjaris T, Datta R (2002) Gene expression studies on developing kernels of maize sucrose synthase (SuSy) mutants show evidence for a third SuSy gene. *Plant molecular biology* **49**, 15-29.
- Chen Z, Wang B, Dong X, *et al.* (2014) An ultra-high density bin-map for rapid QTL mapping for tassel and ear architecture in a large F2 maize population. *BMC genomics* **15**, 433.
- Chuck G, Meeley R, Irish E, Sakai H, Hake S (2007) The maize tasselseed4 microRNA controls sex determination and meristem cell fate by targeting Tasselseed6/indeterminate spikelet1. *Nature genetics* **39**, 1517-1521.
- Chung C-L, Jamann T, Longfellow J, Nelson R (2010) Characterization and fine-mapping of a resistance locus for northern leaf blight in maize bin 8.06. *Theoretical and Applied Genetics* **121**, 205-227.
- Clark RM, Nussbaum Wagler T, Quijada P, Doebley J (2006) A distant upstream enhancer at the maize domestication gene *tb1* has pleiotropic effects on plant and inflorescent architecture. *Nature genetics* **38**, 594-597.
- Cook JP, McMullen MD, Holland JB, *et al.* (2012) Genetic architecture of maize kernel composition in the nested association mapping and inbred association panels. *Plant Physiology* **158**, 824-834.
- Coque M, Gallais A (2006) Genomic regions involved in response to grain yield selection at high and low nitrogen fertilization in maize. *Theoretical and Applied Genetics* **112**, 1205-1220.
- Crémieux L, Bischoff A, Šmilauerová M, *et al.* (2008) Potential contribution of natural enemies to patterns of local adaptation in plants. *New Phytologist* **180**, 524-533.
- 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.
- Delmore K, Breneman R, Lei R, *et al.* (2013) Clinal variation in a brown lemur (*Eulemur* spp.) hybrid zone: combining morphological, genetic and climatic data to examine stability. *Journal of Evolutionary Biology* **26**, 1677-1690.
- Derryberry EP, Derryberry GE, Maley JM, Brumfield RT (2014) HZAR: hybrid zone analysis using an R software package. *Molecular Ecology Resources* **14**, 652-663.
- Doebley J, Stec A (1993) Inheritance of the morphological differences between maize and teosinte: comparison of results for two F2 populations. *Genetics* **134**, 559-570.
- Doebley J, Stec A, Gustus C (1995) Teosinte branched1 and the origin of maize: evidence for epistasis and the evolution of dominance. *Genetics* **141**, 333-346.
- Dordas C, Hasinoff BB, Igamberdiev AU, *et al.* (2003) Expression of a stress-induced hemoglobin affects NO levels produced by alfalfa root cultures under hypoxic stress. *The Plant Journal* **35**, 763-770.
- Fang Z, Pyhäjärvi T, Weber AL, *et al.* (2012) Megabase-scale inversion polymorphism in the wild ancestor of maize. *Genetics* **191**, 883-894.

- Foll M, Gaggiotti O (2008) A Genome-Scan Method to Identify Selected Loci Appropriate for Both Dominant and Codominant Markers: A Bayesian Perspective. *Genetics* **180**, 977-993.
- Frey M, Huber K, Park WJ, *et al.* (2003) A 2-oxoglutarate-dependent dioxygenase is integrated in DIMBOA-biosynthesis. *Phytochemistry* **62**, 371-376.
- Fukunaga K, Hill J, Vigouroux Y, *et al.* (2005) Genetic diversity and population structure of teosinte. *Genetics* **169**, 2241-2254.
- Gallavotti A, Zhao Q, Kyojuka J, *et al.* (2004) The role of barren stalk1 in the architecture of maize. *Nature* **432**, 630-635.
- Gamuyao R, Chin JH, Pariasca-Tanaka J, *et al.* (2012) The protein kinase Pstol1 from traditional rice confers tolerance of phosphorus deficiency. *Nature* **488**, 535-539.
- Gay L, Crochet PA, Bell DA, Lenormand T (2008) Comparing clines on molecular and phenotypic traits in hybrid zones: a window on tension zone models. *Evolution* **62**, 2789-2806.
- Günther T, Coop G (2013) Robust Identification of Local Adaptation from Allele Frequencies. *Genetics* **195**, 205-220.
- Haldane J (1948) The theory of a cline. *Journal of genetics* **48**, 277-284.
- Hirsch CN, Flint-Garcia SA, Beissinger TM, *et al.* (2014) Insights into the effects of long-term artificial selection on seed size in maize. *Genetics* **198**, 409-421.
- Hoffmann AA, Weeks AR (2007) Climatic selection on genes and traits after a 100 year-old invasion: a critical look at the temperate-tropical clines in *Drosophila melanogaster* from eastern Australia. *Genetica* **129**, 133-147.
- Hufford MB, Bilinski P, Pyhäjärvi T, Ross-Ibarra J (2012) Teosinte as a model system for population and ecological genomics. *Trends in Genetics* **28**, 606-615.
- Hung H-Y, Shannon LM, Tian F, *et al.* (2012) ZmCCT and the genetic basis of day-length adaptation underlying the postdomestication spread of maize. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 1913-1921.
- Jiang B, Siregar U, Willeford KO, Luthe DS, Williams WP (1995) Association of a 33-kilodalton cysteine proteinase found in corn callus with the inhibition of fall armyworm larval growth. *Plant Physiology* **108**, 1631-1640.
- Kawecki TJ, Ebert D (2004) Conceptual issues in local adaptation. *Ecology Letters* **7**, 1225-1241.
- Kirkpatrick M, Barton N (2006) Chromosome inversions, local adaptation and speciation. *Genetics* **173**, 419-434.
- Kusano T, Berberich T, Harada M, Suzuki N, Sugawara K (1995) A maize DNA-binding factor with a bZIP motif is induced by low temperature. *Molecular and General Genetics MGG* **248**, 507-517.
- Kyetere D, Ming R, McMullen M, *et al.* (1999) Genetic analysis of tolerance to maize streak virus in maize. *Genome* **42**, 20-26.
- Labbé P, Sidos N, Raymond M, Lenormand T (2009) Resistance gene replacement in the mosquito *Culex pipiens*: fitness estimation from long-term cline series. *Genetics* **182**, 303-312.
- Lauter N, Gustus C, Westerbergh A, Doebley J (2004) The inheritance and evolution of leaf pigmentation and pubescence in teosinte. *Genetics* **167**, 1949-1959.
- Leng P, Ji Q, Tao Y, *et al.* (2015) Characterization of sugarcane mosaic virus Scmv1 and Scmv2 resistance regions by regional association analysis in maize. *PLoS ONE* **10**, e0140617.
- Lipka AE, Gore MA, Magallanes-Lundback M, *et al.* (2013) Genome-wide association study and pathway-level analysis of tocochromanol levels in maize grain. *G3: Genes| Genomes| Genetics* **3**, 1287-1299.
- Liu Y, Zhou M, Gao Z, *et al.* (2015) RNA-Seq Analysis Reveals MAPKKK Family Members Related to Drought Tolerance in Maize. *PLoS ONE* **10**, e0143128.
- Malosetti M, Ribaut JM, Vargas M, Crossa J, Van Eeuwijk FA (2008) A multi-trait multi-environment QTL mixed model with an application to drought and nitrogen stress trials in maize (*Zea mays* L.). *Euphytica* **161**, 241-257.
- Manicacci D, Camus-Kulandaivelu L, Fourmann M, *et al.* (2009) Epistatic interactions between Opaque2 transcriptional activator and its target gene CyPPDK1 control kernel trait variation in maize. *Plant Physiology* **150**, 506-520.
- Maron LG, Guimarães CT, Kirst M, *et al.* (2013) Aluminum tolerance in maize is associated with higher MATE1 gene copy number. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 5241-5246.
- Matsuoka Y, Vigouroux Y, Goodman MM, *et al.* (2002) A single domestication for maize shown by multilocus microsatellite genotyping. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 6080-6084.
- McMullen MD, Simcox KD (1995) Genomic organization of disease and insect resistance genes in maize. *Molecular Plant-Microbe Interactions* **8**, 811-815
- McMullen M, Louie R (1989) The linkage of molecular markers to a gene controlling the symptom response in maize to maize dwarf mosaic virus. *Molecular Plant-Microbe Interactions* **2**, 309-314.

- McMullen MD, Jones MW, Simcox KD, Louie R (1994) Three genetic loci control resistance to wheat streak mosaic virus in the maize inbred Pa405. *MPMI-Molecular Plant Microbe Interactions* **7**, 708-712.
- Meihls LN, Handrick V, Glauser G, *et al.* (2013) Natural variation in maize aphid resistance is associated with 2, 4-dihydroxy-7-methoxy-1, 4-benzoxazin-3-one glucoside methyltransferase activity. *The Plant Cell* **25**, 2341-2355.
- Meyer J, Snook M, Houchins K, *et al.* (2007) Quantitative trait loci for maysin synthesis in maize (*Zea mays* L.) lines selected for high silk maysin content. *Theoretical and Applied Genetics* **115**, 119-128.
- Ming R, Brewbaker J, Pratt R, Musket T, McMullen M (1997) Molecular mapping of a major gene conferring resistance to maize mosaic virus. *Theoretical and Applied Genetics* **95**, 271-275.
- Moeller DA, Tenaillon MI, Tiffin P (2007) Population structure and its effects on patterns of nucleotide polymorphism in teosinte (*Zea mays* ssp. *parviglumis*). *Genetics* **176**, 1799-1809.
- Moeller DA, Tiffin P (2008) Geographic variation in adaptation at the molecular level: a case study of plant immunity genes. *Evolution* **62**, 3069-3081.
- Moose SP, Lauter N, Carlson SR (2004) The Maize macrohairless1 Locus Specifically Promotes Leaf Blade Macrohair Initiation and Responds to Factors Regulating Leaf Identity. *Genetics* **166**, 1451-1461.
- Mullen LM, Hoekstra HE (2008) Natural selection along an environmental gradient: a classic cline in mouse pigmentation. *Evolution* **62**, 1555-1570.
- Mullen LM, Vignieri SN, Gore JA, Hoekstra HE (2009) Adaptive basis of geographic variation: genetic, phenotypic and environmental differences among beach mouse populations. *Proceedings of the Royal Society of London B: Biological Sciences* **278**, 3809-3818.
- Nei M (1973) Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences of the United States of America* **70**, 3321-3323.
- Nieto-Sotelo J, Martínez LM, Ponce G, *et al.* (2002) Maize HSP101 plays important roles in both induced and basal thermotolerance and primary root growth. *The Plant Cell* **14**, 1621-1633.
- Ninamango-Cárdenas FE, Guimarães CT, Martins PR, *et al.* (2003) Mapping QTLs for aluminum tolerance in maize. *Euphytica* **130**, 223-232.
- Nomura T, Ishihara A, Imaishi H, *et al.* (2003) Rearrangement of the genes for the biosynthesis of benzoxazinones in the evolution of Triticeae species. *Planta* **217**, 776-782.
- O'Brien AM, Sawers RJ, Ross-Ibarra J, Strauss SY (2015) Extending the Stress-Gradient hypothesis: increased local adaptation between teosinte and soil biota at the stressful end of a climate gradient. *bioRxiv*, 031195.
- Ogliari JB, Guimarães MA, Geraldi IO, Camargo LEA (2005) New resistance genes in the *Zea mays*: exserohilum turcicum pathosystem. *Genetics and Molecular Biology* **28**, 435-439.
- Olsson K, Ågren J (2002) Latitudinal population differentiation in phenology, life history and flower morphology in the perennial herb *Lythrum salicaria*. *Journal of Evolutionary Biology* **15**, 983-996.
- Pechan T, Jiang B, Steckler D, *et al.* (1999) Characterization of three distinct cDNA clones encoding cysteine proteinases from maize (*Zea mays* L.) callus. *Plant Molecular Biology* **40**, 111-119.
- Peiffer JA, Romay MC, Gore MA, *et al.* (2014) The genetic architecture of maize height. *Genetics* **196**, 1337-1356.
- Poland JA, Bradbury PJ, Buckler ES, Nelson RJ (2011) Genome-wide nested association mapping of quantitative resistance to northern leaf blight in maize. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 6893-6898.
- Pouvreau B, Baud S, Vernoud V, *et al.* (2011) Duplicate maize Wrinkled1 transcription factors activate target genes involved in seed oil biosynthesis. *Plant Physiology* **156**, 674-686.
- Pyhäjärvi T, Hufford MB, Mezouk S, Ross-Ibarra J (2013) Complex patterns of local adaptation in teosinte. *Genome biology and evolution* **5**, 1594-1609.
- Qiu F, Zheng Y, Zhang Z, Xu S (2007) Mapping of QTL associated with waterlogging tolerance during the seedling stage in maize. *Annals of Botany* **99**, 1067-1081.
- Quint M, Mihaljevic R, Dussle C, *et al.* (2002) Development of RGA-CAPS markers and genetic mapping of candidate genes for sugarcane mosaic virus resistance in maize. *Theoretical and Applied Genetics* **105**, 355-363.
- Reiter RS, Coors J, Sussman M, Gabelman W (1991) Genetic analysis of tolerance to low-phosphorus stress in maize using restriction fragment length polymorphisms. *Theoretical and Applied Genetics* **82**, 561-568.
- Ross-Ibarra J, Morrell PL, Gaut BS (2007) Plant domestication, a unique opportunity to identify the genetic basis of adaptation. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 8641-8648.
- Setter TL, Yan J, Warburton M, *et al.* (2010) Genetic association mapping identifies single nucleotide polymorphisms in genes that affect abscisic acid levels in maize floral tissues during drought. *Journal of experimental botany*, erq308.
- Shin EW, Han JS, Jang M, *et al.* (2004) Phosphate adsorption on aluminum-impregnated mesoporous silicates: surface structure and behavior of adsorbents. *Environmental science & technology* **38**, 912-917.



- Sibov S, Gaspar M, Silva M, *et al.* (1999) Two genes control aluminum tolerance in maize: genetic and molecular mapping analyses. *Genome* **42**, 475-482.
- Sigmon B, Vollbrecht E (2010) Evidence of selection at the ramosal locus during maize domestication. *Molecular Ecology* **19**, 1296-1311.
- Simcox KD, Bennetzen JL (1993) The use of molecular markers to study *Setosphaeria turcica* resistance in maize. *Phytopathology* **83**, 1326-1330.
- Simmons CR, Grant S, Altier DJ, *et al.* (2001) Maize rhm1 resistance to *Bipolaris maydis* is associated with few differences in pathogenesis-related proteins and global mRNA profiles. *Molecular Plant-Microbe Interactions* **14**, 947-954.
- Steward N, Kusano T, Sano H (2000) Expression of ZmMET1, a gene encoding a DNA methyltransferase from maize, is associated not only with DNA replication in actively proliferating cells, but also with altered DNA methylation status in cold-stressed quiescent cells. *Nucleic Acids Research* **28**, 3250-3259.
- Studer AJ, Doebley JF (2011) Do large effect QTL fractionate? a case study at the maize domestication QTL teosinte branched1. *Genetics* **188**, 673-681.
- Subbaiah CC, Sachs MM (2000) Maize cap1 encodes a novel SERCA-type calcium-ATPase with a calmodulin-binding domain. *Journal of Biological Chemistry* **275**, 21678-21687.
- Subbaiah CC, Sachs MM (2001) Altered patterns of sucrose synthase phosphorylation and localization precede callose induction and root tip death in anoxic maize seedlings. *Plant Physiology* **125**, 585-594.
- Svistoonoff S, Creff A, Reymond M, *et al.* (2007) Root tip contact with low-phosphate media reprograms plant root architecture. *Nature genetics* **39**, 792-796.
- Szalma S, Buckler IV E, Snook M, McMullen M (2005) Association analysis of candidate genes for maysin and chlorogenic acid accumulation in maize silks. *Theoretical and Applied Genetics* **110**, 1324-1333.
- Takuno S, Ralph P, Swarts K, *et al.* (2015) Independent molecular basis of convergent highland adaptation in maize. *Genetics* **200**, 1297-1312.
- Tanksley SD, McCouch SR (1997) Seed banks and molecular maps: unlocking genetic potential from the wild. *Science* **277**, 1063-1066.
- Tao Y, Jiang L, Liu Q, *et al.* (2013) Combined linkage and association mapping reveals candidates for Scmv1, a major locus involved in resistance to sugarcane mosaic virus (SCMV) in maize. *BMC plant biology* **13**, 162.
- Tian F, Bradbury PJ, Brown PJ, *et al.* (2011) Genome-wide association study of leaf architecture in the maize nested association mapping population. *Nature genetics* **43**, 159-162.
- Tracy WF, Whitt SR, Buckler ES (2006) Recurrent Mutation and Genome Evolution: Example of and the Origin of Sweet Maize. *Crop Science* **46**, S-49-S-54.
- Tzin V, Lindsay PL, Christensen SA, *et al.* (2015) Genetic mapping shows intraspecific variation and transgressive segregation for caterpillar-induced aphid resistance in maize. *Molecular Ecology* **24**, 5739-5750.
- van der Linde K, Hemetsberger C, Kastner C, *et al.* (2012) A maize cystatin suppresses host immunity by inhibiting apoplastic cysteine proteases. *The Plant Cell* **24**, 1285-1300.
- van Heerwaarden J, Doebley J, Briggs WH, *et al.* (2011) Genetic signals of origin, spread, and introgression in a large sample of maize landraces. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 1088-1092.
- Vicente-Carbajosa J, Moose SP, Parsons RL, Schmidt RJ (1997) A maize zinc-finger protein binds the prolamin box in zein gene promoters and interacts with the basic leucine zipper transcriptional activator Opaque2. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 7685-7690.
- Wang H, Nussbaum-Wagler T, Li B, *et al.* (2005) The origin of the naked grains of maize. *Nature* **436**, 714-719.
- Weber A, Clark RM, Vaughn L, *et al.* (2007) Major regulatory genes in maize contribute to standing variation in teosinte (*Zea mays* ssp *parviglumis*). *Genetics* **177**, 2349-2359.
- Weber AL, Briggs WH, Rucker J, *et al.* (2008) The genetic architecture of complex traits in teosinte (*Zea mays* ssp. *parviglumis*): new evidence from association mapping. *Genetics* **180**, 1221-1232.
- Weir BS, Hill WG (2002) Estimating F-Statistics. *Annual Review of Genetics* **36**, 721-750.
- Welcker C, Sadok W, Dignat G, *et al.* (2011) A common genetic determinism for sensitivities to soil water deficit and evaporative demand: meta-analysis of quantitative trait loci and introgression lines of maize. *Plant Physiology* **157**, 718-729.
- Welz H, Xia X, Bassetti P, Melchinger A, Lübberstedt T (1999) QTLs for resistance to *Setosphaeria turcica* in an early maturing Dent × Flint maize population. *Theoretical and Applied Genetics* **99**, 649-655.
- Whipple CJ, Kebrom TH, Weber AL, *et al.* (2011) grassy tillers1 promotes apical dominance in maize and responds to shade signals in the grasses. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 506-512.
- Whitt SR, Wilson LM, Tenailon MI, Gaut BS, Buckler IV ES (2002) Genetic diversity and selection in the maize starch

- pathway. *Proceedings of the National Academy of Sciences of the United States of America Sciences* **99**, 12959-12962.
- Wilson LM, Whitt SR, Ibanez AM, *et al.* (2004) Dissection of maize kernel composition and starch production by candidate gene association. *Plant Cell* **16**, 2719-2733.
- Wright SI, Bi IV, Schroeder SG, *et al.* (2005) The effects of artificial selection on the maize genome. *Science* **308**, 1310-1314.
- Xia X, Melchinger AE, Kuntze L, Lübberstedt T (1999) Quantitative trait loci mapping of resistance to sugarcane mosaic virus in maize. *Phytopathology* **89**, 660-667.
- Yamasaki M, Tenaillon MI, Bi IV, *et al.* (2005) A large-scale screen for artificial selection in maize identifies candidate agronomic loci for domestication and crop improvement. *Plant Cell* **17**, 2859-2872
- Yan J, Lipka AE, Schmelz EA, Buckler ES, Jander G (2014) Accumulation of 5-hydroxynorvaline in maize (*Zea mays*) leaves is induced by insect feeding and abiotic stress. *Journal of experimental botany* **66**, 593-602.
- Yin X, Wang Q, Yang J, *et al.* (2003) Fine mapping of the Ht2 (Helminthosporium turcicum resistance 2) gene in maize. *Chinese Science Bulletin* **48**, 165-169.
- Yu J, Holland JB, McMullen MD, Buckler IV ES (2008) Genetic design and statistical power of nested association mapping in maize. *Genetics* **178**, 539-551.
- Yu J, Li X, Zhu C, *et al.* (2012) Genic and non-genic contributions to natural variation of quantitative traits in maize. *Genome Research*, gr. 140277.140112.
- Zhang H, Uddin MS, Zou C, *et al.* (2014) Meta-analysis and candidate gene mining of low-phosphorus tolerance in maize. *Journal of integrative plant biology* **56**, 262-270.
- Zhang L, Peek AS, Dunams D, Gaut BS (2002) Population genetics of duplicated disease defense genes, hm1 and hm2, in maize (*Zea mays* ssp. *mays* L.) and its wild ancestor (*Zea mays* ssp. *parviglumis*). *Genetics* **162**, 851-860.
- Zila CT, Ogut F, Romay MC, *et al.* (2014) Genome-wide association study of Fusarium ear rot disease in the USA maize inbred line collection. *BMC plant biology* **14**, 372.



**IV. CHAPITRE 3 : RELIER LE POLYMORPHISME  
DES SNP CANDIDATS À DES VARIATIONS  
PHÉNOTYPIQUES**



## **IV.1 INTRODUCTION**

L'une des questions centrales en génétique des populations est la compréhension des processus adaptatifs. Cette compréhension est appuyée par la détection des variants ayant participé à ces adaptations. Dans le Chapitre 2 nous avons employé des approches d'évolution moléculaire pour identifier dans des génomes de téosinte les empreintes laissées par la sélection passée. Nous avons formulé l'hypothèse que de telles empreintes étaient le résultat de processus adaptatifs. Nous avons basé les analyses sur la détection de différenciation entre populations de basses et hautes altitudes. Nous avons également eu recours à une méthode de corrélation avec des variables environnementales pour appuyer les résultats de différenciation. Les variants détectés pointent sur des gènes ou des régions candidates à l'adaptation. Il reste cependant à démontrer leur effet sur la variation phénotypique, qui est la cible de la sélection naturelle. Pour démontrer un tel effet il faut qu'il existe encore, malgré la sélection, une certaine variabilité résiduelle pour pouvoir tester les associations entre variation génotypique et variation phénotypique. Cette démonstration est un premier pas vers une validation fonctionnelle des régions/gènes candidats, qui est l'objectif principal de ce chapitre.

Le principe de la génétique d'association est de rechercher des SNP associés plus fréquemment à des variations phénotypiques qu'attendu par hasard. Contrairement aux études de cartographie de QTL qui s'appuient sur la ségrégation de polymorphismes génétiques dans des populations où les pedigrees sont contrôlés, la génétique d'association recherche des associations avec des variants ségrégeant au sein de larges collections d'individus. Elle permet ainsi d'étudier des populations naturelles. Les marqueurs détectés doivent être cependant fortement liés aux polymorphismes responsables de la variation phénotypique. Ceci est particulièrement vrai pour le maïs, une espèce allogame dont le génome est fortement recombiné et dont le DL chute localement en 200 à 1500 pb (Remington *et al.* 2001). Pour cela, deux stratégies peuvent être employées : une stratégie consistant à utiliser un large panel de marqueurs, souvent fournis par des puces à ADN ou des données de séquençage haut-débit ; ou une approche gène candidats plus ciblée et dans laquelle les polymorphismes étudiés présentent de forts *a priori* d'association aux phénotypes. Ces deux approches ont des buts différents. La stratégie « *genome-wide* » cherche à élucider les bases génétiques d'un caractère morphologique, tandis que l'approche « gène candidat » permet de valider l'impact phénotypique d'un gène candidat sur un caractère particulier. Ces deux approches ne nécessitent pas le même nombre de marqueurs génétiques, les coûts sont aussi très différents, beaucoup plus élevés pour l'approche « *genome-wide* », surtout lorsqu'il s'agit de

séquençage. Les puces quant à elles utilisées dans la stratégie « *genome-wide* » ont un coût ultra-compétitif. Cependant les marqueurs sont prédéfinis et dépendent des panels utilisés lors de la construction de la puce (cf section I.1.3.a). Des biais importants existent donc pour les données issues de puces.

Quelle que soit la stratégie employée, il faut prendre en compte la structure des populations. En effet, la sélection naturelle n'est pas la seule responsable des différences génétiques entre populations. Des processus neutres comme la dérive et les flux de gènes sont aussi responsables de structuration entre populations. Si la variation phénotypique que l'on étudie est liée à la structure de populations, certains variants neutres vont apparaître comme associés à la variation phénotypique de façon fortuite, ce que l'on qualifie de faux positifs. Il est important de prendre en compte la structure des populations dans la génétique d'association afin de limiter les faux positifs (Atwell *et al.* 2010). Ce risque est particulièrement fort chez la téosinte qui possède une structuration relativement élevée et complexe (Moeller *et al.* 2007 ; Wright *et al.* 2005). Des outils statistiques ont été développés pour éliminer ces faux positifs des analyses. Cette structuration a d'abord été incorporée à des modèles de génétiques d'association sous la forme de matrice Q de structure (Pritchard *et al.* 2000b ; Yu *et al.* 2006), calculée par une méthode Bayésienne d'assignation à des groupes génétiques (Pritchard *et al.* 2000a), ou par les premiers axes d'une ACP des données génétiques (Price *et al.* 2006). Cependant, parce que les individus peuvent être apparentés au sein des populations, il est primordial d'ajouter une matrice d'apparentement des individus (Yu *et al.* 2006 ; Zhao *et al.* 2007). Cette matrice peut également prendre en compte une grande part de la structure entre populations, et le maintien de la matrice de structuration Q est à ce jour discuté. Ces matrices sont calculées traditionnellement à partir de marqueurs neutres répartis sur l'ensemble du génome. Il a cependant été proposé de les calculer en excluant le chromosome sur lequel est situé le SNP testé, afin de gagner en puissance de détection des SNP (Rincent *et al.* 2014). Ces méthodes statistiques qui visent à réduire le nombre de faux positifs entraînent une augmentation des faux négatifs, surtout quand le phénotype varie selon une dimension géographique comme la date de floraison (*Arabidopsis thaliana* : Aranzana *et al.* 2005 ; *Zea mays* : Camus-Kulandaivelu *et al.* 2006). Dans ce chapitre, nous nous intéressons à des variations phénotypiques résultant de l'adaptation locale telle que l'adaptation à l'altitude. Les variants génétiques responsables de ces adaptations peuvent être confondus avec la variation neutre suivant des gradients altitudinaux. La prise en compte de la structure de population va donc éliminer les variants liés à la structure de population, et probablement une partie de ceux liés à l'adaptation locale. Dans le but d'éviter les faux

positifs, on se place donc dans une situation qui favorise fortement les faux négatifs. S'il reste de la variation génétique et phénotypique intra-population pour les marqueurs et les caractères étudiés, il est néanmoins possible de détecter un lien statistique indépendamment de la structure des populations.

Les matrices destinées à prendre en compte la structure neutre peuvent être estimées à partir de marqueurs microsatellites. Les microsatellites sont de bons marqueurs pour étudier les mécanismes évolutifs à l'échelle des populations. En effet, ces petites séquences répétées ont un taux de mutation beaucoup plus élevé que les SNP (de  $10^{-4}$  à  $10^{-3}$  par locus et génération pour les microsatellites, Ellegren 2000 ; contre  $10^{-11}$  à  $10^{-8}$  pour les SNP, Baer *et al.* 2007 ; Clark *et al.* 2005). Ils peuvent présenter un grand nombre d'allèles, et évoluent plus vite que les SNP ce qui leur confère un pouvoir de 4 à 12 fois plus élevé que les SNP pour rendre compte de la structure inter individuelle dans les populations (Liu *et al.* 2005; Rosenberg *et al.* 2003).

Dans ce chapitre nous présentons une étude de génétique d'association qui est actuellement en cours. La première partie présente le dispositif général. La deuxième partie est consacrée aux analyses préliminaires des données phénotypiques, tandis que la troisième partie est consacrée à l'acquisition et aux analyses préliminaires des données génétiques. La dernière partie présente le modèle de génétique d'association en cours de construction.

## **IV.2 DISPOSITIF DE GENETIQUE D'ASSOCIATION**

### **IV.2.1 Choix des caractères phénotypiques**

Nous avons choisi 18 caractères phénotypiques qui nous paraissaient pertinents pour l'adaptation à l'altitude. Ces caractères ont été choisis avant de mener l'étude du Chapitre 2, nous avons donc peu de recul sur les caractères et pour les déterminer nous avons analysé les résultats d'un rapport publié par l'INIFAP en 1998 sur des données phénotypiques mesurées en jardins communs sur 4 localités (Chapingo, Mexico ; Iguala, Guerrero ; Tepatitlan, Jalisco ; Zapopan, Jalisco) entre 1986 et 1988. Deux cent huit populations de téosinte *Zea mays ssp parviglumis* et *Zea mays ssp mexicana* dont l'altitude a été notée ont été phénotypées sur 38 caractères morphologiques :

- des caractères de phénologie : nombre de jours jusqu'à germination ; nombre de jours jusqu'à émergence de la 5<sup>e</sup> feuille ; nombre de jours jusqu'à émergence de la 10<sup>e</sup> feuille ; nombre de jours jusqu'à 10% de la floraison mâle ; nombre de jours jusqu'à 50% de la floraison mâle ; nombre de jours jusqu'à 90% de la floraison mâle ; nombre de jours



jusqu'à 50% de la floraison femelle ;

- des caractères de l'appareil végétatif : nombre de feuilles ; longueur des feuilles ; largeur des feuilles ; hauteur de la plante ; hauteur du premier épi ; hauteur du dernier épi ; nombre de nœuds avec épis sur la talle principale ; nombre de nœuds avec épis sur les rameaux latéraux ; nombre de nœud avec épis sur les talles ; nombre de rameaux latéraux ; nombre total de talles ; longueur des rameaux latéraux ; nombre de nœud sur les rameaux latéraux ;

- des caractères de l'appareil reproducteur : nombre d'épis sur le dernier nœud ; nombre d'épis sur l'avant-dernier nœud ; nombre d'épis sur l'avant-avant-dernier nœud ; longueur des épis ; nombre de grains par épi ; poids de 100 grains ; volume de 100 grains ; nombre de ramifications dans la panicule ; longueur de la panicule ; longueur de la partie ramifiée de la panicule ; longueur du rameau principal de la panicule ; longueur des entrenœuds du rameau principal ; longueur des entrenœuds des rameaux latéraux ; largeur de l'épillet ; longueur de l'épillet ; longueur du pédicule de l'épillet ; longueur de la glume ; nombre de veines dans la glume.

Parmi ces caractères, 12 montrent un coefficient de corrélation avec l'altitude supérieur à 0,5 : la largeur des feuilles, la hauteur de la plante, la hauteur du premier épi, la hauteur du dernier épi, le nombre de nœuds avec épis sur les rameaux latéraux, la longueur des rameaux latéraux, le nombre d'épis sur l'avant dernier nœud, le nombre d'épis sur l'avant-avant dernier nœud, la longueur des épis, la longueur de la panicule, la largeur de l'épillet et la longueur de l'épillet. Certains des caractères phénotypiques étudiés précédemment sont connus pour varier avec l'altitude chez la téosinte (cf section I.2.3).

De ces analyses, de la lecture de la littérature et des discussions avec Salvador Hernandez Martinez sur la faisabilité des phénotypages dans la station, nous avons retenu 18 caractères phénotypiques, présentés ci-dessus :

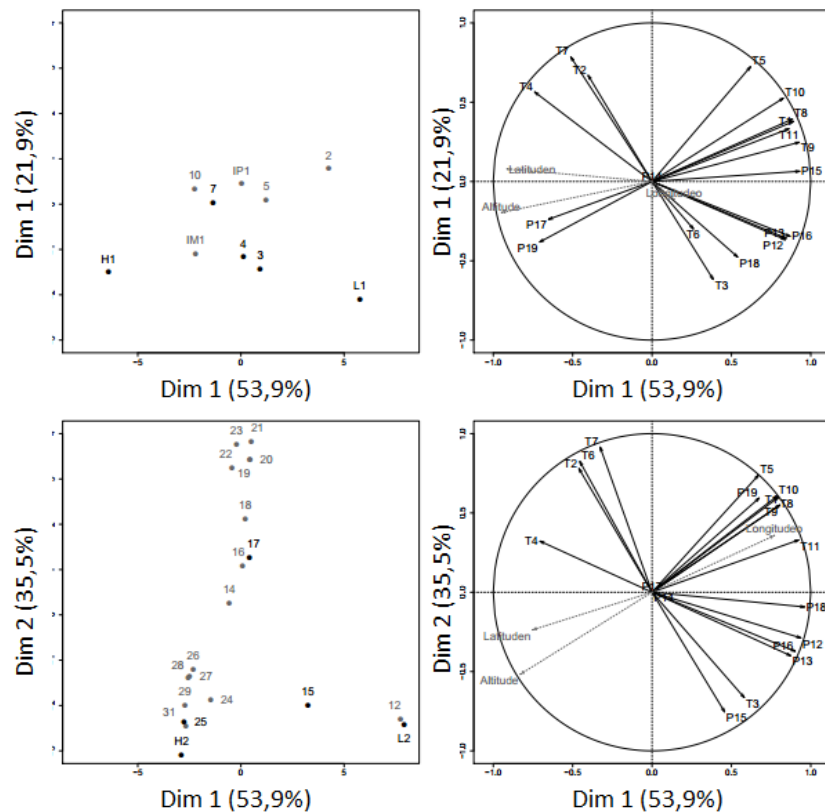
- la date de la floraison femelle : date de l'émergence des soies (en nombre de jours après semis)
- la date de la floraison mâle : date de la maturation des anthères (en nombre de jours après semis)
- la taille de la plante : hauteur de la talle primaire du sol jusqu'à la panicule terminale (en mètre)
- la hauteur du premier épi : du sol jusqu'au premier épi de la talle principale (ou talle de même hauteur si la talle principale n'est pas discernable, en mètre)

- la hauteur du dernier épi : du sol jusqu'au dernier épi de la talle principale (ou talle de même hauteur si la talle principale n'est pas discernable, en mètre)
- le nombre de talles
- le nombre de rameaux latéraux : rameaux partant de la talle principale (ou talle de même hauteur si la talle principale n'est pas discernable)
- le nombre de nœuds avec épis sur la talle principale (ou talle de même hauteur si la talle principale n'est pas discernable)
- la coloration des feuilles (notée de 1 à 6 : mesure moyenne sur l'ensemble des feuilles)
- la longueur des feuilles : mesure sur une feuille au centre de la plante (en centimètre)
- la largeur des feuilles : mesure sur une feuille au centre de la plante (en centimètre)
- le nombre de grains par épi : mesure sur un épi pour les maïs, sur cinq épis pour les téosintes
- le poids de 50 grains (en gramme)
- la longueur du grain : du côté le plus large pour la téosinte (en millimètre)
- la largeur du grain : de l'axe orthogonal au côté le plus large de la téosinte (en millimètre)
- la coloration des grains (notée de 1 à 6 : mesure moyenne sur l'ensemble des grains)
- le nombre de rameaux dans la panicule

Les stomates contrôlent l'évapotranspiration et le taux de photosynthèse. Ces structures peuvent être impliquées dans l'adaptation à des milieux selon le taux d'humidité. Nous avons mesuré la densité en stomates en collaboration avec Juliette de Meaux de l'université de Münster. Pour cela nous avons échantillonné trois feuilles, que nous avons conservées dans du papier buvard imbibé d'eau. Ces échantillons ont été envoyés à l'Institute for Evolution and Biodiversity en Allemagne où neuf photos ont été prises sur chacune des feuilles. La densité en stomates et leur taille sont mesurées par un logiciel d'analyse d'image. L'équipe de Juliette de Meaux a travaillé à améliorer la précision de détection de ce logiciel. Les données de stomates ont été reçues dernièrement et n'ont pas encore été analysées. Elles ne sont donc pas présentées dans la suite du manuscrit.

## IV.2.2 Choix des populations

Onze populations de téosinte ont été évaluées. Nous avons incorporé les quatre populations « extrêmes » présentées dans le chapitre 1 : H1, L1, H2, L2. Nous avons choisi sept populations supplémentaires de façon à maximiser leur variabilité environnementale d'origine. Nous nous sommes appuyés d'une part sur une ACP basée sur les variables climatiques afin de maximiser la variance des environnements d'origine de ces populations, et d'autre part nous avons favorisé l'éloignement géographique entre les populations (Figure 30, Supplementary Figure 1 du chapitre 1). Ce choix a également été contraint par le nombre de graines disponibles dans les stocks de graines des campagnes d'échantillonnage. Les populations choisies sont indiquées en noir sur les figures 1A et 1C.



**Figure 30 : Analyses en Composantes Principales basées sur 19 variables climatiques et 11 populations de téosintes échantillonnées sur le gradient 1 (A et B) et sur 20 populations de téosintes échantillonnées sur le gradient 2 (C et D).** A, C : Projection des populations sur le premier plan de l'ACP. Les populations H1, L1, 3, 4, 7, H2, L2, 15, 17 et 25, utilisées pour le dispositif de génétique d'association sont représentées en noir. (B, D) : Cercles des corrélations des variables sur le premier plan de l'ACP. T1 =Température annuelle moyenne, T2 : Gamme de Température journalière moyenne (Moyenne des mois (max temp-min temp), T3 : Isothermalité  $(T2/T7)*100$ , T4 : Saisonnalité de la température (écart-type\*100), T5 : Température maximale du mois le plus chaud, T6 : Température minimale du mois le plus froid, T7 : Gamme de variation annuelle des température (T5-T6), T8 : Température moyenne du trimestre le plus humide, T9 : Température moyenne du trimestre le plus sec, T10 : Température moyenne du trimestre le plus chaud, T11 : Température moyenne du trimestre le plus froid, P12 : Précipitations annuelles, P13 : Précipitation du mois le plus humide, P14 : Précipitation du mois le plus sec, P15 : Saisonnalité de la précipitation (coefficient de variation), P16 : Précipitation du trimestre le plus humide, P17 : Précipitation du trimestre le plus sec, P18 : Précipitation du trimestre le plus chaud, P19 : Précipitation du trimestre le plus froid.

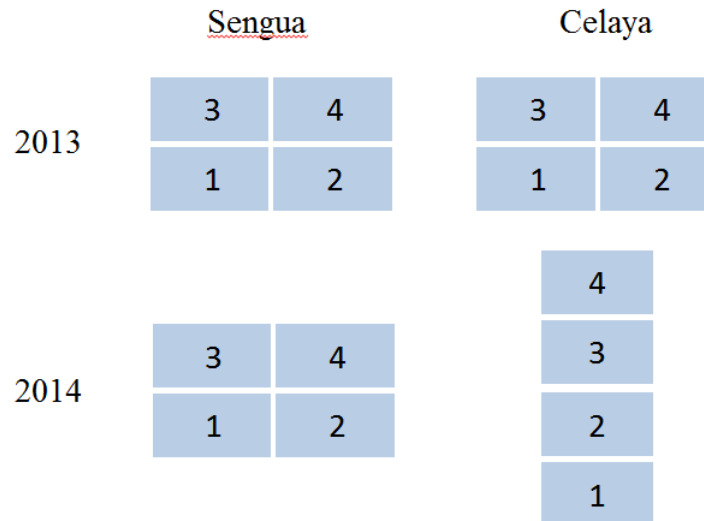
### **IV.2.3 Dispositif expérimental**

Nous avons mis en place un dispositif permettant d'évaluer la valeur phénotypique de plantes issues des onze populations sélectionnées. Les individus phénotypés sont issus de grains prélevés dans ces populations naturelles sur des plantes qui ne sont pas fixées génétiquement. Il est donc impossible de phénotyper plusieurs fois le même génotype dans notre dispositif, ce qui induit de fortes imprécisions de phénotypage. Pour pallier à cette difficulté, nous avons choisi (1) de travailler sur de gros effectifs, (2) de travailler, pour certaines analyses, sur des moyennes de plantes demi-sœurs issues de grains récoltés sur une même plante-mère, ces demi-sœurs étant phénotypées et génotypées individuellement et considérées comme des répliquats des plantes-mères.

Afin de contrôler au maximum les effets environnementaux et de déterminer si les variations génétiques que nous observions étaient similaires dans différents environnements (faibles interactions génotype x environnement), nous avons réalisé quatre essais, deux années (2013 et 2014) x deux lieux. En regard des différences environnementales que l'on retrouve dans les localités où poussent les téosintes de basse altitude et celles de haute altitude, il est peu probable de faire arriver à floraison les populations d'un extrême transplantées dans l'autre extrême. Nous avons donc choisi d'implanter les essais dans deux sites à des altitudes relativement intermédiaires. Pour des raisons logistiques, et afin de minimiser les effets d'interactions entre génotypes et environnement, nous avons choisi deux sites proches l'un de l'autre, vraisemblablement soumis à des conditions environnementales comparables. Les essais se sont déroulés sur deux sites expérimentaux appartenant à l'INIFAP dans l'état de Guanajuato: l'un situé à Celaya (20° 31' 20'' Nord, 100° 48' 44'' Ouest) à 1750 mètres d'altitude et l'autre appelé Sengua situé à San Luis de la Paz (21° 17' 55'' Nord, 100° 30' 59'' Ouest) à 2017 mètres d'altitudes. Les deux sites sont situés dans une zone de transition entre un climat continental et un climat steppique, avec des hivers doux et secs et des étés chauds et humides. Le site de Sengua est cependant plus sec que le site de Celaya. Afin de contrôler pour une éventuelle hétérogénéité environnementale locale, chaque essai a été structuré en quatre blocs. Pour chaque plante-mère, un descendant a été placé dans chaque bloc. Les mêmes plantes-mères ont été utilisées pour les 4 essais si les stocks de graines le permettaient. Pour les populations dans lesquelles ce n'était pas possible, certaines plantes-mères utilisées la première année ont été remplacées par d'autres plantes-mères pour les essais de la deuxième année. Les deux essais de chaque année comportaient toujours les mêmes plantes-mères pour toutes les populations. Les plantes-mères et leurs descendants utilisés dans le

dispositif sont répertoriés dans l'Annexe IV.1. Enfin, nous avons introduit au sein du dispositif des plantes témoins, cinq lignées de maïs tropical génétiquement fixées, afin d'estimer les variations environnementale au sein et entre les essais. Les lignées tropicales utilisées ont été fournies par l'INIFAP (LPC-15, B-16 et B-45) et le CIMMYT (CMK\_142 et CML-295).

Chaque essai comporte 800 plantes, 60 individus de téosinte pour chacune des onze populations (660 téosintes) et 28 individus de maïs pour chacune des cinq lignées (140 maïs). Ces individus ont été répartis dans quatre blocs de 200 individus organisés, chacun, en 10 colonnes et 20 lignes. Chaque bloc comporte une plante issue de chacune des quinze plantes-mères choisies par population, et sept plantes pour chaque lignée de maïs. A l'intérieur d'un bloc, les individus sont disposés de façon aléatoire tout en optimisant la répartition. Les populations de téosintes sont représentées une à deux fois par ligne et pas plus d'une fois par colonne. Les lignées de maïs sont représentées au maximum une fois par ligne et par colonne. Ces règles de répartition ont servi pour générer les plans de chacun des blocs d'un essai. Durant l'année 2013, les plans d'expérimentation ont été identiques à Celaya et à Sengua (il y a eu un problème de communication entre notre laboratoire et l'INIFAP). Ces plans ont été ensuite complètement redessinés, sur les mêmes critères, pour les essais de 2014, et ont été différents pour les deux lieux. Les blocs ont été disposés en carré (Celaya 2013, Sengua 2013 et 2014 ; de bas en haut et de gauche à droite bloc 1, bloc 2, bloc 3 et bloc 4. Figure 31) car les terrains expérimentaux ne présentaient *a priori* pas de gradients environnementaux. L'arrivée tardive de la saison des pluies en 2013 nous a cependant contraints à irriguer l'essai de Celaya en 2013, conduisant à un gradient d'humidité au sein de l'essai. Les blocs de l'essai de Celaya 2014 ont été disposés en ligne pour des contraintes de place au sein de l'unité expérimentale (du bloc 1 au bloc 4, Figure 31).



**Figure 31 : Disposition des blocs dans les essais.** Chaque bloc est numéroté et comporte 200 plantes disposées en 10 colonnes et 20 lignes.

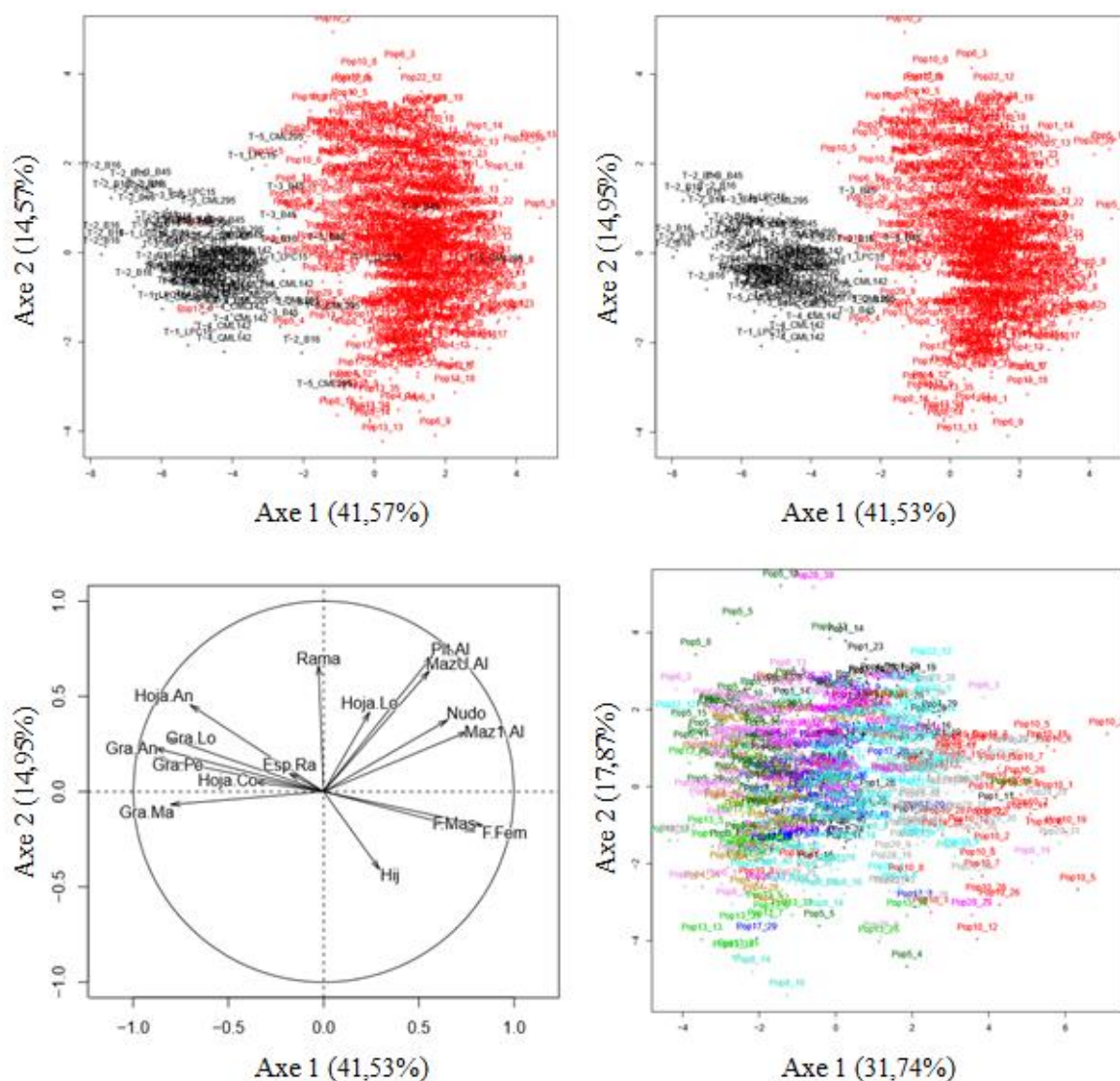
Les plans d'expérimentations ont été fournis aux collaborateurs mexicains en charge du dispositif. Les grains ont été semés en pots puis transplantés *in situ*. Afin de prendre en compte le risque de non germination, les collaborateurs ont semé douze grains par plante-mère par année au lieu des huit nécessaires. Dans les cas où moins de huit grains ont germé, l'emplacement a été rempli si possible par un individu provenant de la même population mais d'une autre plante-mère pour laquelle plus de huit grains avaient germé. Si aucun individu excédentaire n'était disponible pour la population, l'emplacement a été rempli par une plante provenant d'une autre population. Ces individus ne seront pas pris en compte pour la génétique d'association, mais ont servi à maintenir la densité des plantes au champ et ont été incorporés aux analyses phénotypiques préliminaires.

### **IV.3 ANALYSES DES DONNEES PHENOTYPIQUES**

#### **IV.3.1 Nettoyages des données phénotypiques**

Des ACP des individus sur les caractères phénotypiques ont été réalisées afin de repérer d'éventuelles erreurs de semis dans les champs (Figure 32). En effet, les compartiments maïs et téosintes se distinguent bien et ont permis de réattribuer certains individus à leur véritable appartenance (deux plantes inversées, lignes décalées ou inversées) (Figure 32-A, B). Une ACP a ensuite été réalisée sur les individus de téosintes uniquement. L'axe 1 sépare assez bien les populations de téosinte. Il est intéressant de noter que les populations extrêmes sont opposées sur cet axe (populations de basse altitude L1 en marron clair et L2 en vert clair, populations de haute altitude H1 en rouge et H2 en turquoise, Figure 32-D). Cet axe rend

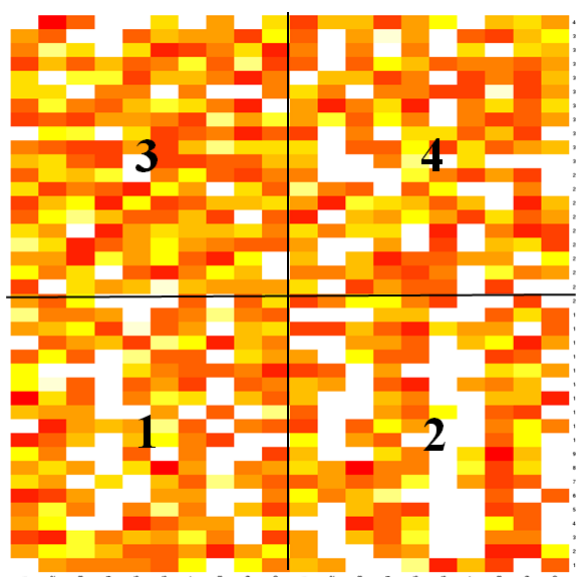
donc compte de variations liées à l'altitude.



**Figure 32 : ACP des individus du champ de SENGUA en 2014.** A, B. Les maïs sont représentés en noir, les téosintes en rouge. A. Données brutes. B. Données après réattribution des noms. C. Cercle des corrélations phénotypiques pour l'ACP comprenant maïs et téosintes avec les données corrigées. D. Les individus de l'ACP avec uniquement les téosintes sont représentés selon leur population d'origine par un code couleur.

### IV.3.2 Modélisation des effets environnementaux intra-essai

Les conditions environnementales, climatiques et/ou biotiques, peuvent varier grandement entre les essais expérimentaux. A l'échelle la plus fine, il peut exister une hétérogénéité importante à l'intérieur d'un même essai. Cette hétérogénéité éventuelle a été visualisée au moyen de « heatmaps » (Figure 33). Pour chacun des caractères, aucun des essais ne présente une structuration spatiale assez forte pour être discernable sur les « heatmaps ».



**Figure 33 : Visualisation des valeurs de la largeur de feuille au sein de l'essai Sengua 2014.** Les blocs sont délimités par les lignes noires et numérotés. Chaque individu est représenté par un rectangle. Les valeurs phénotypiques sont représentées selon un code couleur croissant. Les individus absents (mortalité ou non phénotypé) sont représentés en blanc.

L'hétérogénéité de chaque essai a été analysée par deux modèles en utilisant les lignées témoins. Le premier modèle (M1) prend en compte un effet bloc, et le second modèle (M2) un effet des lignes et des colonnes.

$$\text{M1 Phenotype} = \text{population} + \text{bloc} + \text{résiduelle}$$

$$\text{M2 Phenotype} = \text{population} + \text{ligne} + \text{colonne} + \text{résiduelle}$$

Les deux modèles ont ensuite été comparés avec le critère d'information bayésien (BIC). Trois des caractères présentaient trop de données manquantes pour le modèle M2. Sur les 15 caractères testés, 14 présentent un BIC plus petit pour le modèle M1 dans l'ensemble des essais où la comparaison était possible. Seul le caractère « longueur des grains » présente un BIC plus petit pour le modèle M2 dans les essais Celaya 2014 (Tableau 5). Devant l'importance du signal en faveur du modèle M1 et pour un souci d'homogénéité des analyses, nous avons choisi de prendre en compte l'hétérogénéité intra-essai par un effet bloc.



**Tableau 5 : Comparaison des modèles M1 et M2.**

	Celaya 2013		SENGUA 2013		Celaya 2014		SENGUA 2014	
	M1	M2	M1	M2	M1	M2	M1	M2
F.Fem <sup>1</sup>	<b>402,42</b>	411,63	<b>597,71</b>	619,55	<b>833,56</b>	1002,68	<b>725,12</b>	886,35
F.Mas <sup>2</sup>	<b>399,51</b>	407,15	<b>634,61</b>	650,82	<b>836,12</b>	1000,65	<b>734,19</b>	799,84
Plt.Al <sup>3</sup>	<b>21,21</b>	45,86	<b>58,29</b>	78,34	<b>74,74</b>	222,79	<b>27,75</b>	137,17
Maz1.Al <sup>4</sup>	<b>34,62</b>	51,17	<b>-85,59</b>	-64,43	<b>-4,02</b>	64,91	<b>-18,60</b>	1,70
MazU.Al <sup>5</sup>	<b>78,10</b>	98,39	<b>-29,68</b>	-8,97	<b>-37,21</b>	-14,38	<b>-55,42</b>	-97,13
Hoja.Lo <sup>6</sup>	<b>-82,15</b>	-68,50	<b>-87,43</b>	-74,36	<b>-64,50</b>	80,52	<b>-117,71</b>	6,64
Hoja.An <sup>7</sup>	<b>-43,72</b>	-21,83	<b>-471,86</b>	-459,03	<b>438,93</b>	574,89	<b>345,13</b>	474,00
Hoja.Co <sup>8</sup>	<b>126,60</b>	132,99	<b>24,73</b>	47,79	<b>230,69</b>	438,73	<b>50,93</b>	180,89
Nudo <sup>9</sup>	<b>128,54</b>	136,03	<b>97,80</b>	107,93	<b>305,97</b>	395,61	<b>187,70</b>	360,42
Esp.Ra <sup>10</sup>	<b>365,11</b>	385,53	<b>514,84</b>	534,93	<b>704,39</b>	892,96	<b>558,60</b>	689,88
Gra.Ma <sup>11</sup>	<b>682,54</b>	687,99	<b>1032,70</b>	1044,05	<b>1335,61</b>	1411,49	<b>1029,53</b>	1109,42
Gra.Pe <sup>12</sup>	<b>321,81</b>	329,41	<b>411,40</b>	440,37	<b>650,41</b>	683,18	<b>481,97</b>	547,67
Gra.Lo <sup>13</sup>	<b>197,12</b>	209,60	<b>316,49</b>	333,47	419,40	<b>410,26</b>	<b>336,70</b>	405,59
Gra.An <sup>14</sup>	<b>188,45</b>	202,09	<b>296,60</b>	314,28	<b>309,05</b>	336,13	<b>267,25</b>	359,42
Gra.Co <sup>15</sup>	<b>113,27</b>	122,51	<b>133,74</b>	158,93	<b>136,38</b>	345,89	78,19	NA

Les BIC les plus faibles entre les modèles M1 et M2 sont représentés en gras. NA : test non réalisé car ajustement pratiquement parfait.

<sup>1</sup> date de la floraison femelle

<sup>2</sup> date de la floraison mâle

<sup>3</sup> taille de la plante

<sup>4</sup> la hauteur du premier épi

<sup>5</sup> la hauteur du dernier épi

<sup>6</sup> longueur des feuilles

<sup>7</sup> largeur des feuilles

<sup>8</sup> coloration des feuilles

<sup>9</sup> nombre de nœuds avec épis sur la talle

<sup>10</sup> nombre de rameaux dans la panicule

<sup>11</sup> nombre de grains par épi

<sup>12</sup> poids de 50 grains

<sup>13</sup> longueur des grains

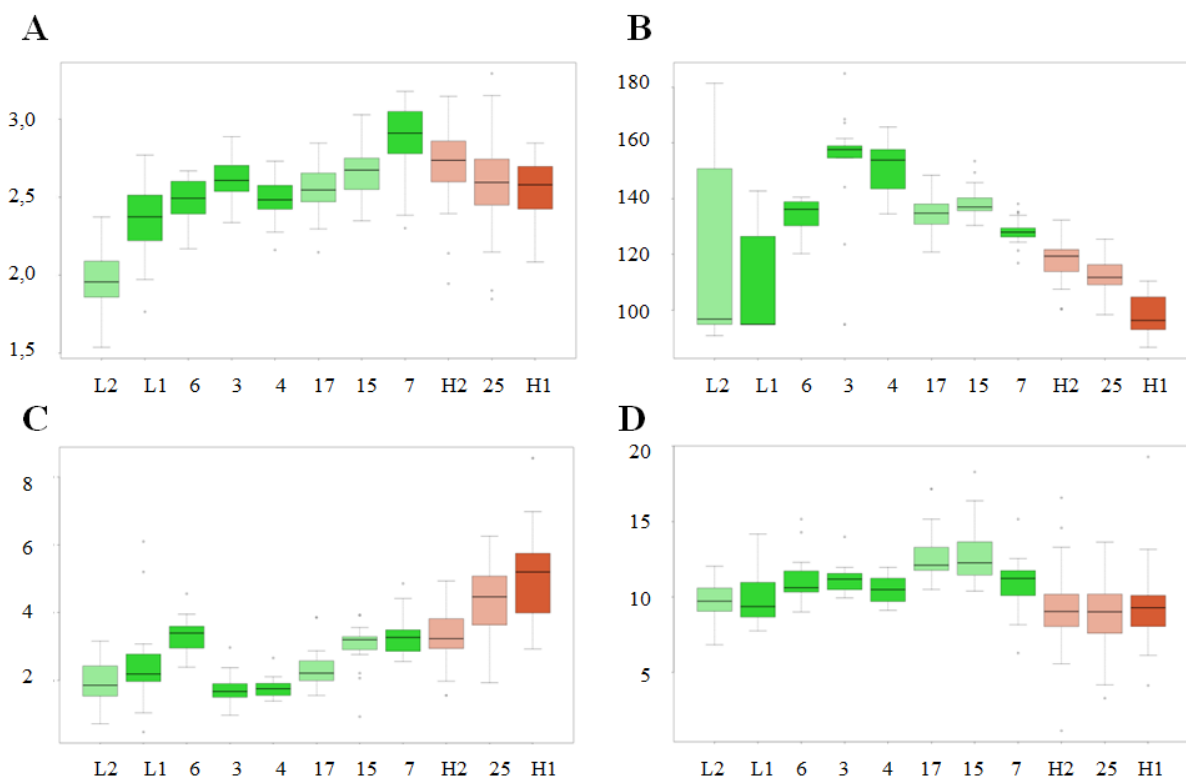
<sup>14</sup> largeur des grains

<sup>15</sup> coloration des grains

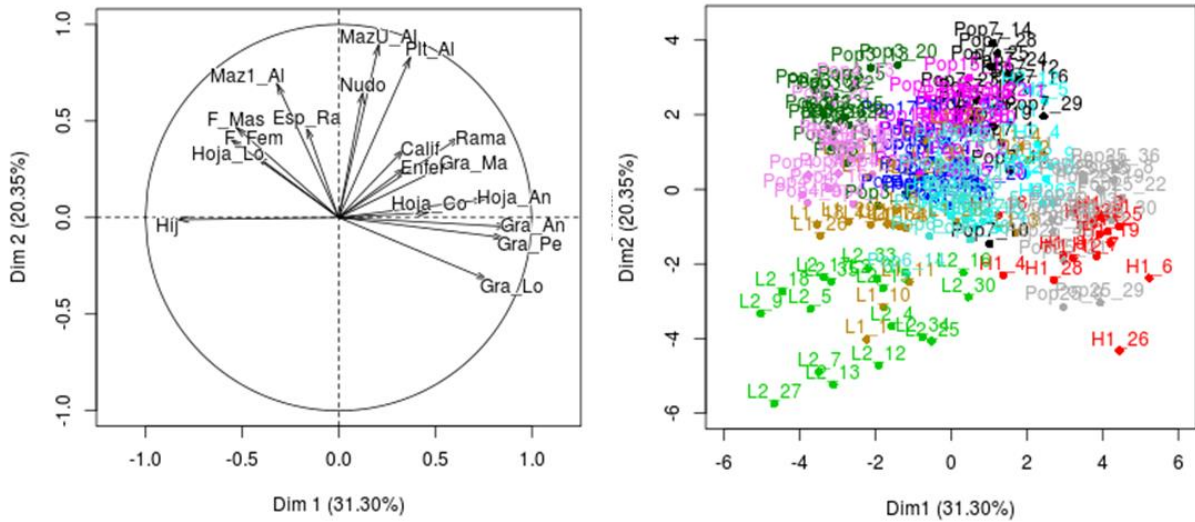
### **IV.3.3 Analyses des données phénotypiques**

Pour chaque caractère, des moyennes phénotypiques ont été calculées au niveau des plantes mères et des populations : les effets de l'année, du site et du bloc ont été retranchés sous R à l'aide de la fonction *predict* du package *asreml*. La distribution des valeurs phénotypiques corrigées au niveau des plantes mères est présentée au moyen de boxplots (Figure 34). Les individus d'une même population présentent globalement des phénotypes homogènes (Figure 35). Il est intéressant de noter que les populations *mexicana* (H1, H2, 25) ont des phénotypes assez similaires mais non clairement distingués de ceux des populations *parviglumis* (Figure 35). Certains caractères phénotypiques sont fortement corrélés entre eux (Figure 36). Les

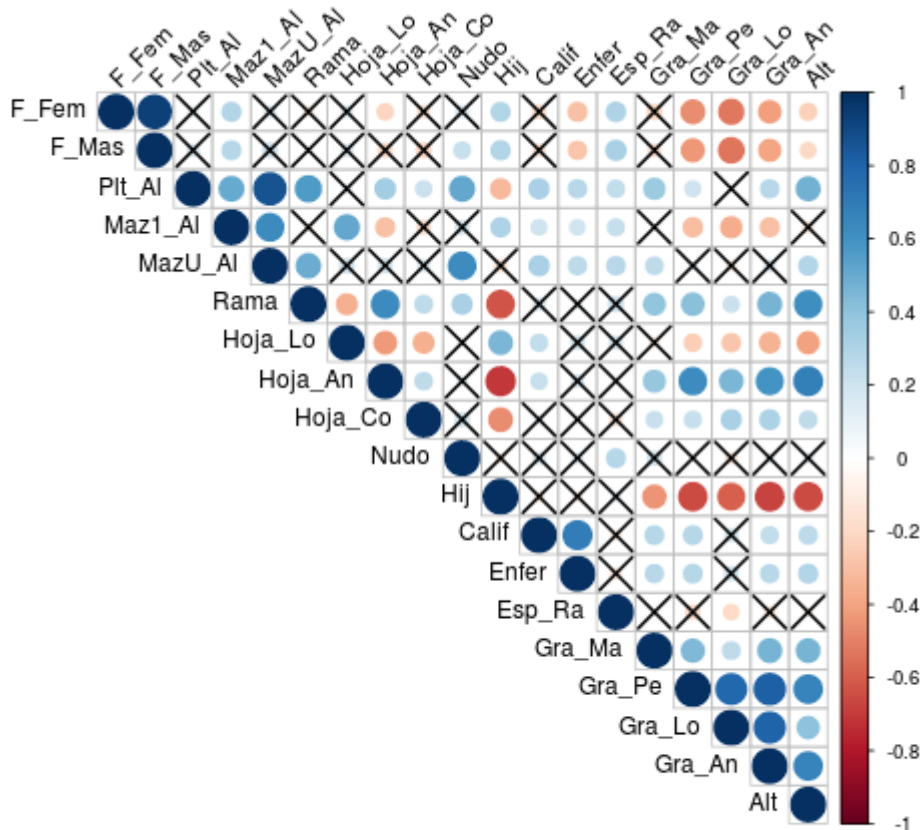
caractères de date de floraison mâle et femelle sont comme attendu très fortement corrélés entre eux, tout comme les caractères de poids ou taille des grains. Ces deux ensembles de caractères sont corrélés négativement entre eux (Figure 36). Ils varient en fonction de l'altitude (Figure 34-B et C). En effet, à l'exception des populations L1 et L2, qui n'ont pas ou très peu fleuri et pour lesquelles nous ne disposons pas d'assez de données, nous pouvons observer que les floraisons mâle et femelle sont de plus en plus précoces en lorsque les plantes viennent d'altitude élevée (Figure 34-B). Les caractères de l'architecture générale de la plante semblent aussi varier en fonction de l'altitude de la population d'origine, comme attendu par les nombreuses observations de changement de morphologie entre téosinte de haute et de basse altitude (Figure 35). Plus les plantes proviennent de haute altitude, plus elles sont grandes et moins elles possèdent de talles, jusqu'à une certaine altitude où ces deux caractères semblent se stabiliser (Figure 34). Les caractères de morphologie des feuilles ne présentent eux ni d'effet plantes-mères (Tableau 6) ni de patron discernable avec l'altitude (Figure non montrée).



**Figure 34 : Boxplots (ou boîtes à moustaches) par population des valeurs phénotypiques ajustées par plante mère pour les effets de l'année, du site et du bloc.** Les populations sont ordonnées par altitude croissante. Les populations *parviglumis* sont représentées en vert, les *mexicana* en rouge. Les populations provenant du gradient 1 sont représentées en foncé, celles du gradient 2 en clair. A. Taille de la plante (m). B. Date de floraison mâle (jours après semis). C. Poids de 50 grains (g). D. Nombre d'épillettes dans la panicule.



**Figure 35 : ACP des caractères phénotypiques au niveau des plantes-mères.** Les plantes-mères sont représentées par un code couleur selon leur population d'origine (L1 en marron clair, L2 en vert clair, 3 en vert foncé, 4 en rose, 6 en bleu clair, 7 en noir, 15 en violet, 17 en bleu foncé, 25 en gris, H2 en turquoise et H1 en rouge).



**Figure 36 : Corrélations des caractères phénotypiques et de l'altitude.** La taille et la coloration du cercle indiquent la valeur de la corrélation. Les corrélations avec une P-valeur supérieures à 0.01 sont indiquées avec une croix noire.

La pigmentation des grains et celle des feuilles semblent être des caractères adaptatifs liés à l'altitude chez la téosinte (Lauter *et al.* 2004). Cependant, nos mesures ont été réalisées à l'œil nu par différents observateurs et sont particulièrement subjectives. Pour la coloration des feuilles, les différences étaient très subtiles, et les phénotypes obtenus peu variables. Inversement, les grains présentaient des colorations très variées qui nous ont permis de faire des catégories de couleurs plus nombreuses. Les caractères quantitatifs de taille des grains sont fortement et positivement corrélés (Figure 36) et présentent tous le même patron : les valeurs (poids, longueur, largeur) augmentent fortement avec l'altitude pour les trois premières populations, puis diminuent entre les populations 6 et 3 pour augmenter à partir de la population 3 avec l'altitude (Figure 34-C). Le taux de germination serait relié à la taille du grain (Moles & Westoby 2004). Les conditions plus extrêmes en haute altitude, notamment le froid qui limite la germination, ont pu être des pressions de sélection en faveur de grains plus gros pour augmenter la probabilité de germination.

#### **IV.3.4 Effets populations et plantes-mères**

Un travail préliminaire a été réalisé pour tester l'effet des populations, des plante-mères, des années, des sites et des blocs. Deux modèles comprenant les effets environnementaux et la structuration intégrée par l'effet fixe population (M3) ou l'effet fixe population et l'effet aléatoire plante-mère (M4) ont été analysés. Ces deux modèles comprennent un effet bloc hiérarchisé dans la localité et des effets d'interactions localité x année, localité x population et année x population :

(M3) Phénotype = Localité + Année + Localité(Bloc) + Population +  
Localité:Année + Localité:Population + Année:Population + **Résiduelle**

(M4) Phénotype = Localité + Année + Localité(Bloc) + Population + **Plante-mère** +  
Localité:Année + Localité:Population + Année:Population + **Résiduelle**

L'effet plante-mère est ici considéré en tant qu'effet aléatoire et a donc été testé par un test de ratio des vraisemblances entre les modèles (M3) et (M4) qui suit une loi du chi-deux à 1 degré de liberté.

$$LRT=2*(\log(Vraisemblance(M4))-\log(Vraisemblance(M3)))$$

Les effets fixes des modèles (M3) ou (M4) ont été testés par un test de Wald pour chacun des

caractères (Tableau 6).

Le Tableau 6 montre que le modèle (M4) explique significativement mieux les données dans la majorité des cas (Tableau 6). Il est donc utile de déclarer un effet plante-mère, ce qui suggère des effets génétiques et/ou maternels importants. Il montre aussi des effets Localité, Année et Population significatifs pour l'ensemble des caractères (Tableau 6). Ceux-ci sont donc très importants à prendre en compte pour l'analyse. Par ailleurs, les interactions Localité x Année, Localité x Population, Année x Population sont souvent significatives, pour la majorité des caractères et pour les deux premières interactions. L'existence d'une interaction Localité x Population peut indiquer de fort effets « Génotype x Environnement ». L'effet Bloc est quant à lui globalement assez peu significatif, notamment pour les caractères liés à la valeur sélective comme la date de floraison et les caractères des grains.

**Tableau 6 : P-values de la comparaison des modèles M3 et M4, et des effets dans le modèle M3 ou M4, pour 17 caractères phénotypiques.**

	Modèle	Test LRT	Localité	Année	Population	at(bloc, 1):	at(bloc, 2):	at(bloc, 3):	Localité:	Année:	Localité:
						Localité	Localité	Localité	Année	Population	Population
F.Fem <sup>1</sup>	M4	4,88E-12	4,86E-25	5,65E-125	4,55E-73	8,37E-01	9,59E-01	5,60E-01	7,84E-09	4,39E-17	2,30E-13
F.Mas <sup>2</sup>	M4	8,44E-06	1,25E-05	9,25E-66	6,64E-74	1,88E-01	9,50E-01	3,56E-01	8,90E-01	1,08E-11	9,81E-09
Plt.Al <sup>3</sup>	M4	2,64E-05	7,07E-298	1,60E-25	1,46E-32	1,00E-01	1,45E-03	4,82E-03	3,11E-24	6,15E-02	2,99E-12
Maz1.Al <sup>4</sup>	M4	1,69E-02	3,52E-171	2,22E-03	4,42E-09	1,83E-004	4,02E-02	1,16E-02	5,02E-18	5,08E-02	8,66E-03
MazU.Al <sup>5</sup>	M4	6,94E-04	6,97E-280	2,86E-40	4,47E-24	1,57E-06	2,63E-02	1,15E-02	6,14E-28	4,46E-02	1,15E-15
Rama <sup>6</sup>	M4	6,11E-06	2,93E-14	9,07E-42	2,21E-37	3,82E-01	8,89E-01	2,66E-01	5,71E-23	3,69E-03	1,41E-06
Hoja.Lo <sup>7</sup>	M3	1,09E-01	1,25E-224	2,12E-16	1,23E-48	3,55E-03	1,37E-01	4,47E-03	5,03E-05	5,88E-17	1,00E-15
Hoja.An <sup>8</sup>	M3	9,67E-02	7,12E-12	0,00E+00	1,23E-81	9,71E-03	1,58E-01	9,00E-01	8,66E-06	1,33E-39	4,60E-03
Hoja.Co <sup>9</sup>	M3	7,81E-01	7,60E-02	4,52E-13	2,46E-09	8,52E-06	2,07E-01	1,52E-03	3,06E-01	3,63E-01	1,01E-02
Nudo <sup>10</sup>	M4	3,22E-02	8,73E-07	6,73E-21	2,41E-10	8,84E-02	1,32E-01	4,89E-01	1,99E-01	1,46E-05	5,48E-09
Hij <sup>11</sup>	M4	4,97E-03	8,85E-49	2,07E-03	3,86E-53	2,05E-05	2,61E-03	5,94E-02	6,78E-18	2,14E-02	3,38E-09
Esp.Ra <sup>12</sup>	M3	8,01E-01	7,34E-08	1,04E-61	3,81E-31	7,19E-01	2,03E-02	8,83E-07	9,13E-01	1,09E-31	1,19E-02
Gra.Ma <sup>13</sup>	M4	2,70E-11	7,13E-09	1,77E-21	1,50E-11	7,31E-01	3,65E-02	4,12E-02	5,95E-01	5,12E-04	2,40E-01
Gra.Pe <sup>14</sup>	M4	4,62E-08	9,60E-02	2,31E-05	1,63E-56	3,19E-01	3,47E-01	3,01E-01	1,61E-02	4,25E-01	1,64E-18
Gra.Lo <sup>15</sup>	M4	3,92E-12	1,96E-01	4,29E-42	8,84E-56	4,99E-01	6,66E-02	7,67E-02	2,08E-03	5,31E-08	1,87E-03
Gra.An <sup>16</sup>	M4	2,07E-02	1,13E-01	5,01E-06	2,08E-39	3,51E-01	4,19E-01	1,06E-01	5,91E-05	1,94E-05	7,05E-03
Gra.Co <sup>17</sup>	M4	4,32E-02	2,65E-01	8,76E-57	1,57E-10	1,31E-01	3,33E-01	1,31E-01	9,20E-01	1,94E-01	1,20E-11

Les effets significatifs sont représentés en rouge pour un seuil de 0,05. Il n'y a pas eu de correction pour les tests multiples.

<sup>1</sup> date de la floraison femelle

<sup>2</sup> date de la floraison mâle

<sup>3</sup> taille de la plante

<sup>4</sup> la hauteur du premier épi

<sup>5</sup> la hauteur du dernier épi

<sup>6</sup> nombre de rameaux latéraux

<sup>7</sup> longueur des feuilles

<sup>8</sup> largeur des feuilles

<sup>9</sup> coloration des feuilles

<sup>10</sup> nombre de nœuds avec épis sur la talle

<sup>11</sup> nombre de talles

<sup>12</sup> nombre de rameaux dans la panicule

- <sup>13</sup> nombre de grains par épi
- <sup>14</sup> poids de 50 grains
- <sup>15</sup> longueur des grains
- <sup>16</sup> largeur des grains
- <sup>17</sup> coloration des grains

## **IV.4 ACQUISITION ET ANALYSE DES DONNEES GENETIQUES.**

### **IV.4.1 Prélèvements et extractions d'ADN**

Seuls les individus de téosintes, pour lesquels nous voulons génotyper les SNP candidats, ont été prélevés. Les prélèvements de feuilles ont été effectués au mois de Juin pour l'année 2013 et au mois de Juillet pour l'année 2014. Ils ont été faits en 3 exemplaires : un à destination du laboratoire du Moulon, un pour les collaborateurs de l'UNAM et un comme réserve, stocké dans le laboratoire du Langebio (Mexique). Pour chacun de ces exemplaires un morceau de feuille adulte d'une dizaine de centimètres a été prélevé et stocké dans une enveloppe. Au bout de chaque colonne, les enveloppes étaient mises dans un sachet en plastique double zippé, le tout conservé dans de la glace. L'exemplaire à destination du laboratoire Français a été ensuite lyophilisé dans le laboratoire du Langebio, puis conservé dans des sachets plastiques double zippés en présence de silicagel. . Ce jeu de prélèvements a été acheminé en France par avion. Les deux autres jeux de prélèvements ont été plongés dans de l'azote liquide puis conservés dans des congélateurs à -80°C.

Pour le prélèvement acheminé en France, 30 mg de feuille lyophilisée ont été découpés et placés dans un puits d'une plaque DEEPWELL format 96 contenant deux billes d'acier de 4mm de diamètre. Pour faciliter l'étape de broyage, les feuilles ont été prédécoupées aux ciseaux. L'ADN de 2 415 individus a ensuite été extrait selon un protocole « maison » d'extraction en plaques décrit au Chapitre 2.

Les SNP candidats à tester pour la génétique d'association sont les mêmes que ceux du Chapitre 2 (cf section III.2).

### **IV.4.2 Description des marqueurs microsatellites**

Quarante-six marqueurs microsatellites ont été génotypés (Tableau 7) afin de décrire la structure neutre de l'échantillonnage. Les microsatellites utilisés ont des motifs de répétition de 2, 3, 4 ou 5 nucléotides (1 motif de 2 nuc., 25 motifs de 3 nuc., 16 motifs de 4 nuc., 1 motif de 5 nuc. et 3 motifs dégénérés de 4, 3 ou 2 nuc, Tableau 7). Les microsatellites comportant des motifs de deux bases évoluent plus vite que ceux comportant des motifs plus importants, et sont donc plus informatifs dans des études de structure de populations (Guichoux *et al.* 2011). Généralement plus petits par rapport aux autres, ils permettent de construire des multiplexages avec un plus grand nombre de marqueurs (Guichoux *et al.* 2011). Cependant, nous avons évité ces microsatellites car ils évoluent trop vite chez le maïs



(Vigouroux *et al.* 2002) et s'éloignent du modèle de mutations infinies car ils présentent de l'homoplasié (Excoffier & Hamilton 2003). De plus, ces marqueurs sont sujets au phénomène de stuttering (Guichoux *et al.* 2011), c'est-à-dire au déplacement d'une base de la polymérase lors de la PCR, ce qui complique la lecture et l'identification des allèles. Les 46 microsatellites choisis ont déjà été utilisés au laboratoire (Camus-Kulandaivelu 2006). La connaissance des allèles déjà identifiés chez le maïs a guidé le génotypage des individus de téosinte (Tableau 7).

**Tableau 7 : Description et Génotypage des 46 marqueurs microsatellites utilisés pour décrire la structure des populations et les apparentements entre individus.**

Trio <sup>1</sup>	SSR	Lab name	Fluo	Chro <sup>2</sup>	Motif	Nb allèles maïs	Range	Nb allèles téosinte	Taux de réussite (%)
1	phi046	40	FAM	3	ACGC	2	62-66	8	86,3
2	phi427434	492	FAM	2	ACC	7	123-144	11	74,9
	phi112	70	VIC	7	AG	21	125-163	19	77,6
	umc1496	1296	NED	5	GCA	13	135-164	22	71,2
3	phi331888	845	FAM	5	AAG	10	124-138	13	87,9
	phi029	214	VIC	8	AG/AG CG	10	145-165	23	80,7
	phi031	6	NED	6	GTAC	6	186-228	L <sup>a</sup>	-
4	phi121	76	FAM	8	CCG	4	97-106	L	-
	phi127	80	VIC	2	AGAC	12	97-132	14	85,5
	phi065	228	NED	9	CACTT	4	130-149	6	91,1
5	phi059	48	FAM	10	ACC	6	115-159	10	74,9
	phi109188	1293	VIC	5	AAAG	13	146-177	18	82,4
	phi402893	469	NED	2	AGC	14	207-243	17	75,1
6	umc1319	718	FAM	10	ACC	5	113-125	6	94,1
	phi308707	1177	VIC	1	AGC	7	114-132	8	84,3
	phi064	51	NED	1	ATCC	17	73-112	19	75,8
7	phi084	45	FAM	10	GAA	4	149-160	10	85,7
	phi104127	497	VIC	3	ACCG	7	156-168	11	75,6
	phi453121	496	NED	3	ACC	10	207-228	L	-
8	phi108411	1279	FAM	9	AGCT	6	117-139	L	-
	phi034	216	VIC	7	CCT	9	119-145	16	84,7
	phi330507	846	NED	5	CCG	4	132-142	6	76,4
9	phi062	49	FAM	10	ACG	8	150-179	11	90,7
	phi024	32	VIC	5	CCT	5	161-176	12	87,1
	phi032	215	NED	9	AAAG	6	232-240	13	79,1
10	phi002	208	FAM	1	AACG	2	71-75	8	91,9
	phi089	63	VIC	6	ATGC	5	85-94	12	92
	phi014	25	NED	8	GGC	7	148-169	10	90,3
11	phi078	59	FAM	6	AAAG	21	122-214	B <sup>b</sup>	-
	phi116	73	VIC	7	ACTG/ ACG	14	140-175	20	89,5

	phi033	34	NED	9	AAG	9	234-261	14	88,4
12	phi109275	998	FAM	1	AGCT	13	120-140	19	63,2
	umc1133	1298	VIC	6	ATAC	7	86-103	17	69,8
	phi233376	641	NED	8	CCG	9	137-157	18	63
13	phi448880	1280	FAM	9	AAG	4	171-186	6	93,1
	dupssr34	19	VIC	4	TTG	26	114-176	39	88,2
	phi213984	1291	NED					B	
				4	ACC	3	284-302		-
14	phi102228	1278	FAM	3	AAGC	5	123-135	14	85,9
	phi072	56	VIC	4	AAAC	14	132-266	L	-
	phi051	43	NED	7	AGG	8	137-147	13	66,8
15	phi053	44	FAM	3	ATAC	12	167-213	19	95,1
	phi115	248	VIC		AT/AT				
				8	AC	3	290-310	2	97
	phi227562	1176	NED	1	ACC	9	307-332	13	82,2
16	phi335539	1289	VIC	1	CCG	3	91-98	10	40,7
	phi308090	1292	FAM	4	AGC	4	210-223	11	82,2
	phi389203	857	VIC	6	AGC	4	301-313	11	92,2

<sup>1</sup> Trio de multiplexage

<sup>2</sup> Chromosome

<sup>a</sup> Echec à la lecture

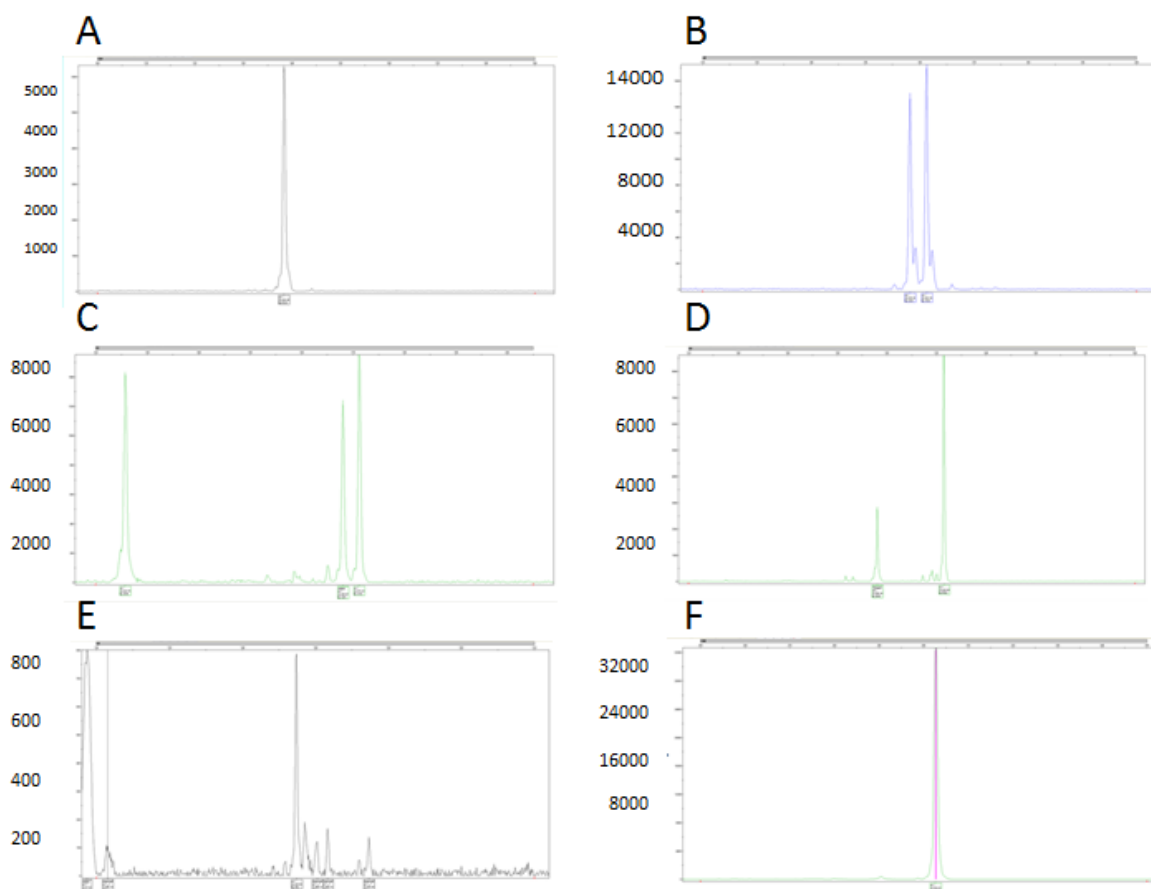
<sup>b</sup> Echec au *binnage*

#### **IV.4.3 Génotypage des microsatellites**

Tous les individus utilisés pour la génétique d'association ont été génotypés pour les microsatellites. Afin d'optimiser le nombre de manipulations et de réduire les coûts de séquençage, les microsatellites ont été multiplexés par groupes de trois. Les amorces *forward* sont couplées avec un fluorochrome bleu (FAM), vert (VIC) ou jaune (NED) (Tableau 7). Le microsatellite 40 a été traité en tant que microsatellite test et n'a pas été multiplexé (Tableau 7). Quatorze multiplexages ont été réalisés par lots de trois microsatellites attachés chacun à un fluorochrome différent (Tableau 7). Un dernier lot de trois microsatellites a été multiplexé avec un microsatellite couplé au marqueur bleu, et deux microsatellites couplés au marqueur vert (Tableau 7). L'amplification par PCR des microsatellites sur l'ensemble des individus (1491 au total) a été réalisée au Moulon.

Le génotypage par fluorescence a ensuite été sous-traité à la plateforme GENTYANE (Clermont Ferrand) qui nous a renvoyé des données brutes de pics de fluorescence. L'attribution des allèles en fonction des pics de fluorescence se fait en deux étapes : l'« *allele calling* » puis le « *binnage* ». La première étape, ou « *allele calling* » consiste à attribuer une valeur numérique de taille du microsatellite en fonction du pic de fluorescence. Les lectures

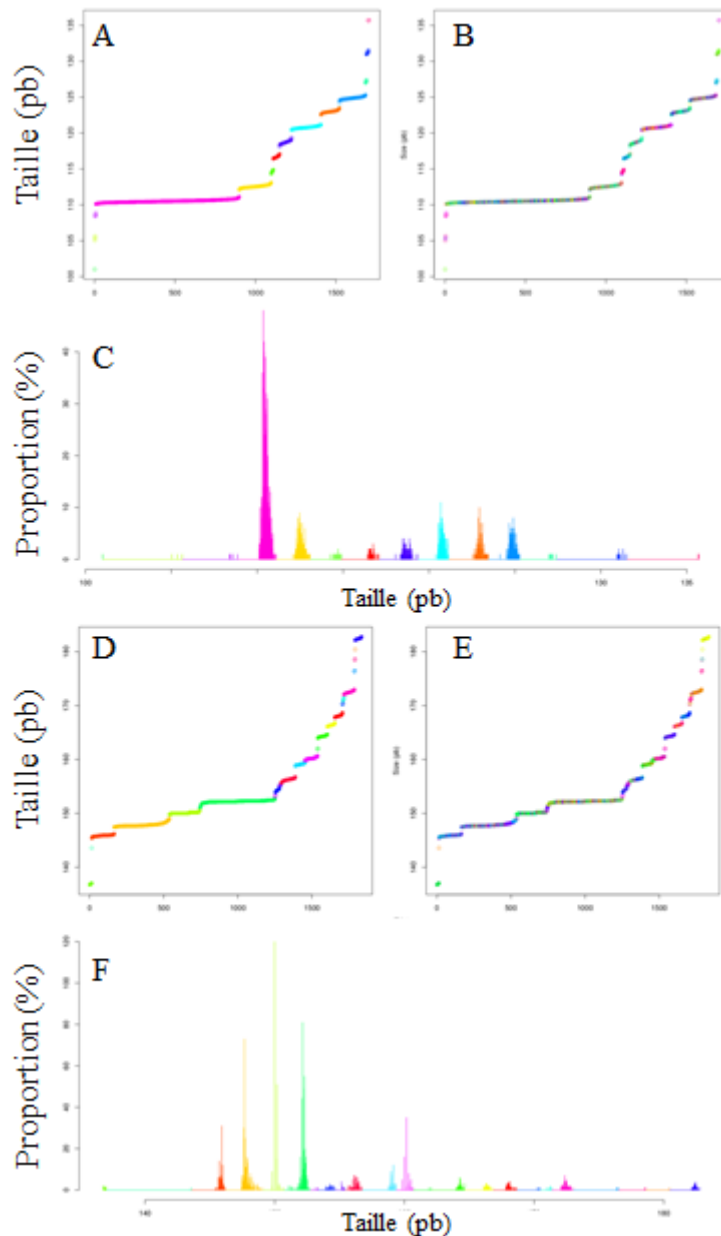
ont été effectuées au laboratoire du Moulon à l'aide du logiciel GeneMapper® Software Applied Biosystems®. Ce logiciel corrige automatiquement les erreurs dues à la saturation des pics et calcule une valeur numérique en fonction de la taille du pic. Cette lecture est semi-automatique, et l'utilisateur doit valider manuellement la valeur et la pertinence du pic (Figure 37). Un profil présentant un seul pic est considéré comme un homozygote sur le pic (Figure 37-A), et un profil présentant deux pics un hétérozygote (Figure 37-B). Certaines lectures peuvent être compliquées par des problèmes de différentes natures (qualité de l'ADN, redondance du génome, problèmes liés au multiplexage...). Un code qualité est assigné aux profils présentant des ambiguïtés (Figure 37). Nous détaillons ci-après ces codes. Certains individus peuvent présenter 3 pics, dans ce cas le code A est assigné au plus petit pic (Figure 37-C). Certains individus présentent deux pics dont l'un est plus faible que l'autre (entre un quart et un tiers de l'intensité du pic le plus élevé). Dans ces cas-là l'individu a été considéré comme hétérozygote mais l'allèle ayant un signal plus faible est signalé par la lettre B (Figure 37-D). Certains profils peuvent présenter une intensité de fluorescence très faible bien que les pics soient nets, le profil entier est alors assigné avec le code C (Figure 37-E). Quand l'intensité du signal de fluorescence sature le lecteur, la mesure exacte de la position du pic est rendue plus difficile. La position du pic est évaluée au milieu de la zone saturante et le pic est signalé par un code D (Figure 37-F). Certains profils présentent des pics très proches que le logiciel GeneMapper ne détecte pas comme deux pics, dans ce cas-là le seul pic détectable est assigné avec un code E auquel la taille du pic non cliquable est juxtaposée. Certains marqueurs présentent des allèles douteux, c'est-à-dire des profils conjoints pour lesquels des individus présentent plusieurs pics pouvant être des allèles potentiels. Ces allèles ont été codés par la lettre F. Deux lecteurs différents ont lu chaque profil, et en cas de désaccord un troisième lecteur a tranché. Parmi les 46 marqueurs, 5 présentaient des lectures très compliquées et ont été abandonnés (Tableau 7). Parmi les 41 microsatellites restants génotypés sur 1 491 individus, 61 131 profils ont pu être interprétés, dont 53 419 ne présentant pas de problèmes particuliers. Les 11 046 profils dont la lecture s'annonce problématique ont été éliminés des premières analyses.



**Figure 37 : Interprétation des pics de fluorescence.** A. Profil de l'individu C14\_10\_2\_1 pour le marqueur 215. L'individu présente un profil typique d'un homozygote. B. Profil de l'individu C14\_10\_2\_1 pour le marqueur 49. L'individu présente un profil typique d'un hétérozygote. C. Profil de l'individu C14\_10\_3\_1 pour le marqueur 32. Le second pic est notifié par le code A. D. Profil de l'individu C14\_10\_2\_1 pour le marqueur 19. Le pic de gauche est notifié par un code B. E. Profil de l'individu C14\_5\_26\_2 pour le marqueur 25. L'individu, homozygote, présente un pic de valeur similaire au bruit de fond, le profil est notifié par un code C. F. Profil de l'individu C14\_10\_11\_4 pour le marqueur 497. Le pic est saturé et donc notifié par un code D.

La deuxième étape du génotypage, le « *binning* », consiste à attribuer une valeur finie correspondant à l'identité de l'allèle à partir des valeurs continues obtenues après l'« *allele calling* ». Cette étape est primordiale et nécessite une attention particulière afin de ne pas faire d'erreur de génotypage (Guichoux *et al.* 2011). Le logiciel Autobin (Guichoux *et al.* 2011) a été utilisé afin d'attribuer automatiquement un allèle particulier en fonction des distributions des valeurs numériques. Les valeurs numériques sont classées par ordre croissant, un nouvel allèle est attribué automatiquement quand le pas entre deux valeurs est supérieur à 1, et dans certains cas pour un pas entre 0,3 et 1, selon la valeur numérique brute et la répartition des allèles attribués. Une vérification suivie d'une correction manuelle ont été réalisées en visualisant la distribution des valeurs brutes sur R (Figure 38). Pour cela, les valeurs brutes

ont été coloriées selon l'allèle attribué (Figure 38-A, C, D, F) ou selon la population de l'individu porteur de l'allèle (Figure 38-B, E). Deux microsatellites présentant des distributions continues, et dont le « *binnage* » s'est avéré infaisable, ont été abandonnés à cette étape (Tableau 7). Les microsatellites « *binnés* » avec succès présentent de 2 à 39 allèles, avec un pourcentage d'individus génotypés allant de 40,7 à 97% (Tableau 7). Le microsatellite 1289, présentant un taux de réussite de 40,7% a été enlevé des analyses suivantes. Le taux de réussite minimal a donc été réactualisé à 63%.



**Figure 38 : Distribution des valeurs brutes de tailles d'allèles.** Les allèles sont rangés par taille croissante. A, B, C. Microsatellite 80. D, E, F. Microsatellite 214. A, B, D, E. Les allèles attribués sont représentés chacun par une couleur différente. C, F. Les allèles sont coloriés selon la population d'origine de l'individu porteur de l'allèle.

#### **IV.4.4 Modélisation de la structuration neutre**

La structuration des populations entraîne un excès de faux positifs dans les dispositifs de génétique d'association, il est donc important de la prendre en compte. Notre dispositif présente deux niveaux de structuration : un niveau global correspondant à des différences dues à l'histoire démographique et évolutive des populations, et un niveau plus fin représentant les relations de parentés entre les individus. La structuration des populations a été définie à partir des données de microsatellites. Seuls les individus présentant plus de 30 % de locus génotypés ont été intégrés à l'analyse. Nous avons utilisé une méthode Bayésienne de classification implémentée dans le logiciel STRUCTURE (Pritchard *et al.* 2000a). Cette méthode basée sur des simulations Markov Chain Monte Carlo (MCMC) définit la proportion d'appartenance d'un individu à un nombre prédéfini de K clusters. Tous les K entre 2 et 11 ont été testés pour 10 répétitions indépendants de 500 000 itérations après une période de burn-in de 50 000 MCMC. STRUCTURE a été utilisé sur les 1 125 individus génotypés sur au moins 30 locus. L'algorithme LargeKGreedy du programme CLUMPP (Jakobsson & Rosenberg 2007) a été utilisé afin de combiner l'information des 10 répétitions pour chaque valeur de K. La visualisation des données de structure a été effectuée grâce au logiciel DISTRUCT (Rosenberg 2004) (Figure 39). La méthode d'Evanno permet de comparer le gain de vraisemblance  $\Delta K$  entre des modèles avec des K consécutifs (Evanno *et al.* 2005). La valeur avec le  $\Delta K$  le plus élevé est considérée comme le meilleur nombre de clusters reflétant la structure des populations. La méthode d'Evanno indique que le meilleur nombre de clusters est 2, mais dévoile un gain de vraisemblance non négligeable pour K=5 (Figure 39). Pour K=2, les individus ségrègent par sous-espèce *parviglumis* et *mexicana* (Figure 40). Lorsque 3 clusters sont utilisés, le cluster *mexicana* reste entier tandis que le compartiment *parviglumis* est divisé en deux blocs suivant le gradient d'origine de l'individu (Figure 40). Lorsque 5 clusters sont utilisés, les *mexicana* sont divisés en deux compartiments reflétant le gradient d'origine. Les *parviglumis* sont divisés en trois clusters dont l'interprétation biologique ou géographique est difficile (Figure 40). Lorsque K=11 est utilisé, soit le nombre de populations d'origine, chaque population est globalement représentée par un cluster. Dans les cas K=3, K=5 et K=11, les populations présentent un patron qui leur est propre (Figure 40). Ce patron a servi d'outil diagnostique pour repérer 43 individus faussement attribués à une population, suite à des erreurs de transplantations dans le champ ou des erreurs d'étiquetage lors des prélèvements. Pour prendre en compte la structure des populations, nous allons incorporer la matrice Q d'appartenance des individus aux 5 clusters définis à K=5, en

effet fixe dans le modèle d'analyse de variance.

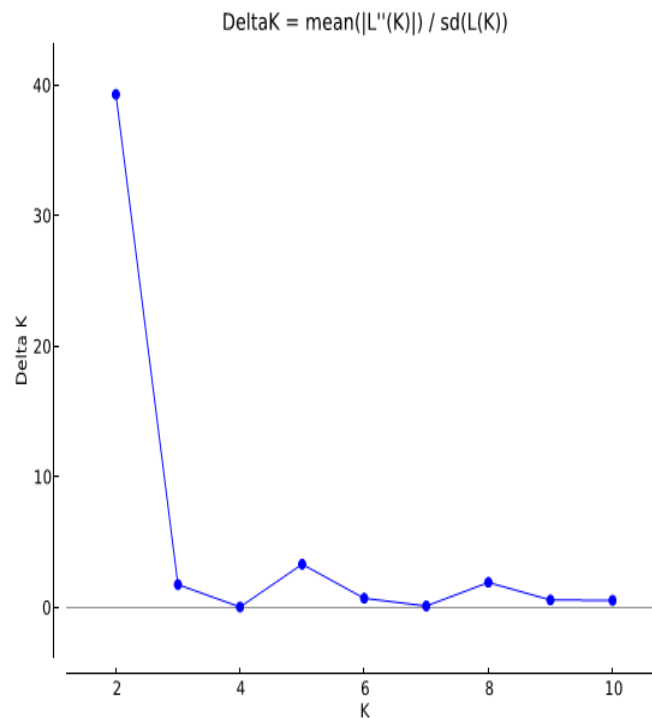
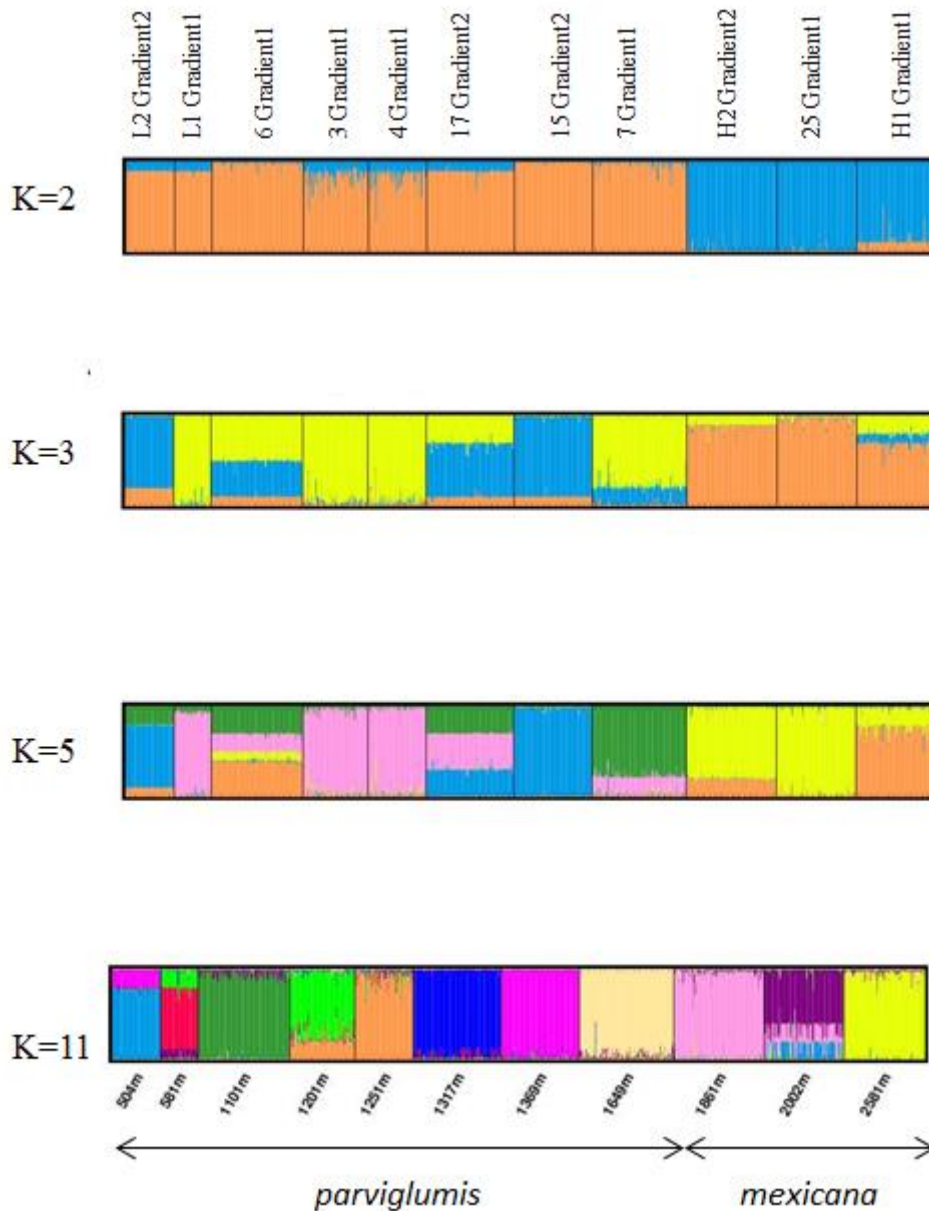


Figure 39 : Gain de vraisemblance entre modèles avec K successifs.

Les individus issus d'une plante-mère sont au minimum demi-frères. Nous ne disposons pas de l'identité des pères, ce qui nous empêche de détecter les pleins-frères, ou d'autres demi-frères venant du même père mais de mères différentes. Ces niveaux d'apparentement ainsi que d'autres plus éloignés peuvent être estimés au moyen d'une matrice d'apparentement. La structuration fine des individus sera donc prise en compte en effet aléatoire par une matrice d'apparentement (Yu *et al.* 2006). Cette matrice sera calculée à partir des données microsatellites en utilisant le programme SPAGeDI (Hardy & Vekemans 2002). Ce programme permet d'estimer une matrice d'apparentement entre individus à partir de marqueurs co-dominants. Il est possible que la matrice d'apparentement suffise à prendre en compte la structure totale du dispositif (Zhao *et al.* 2007), dans ce cas la matrice de structure Q ne sera pas intégrée au modèle de génétique d'association.



**Figure 40 : Structuration des individus issus de onze populations de téosintes.** Chaque individu est représenté par une barre verticale. Les individus sont classés par population, elles-mêmes ordonnées par altitude croissante. Chaque cluster est représenté par une couleur différente. Ces couleurs sont indépendantes entre les différents K.

## IV.5 CONSTRUCTION DES MODELES DE GENETIQUE D'ASSOCIATION

Le phénotype d'une plante est dû à l'action combinée du génotype et des effets de l'environnement. Les modèles de génétique d'association permettent de décomposer cette part génétique et cette part environnementale en se basant sur un grand nombre d'observations. Nous allons tester les SNP candidats selon deux approches différentes : une approche plante à



plante, et une approche au niveau des plantes-mères. Quelle que soit l'approche sélectionnée, il sera primordial de prendre en compte la structure des populations dans les modèles. La structure sera prise en compte par un modèle inspiré de (Yu, Pressoir et al. 2006):

(MGA) Phenotype = Localité + Année + Localité(Bloc) + Localité :Année +  
Matrice de STRUCTURE + SNP + **Matrice d'apparentement** + **Résiduelle**

$$(MGA) \quad Y = X\theta + S\alpha + Qv + \mathbf{Zu} + \mathbf{e}$$

Avec  $Y$  le vecteur des observations phénotypiques,  $X\theta$  représente les effets fixes environnementaux (localité, année, bloc),  $S\alpha$  l'effet fixe du SNP testé,  $Qv$  l'effet fixe de la structuration des populations,  $\mathbf{Zu}$  représente l'effet aléatoire de l'apparentement et  $\mathbf{e}$  est un vecteur des effets résiduels. Les variances des effets aléatoires sont estimées par :

$$\text{Var}(u)=2KV_g$$

avec  $K$  la matrice d'apparentement,  $V_g$  la variance génétique ; et

$$\text{Var}(e)=RV_r$$

avec  $R$  une matrice diagonale indiquant le nombre de répétition de mesures phénotypiques pour un génotype et  $V_r$  la variance résiduelle.

L'alternative « plante-mère » consiste à chercher les associations au niveau des plantes-mères, en utilisant les moyennes phénotypiques corrigées pour les effets environnementaux décrites à la section IV.3, et en utilisant des inférences de génotype à partir des descendants des plantes-mères. Ce modèle, (MGA<sub>bis</sub>), dérive du modèle (MGA) :

$$(MGA_{bis}) \quad Y_{pm} = S_{pm}\alpha_{pm} + Q_{pm}v_{pm} + \mathbf{Z}_{pm}\mathbf{u}_{pm} + \mathbf{e}_{pm}$$

Les deux modèles sont composés à la fois d'effets fixes et d'effets aléatoires (en gras dans le modèle) et seront analysés avec le package ASReml-R de R qui estime conjointement ces deux sortes d'effets.

Les modèles (MAG) et (MAG<sub>bis</sub>) sont des modèles linéaires mixtes univariés qui testent les associations entre un SNP et un caractère phénotypique uniques. Les paramètres du modèle sont estimés pour chaque combinaison SNP-caractère. Les statistiques multivariées permettent de tester plusieurs phénotypes et SNP en même temps. Ceci est particulièrement intéressant quand les caractères sont corrélés comme dans notre dispositif (Figure 36). En

effet, ces statistiques permettent non seulement un gain de puissance, mais également de détecter une association entre un SNP qui affecte un phénotype parmi de nombreux phénotypes corrélés (Zhou & Stephens 2014). Les statistiques multivariées présentent cependant des limites computationnelles, et bien que le nombre de SNP testé peut se compter en centaine de milliers, le nombre de phénotypes analysés conjointement reste limité à une petite dizaine lorsqu'un effet aléatoire comme la matrice de kinship est déclaré (Zhou & Stephens 2014).

Ces deux types de modèles complémentaires nous permettront de mettre en évidence les associations significatives entre la variation des SNP et celle des phénotypes, une fois correction faite pour la structure du plan d'expérience (en prenant en compte les deux sites, les deux années, l'organisation en blocs), pour la structure génétique des populations d'origine (afin d'éviter la confusion entre une association réelle et un lien statistique dû à la dérive génétique et la structure des populations) et enfin pour l'apparementement entre individus (afin d'éliminer les associations qui seraient dues à une structure génétique entre apparentés issus de la même mère ou de mères différentes à l'intérieur des populations). Ces analyses seront réalisées prochainement.

#### **IV.6 DISCUSSION**

Dans ce dispositif de génétique d'association nous avons mesuré deux caractères liés à la date de floraison, sept caractères liés à l'architecture végétative des plantes, trois liés à l'architecture de l'inflorescence, trois liés à la morphologie des grains, deux liés à la coloration, et pour finir la densité en stomate. Des études de génétiques d'association sur les téosintes *parviglumis* et/ou *mexicana* ont pu mettre en évidence l'effet de gènes candidats sur des caractères liés à l'architecture végétative, à l'architecture de l'inflorescence, aux caractéristiques des grains (Weber *et al.* 2008), ainsi qu'à la pigmentation de la plante (Lauter *et al.* 2004). Nous n'avons pas encore effectué l'analyse de génétique d'association, et nous ne pouvons pas présenter des résultats biologiques concrets. Cependant, les caractères étudiés ici par les modèles M3 ou M4 présentent tous un effet population, et la plupart du temps un effet plante-mère (13 sur 17 caractères, Tableau 6). Ces deux effets représentent la part génétique et ces résultats sont encourageants pour la suite des analyses, même si le fait d'incorporer la structure des populations dans l'analyse risque de réduire la puissance des tests.

Les analyses ne sont pas terminées, et ce travail va être mené à terme dans les

semaines qui suivront le rendu de ce manuscrit. La construction du modèle est encore en cours de discussion. Il existe plusieurs solutions pour intégrer les effets environnementaux et la structure des populations.

Deux alternatives s'offrent à nous pour tester les SNP : un modèle plante à plante (MGA) et un modèle à partir des plantes-mères (MGAbis). Ce dernier modèle est fréquemment utilisé quand des lignées sont évaluées. Il apporte un gain de précision des estimations de valeurs phénotypiques car celles-ci sont estimées en moyennant les descendants avec prise en compte des effets environnementaux. Cependant dans notre cas, ce modèle s'accompagne d'une perte de précision importante au niveau des génotypes car les téosintes ne sont pas fixées et qu'il est difficile d'inférer le génotype d'une plante-mère à l'aide d'un nombre restreint de descendants (maximum 8). Le modèle (MGAbis) est computationnellement plus intéressant, mais alors que nous voulons tester seulement 270 SNP, le gain de calcul est négligeable par rapport à l'inférence des probabilités de génotype des plantes-mères.

La structure des populations peut être intégrée de différentes façons. L'analyse de la structuration des onze populations de téosinte donne des résultats cohérents avec la littérature (Figure 40). Un premier niveau de structure sépare nettement les deux sous-espèces (Figure 40,  $K=2$ ) comme reporté par Pyhäjärvi *et al.* (2013). Alors que Ross-Ibarra *et al.* (2009) reportent une structuration plus importante chez *mexicana* que chez *parviglumis*, nous observons 2 groupes au sein de *parviglumis* tandis que *mexicana* conserve son identité à un niveau de structure supérieur (Figure 40,  $K=3$ ). Nous avons cependant moins de populations de *mexicana* (3) que de *parviglumis* (8) dans le dispositif, et les populations de *parviglumis* proviennent de deux endroits géographiquement distants, la région de Jalisco et celle de Balsas, pour lesquels des différences génétiques ont déjà été observées (Hufford *et al.* 2012a; Moeller *et al.* 2007). Les deux groupes observés dans *parviglumis* correspondent assez bien à ces deux régions, et donc aux deux gradients (Figure 40,  $K=3$ ). Les populations 3 et 4 se comportent comme une unique population quand on regarde la structure globale (Figure 40,  $K=2$ ,  $K=3$ ,  $K=5$ ) et semblent admixées lorsque l'on s'intéresse à  $K=11$  (Figure 40). Ces deux populations sont proches d'un point de vue géographique (Introduction, Figure 21) et poussent dans des conditions environnementales similaires (Figure 30). Au vu de ces résultats, nous pouvons utiliser  $K=2$  ou intégrer un effet sous-espèce pour prendre en compte le plus gros niveau de structuration. Nous envisagerons également  $K=3$  et  $K=5$  qui semble en adéquation avec nos données (Figure 39). Les groupes obtenus avec  $K=5$  sont cependant

difficilement interprétables. Il est également possible d'utiliser  $K=11$  qui définit bien les populations (Figure 40), ou bien d'intégrer directement un effet population au modèle. La matrice d'apparentement prendra en compte l'apparentement au sein des populations. Il est également possible qu'elle prenne en compte la structure d'un niveau plus élevée, et que la déclaration d'un niveau de structure plus élevée (matrice de structure, population ou sous-espèce) ne soit alors plus nécessaire (Zhao *et al.* 2007).

La valeur génétique se décompose en trois termes : un effet additif propre à la moyenne des allèles du génotype, un effet de dominance dépendant de la combinaison des allèles à un locus hétérozygote, et un effet épistasique dû à l'interaction des allèles de gènes différents entre eux. Notre dispositif, composé de populations naturelles pour lesquelles les allèles ne sont pas forcément fixés, présente à la fois des homozygotes et des hétérozygotes. Comparer les effets des allèles selon leur forme homozygote ou hétérozygote va nous permettre de tester la dominance d'un allèle. Il sera également possible de tester l'épistasie entre locus, c'est-à-dire l'action combiné de plusieurs locus dans le cadre des statistiques multivariées.

Les graines utilisées dans le dispositif sont issues de populations naturelles poussant dans des environnements variés. Le dispositif est donc sujet à des effets maternels, c'est-à-dire que le génotype de la mère peut influencer sur le développement et/ou le phénotype des descendants (Donohue 2009). Les effets maternels dépendent de l'environnement dans lequel la mère a poussé, et non de l'environnement actuel de la plante évaluée (Wolf & Wade 2009). Ils peuvent donc biaiser les associations entre phénotype et génotype d'une plante car ils ne sont pas pris en compte dans les effets environnementaux du modèle. Il est possible pour réduire les effets maternels de semer des graines issues de plantes-mères dans un environnement commun (première génération), d'autoféconder ces individus et d'utiliser la descendance obtenue dans le dispositif d'association (seconde génération). Les associations statistiques entre la moyenne phénotypique des descendants de seconde génération et le génotype des plantes de première génération sont ensuite recherchées. Ceci permet également de se procurer des répétitions des mesures phénotypiques de la même plante via ses descendants. Les téosintes sont très complexes à autoféconder du fait du faible nombre de grains par épi et de la dispersion des épis sur la plante. De plus, elles présentent une forte dépression de consanguinité qui aurait mis en péril la suite du dispositif. Cette étape de production de semence n'a donc pas été réalisée. Dans ce dispositif la descendance d'un nombre limité de plantes échantillonnées a été utilisée. Dans le modèle MGAbis, les effets

maternels risquent particulièrement d'être confondus à l'effet SNP, tandis que dans le modèle MGA, plante à plante, ils peuvent être en partie absorbés dans la matrice d'apparentement qui prendra en compte les relations entre descendants de la même plante-mère.

Les analyses préliminaires indiquent la présence de fortes interactions « Génotype x Environnement ». Les polymorphismes responsables de ces effets sont difficilement détectés par les modèles de génétique d'association (Sasaki *et al.* 2015). Ceci est particulièrement vrai lorsque les locus impliqués de ces interactions présentent des séries alléliques (Sasaki *et al.* 2015), ce qui est courant chez les téosintes. Pour détecter les polymorphismes responsables de cette interaction Sasaki *et al.* (2015) proposent d'intégrer aux modèles de génétique d'association non pas les SNP en effet fixe mais plutôt la variance locale comparée à l'ensemble du génome. Cependant, notre faible densité en SNP limite notre marge de manœuvre.

En utilisant peu de SNP dans un dispositif de génétique d'association nous nous exposons à détecter une part limitée l'information. En effet, les téosintes sont des plantes allogames, avec de larges populations et présentent un faible déséquilibre de liaison. Le déséquilibre de liaison chez les populations traditionnelles du maïs est estimé à environ 1kb (Tenaillon *et al.* 2001) ; il doit être plus élevé que dans le compartiment sauvage. Pour disséquer entièrement les bases génétiques d'un caractère chez les maïs traditionnels, il faudrait aux alentours de 750 000 SNP (Flint-Garcia *et al.* 2003), et un nombre encore plus important chez les téosintes. Cependant, et justement parce que le déséquilibre de liaison est faible, l'approche gènes-candidats est particulièrement intéressante dans notre dispositif car un SNP détecté est supposément très proche du polymorphisme causal (Flint-Garcia *et al.* 2003). Le déséquilibre de liaison peut varier localement, en particulier du fait d'effets de la sélection et de l'hétérogénéité de la recombinaison. En effet, lorsqu'un allèle est sous sélection il entraîne les locus à proximité, ce qui augmente le déséquilibre de liaison de manière d'autant plus forte que le taux de recombinaison est faible. Cet effet peut être très variable car le taux de recombinaison est très hétérogène chez le maïs, variant d'un facteur 1000, et peut être particulièrement faible pour certaines régions géniques (Rodgers-Melnick *et al.* 2015)

## **REFERENCES BIBLIOGRAPHIQUES**

- Aranzana MJ, Kim S, Zhao K, *et al.* (2005) Genome-Wide Association Mapping in *Arabidopsis* identifies previously known flowering time and pathogen resistance genes. *PLoS Genet* **1**, e60.
- Atwell S, Huang YS, Vilhjálmsson BJ, *et al.* (2010) Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* **465**, 627-631.
- Beadle GW (1980) The ancestry of corn. *Scientific American* **242**, 112-119.
- Camus-Kulandaivelu L, Veyrieras J-B, Madur D, *et al.* (2006) Maize adaptation to temperate climate: relationship between population structure and polymorphism of *Dwarf8* gene. *Genetics* **172**, 2449-2463.
- Doebley J, Stec A (1991) Genetic analysis of the morphological differences between maize and teosinte. *Genetics* **129**, 285-295.
- Doebley J, Stec A (1993) Inheritance of the morphological differences between maize and teosinte: comparison of results for two F2 populations. *Genetics* **134**, 559-570.
- Doebley J, Stec A, Hubbard L (1997) The evolution of apical dominance in maize. *Nature* **386**, 485-488.
- Doebley J, Stec A, Wendel J, Edwards M (1990) Genetic and morphological analysis of a maize teosinte F2 population - implications for the origin of maize *Proceedings of the National Academy of Sciences of the United States of America* **87**, 9888-9892.
- Donohue K (2009) Completing the cycle: maternal effects as the missing link in plant life histories. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **364**, 1059-1074.
- Dorweiler J, Stec A, Kermicle J, Doebley J (1993) *Teosinte glume architecture 1*: a genetic locus controlling a key step in maize evolution. *Science* **262**, 233-235.
- Ellegren H (2000) Microsatellite mutations in the germline: implications for evolutionary inference. *Trends in Genetics* **16**, 551-558.
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**, 2611-2620.
- Excoffier L, Hamilton G (2003) Comment on "Genetic structure of human populations.". *Science* **300**, 1877.
- Flint-Garcia SA, Thornsberry JM, IV B (2003) Structure of linkage disequilibrium in plants\*. *Annual review of plant biology* **54**, 357-374.
- Guichoux E, Lagache L, Wagner S, *et al.* (2011) Current trends in microsatellite genotyping. *Molecular Ecology Resources* **11**, 591-611.
- Hardy OJ, Vekemans X (2002) SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes* **2**, 618-620.
- Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23**, 1801-1806.
- Lauter N, Gustus C, Westerbergh A, Doebley J (2004) The inheritance and evolution of leaf pigmentation and pubescence in teosinte. *Genetics* **167**, 1949-1959.
- Liu N, Chen L, Wang S, Oh C, Zhao H (2005) Comparison of single-nucleotide polymorphisms and microsatellites in inference of population structure. *BMC Genetics* **6**, 26.
- Moeller DA, Tenailon MI, Tiffin P (2007) Population structure and its effects on patterns of nucleotide polymorphism in teosinte (*Zea mays* ssp. *parviglumis*). *Genetics* **176**, 1799-1809.
- Moles AT, Westoby M (2004) Seedling survival and seed size: a synthesis of the literature. *Journal of Ecology* **92**, 372-383.
- Price AL, Patterson NJ, Plenge RM, *et al.* (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics* **38**, 904-909.
- Pritchard JK, Stephens M, Donnelly P (2000a) Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-959.
- Pritchard JK, Stephens M, Rosenberg N, Donnelly P (2000b) Association mapping in structured populations. *American Journal of Human Genetics* **67**, 170-181.
- Pyhäjärvi T, Hufford MB, Mezouk S, Ross-Ibarra J (2013) Complex patterns of local adaptation in teosinte. *Genome biology and evolution* **5**, 1594-1609.
- Remington DL, Thornsberry JM, Matsuoka Y, *et al.* (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 11661-11666.

- States of America* **98**, 11479-11484.
- Rincent R, Moreau L, Monod H, *et al.* (2014) Recovering power in association mapping panels with variable levels of linkage disequilibrium. *Genetics* **197**, 375-387.
- Rodgers-Melnick E, Bradbury PJ, Elshire RJ, *et al.* (2015) Recombination in diverse maize is stable, predictable, and associated with genetic load. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 3823-3828.
- Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes* **4**, 137-138.
- Rosenberg NA, Li LM, Ward R, Pritchard JK (2003) Informativeness of Genetic Markers for Inference of Ancestry\*. *The American Journal of Human Genetics* **73**, 1402-1422.
- Ross-Ibarra J, Tenaillon MI, Gaut BS (2009) Historical divergence and gene flow in the genus *Zea*. *Genetics* **181**, 1399-1413.
- Sasaki E, Zhang P, Atwell S, Meng D, Nordborg M (2015) "Missing" G x E Variation Controls Flowering Time in *Arabidopsis thaliana*. *PLoS Genet* **11**, e1005597.
- Studer AJ, Doebley JF (2012) Evidence for a natural allelic series at the maize domestication locus *teosinte branched1*. *Genetics* **191**, 551-558.
- Tenaillon MI, Sawkins MC, Long AD, *et al.* (2001) Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays ssp mays* L.). *Proceedings of the National Academy of Sciences of the United States of America* **98**, 9161-9166.
- Vigouroux Y, Jaqueth JS, Matsuoka Y, *et al.* (2002) Rate and pattern of mutation at microstaellite loci in maize. *Molecular biology and evolution* **19**, 1251-1260.
- Wang H, Nussbaum-Wagler T, Li B, *et al.* (2005) The origin of the naked grains of maize. *Nature* **436**, 714-719.
- Weber A, Clark RM, Vaughn L, *et al.* (2007) Major regulatory genes in maize contribute to standing variation in teosinte (*Zea mays ssp parviglumis*). *Genetics* **177**, 2349-2359.
- Weber AL, Briggs WH, Rucker J, *et al.* (2008) The genetic architecture of complex traits in teosinte (*Zea mays ssp. parviglumis*): new evidence from association mapping. *Genetics* **180**, 1221-1232.
- Wolf JB, Wade MJ (2009) What are maternal effects (and what are they not)? *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **364**, 1107-1115.
- Wright SI, Bi IV, Schroeder SG, *et al.* (2005) The effects of artificial selection on the maize genome. *Science* **308**, 1310-1314.
- Yu J, Buckler ES (2006) Genetic association mapping and genome organization of maize. *Current Opinion in Biotechnology* **17**, 155-160.
- Yu J, Pressoir G, Briggs WH, *et al.* (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics* **38**, 203-208.
- Zhao K, Aranzana MJ, Kim S, *et al.* (2007) An *Arabidopsis* example of association mapping in structured samples. *PLoS Genet* **3**, e4.
- Zhou X, Stephens M (2014) Efficient algorithms for multivariate linear mixed models in genome-wide association studies. *Nature methods* **11**, 407.







## **V. DISCUSSION GENERALE**



## **V.1 LES TEOSINTES COMME MODELE D'ADAPTATION LOCALE.**

Les téosintes poussent dans des environnements très différents au sein du Mexique: certaines populations se trouvent sous des climats subtropicaux chauds et humides alors que d'autres sont situées sous des climats semi-tempérés, plus froids et secs (Hufford *et al.* 2012). Les téosintes sont donc soumises à de grands différentiels de pression de sélection. Les sous-espèces *parviglumis* et *mexicana* ont un effectif efficace important (de l'ordre de 120 000 à 160 000 ; Ross-Ibarra *et al.* 2009) et un niveau de polymorphisme élevé (Fukunaga *et al.* 2005). Elles sont allogames, et ont un cycle de vie court. Elles présentent donc un fort potentiel de réponse à la sélection. De plus leurs génomes sont d'une complexité moyenne par rapport à l'ensemble des Angiospermes ; elles offrent donc un complément intéressant par rapport aux études réalisées à partir des espèces du genre *Arabidopsis* qui présentent des génomes beaucoup plus simples. D'un point de vue pratique, les ressources développées et les connaissances obtenues chez le maïs (génome de référence, annotation des gènes, très nombreuses détections de QTL) facilitent les études génomiques et génétiques de la téosinte. Ceci est un véritable avantage pour étudier l'adaptation locale en population naturelles. Pyhäjärvi *et al.* (2013) ont par exemple mis à profit l'existence de la puce 50K pour mener une étude de grande ampleur sur l'adaptation locale chez les téosintes. Cette étude a mis en avant l'intensité de l'adaptation locale chez les téosintes, mais surtout l'importance potentielle de plusieurs inversions chromosomiques dans les processus adaptatifs.

Cependant les téosintes montrent une structuration importante qui coïncide avec l'altitude (analyse de structure, chapitre 3 ; Moeller *et al.* 2007; Pyhäjärvi *et al.* 2013). Ces effets démographiques peuvent induire une variation génétique de distribution similaire à celle engendrée par la sélection. Les ignorer peut conduire à une surestimation de la sélection et à la détection de nombreux faux positifs (Wright & Gaut 2004). La structuration génétique a été décrite et prise en compte par différentes mesures dans les travaux présentés dans ce manuscrit : par la conception de l'étude (séquençage de populations intermédiaires pour prendre en compte la différenciation entre les sous-espèces) ou par l'utilisation de méthodes incorporant cette structuration (BayeScan, BAYENV2, matrice de structure et/ou d'apparement dans le modèle de génétique d'association). Nous avons donc sciemment mis de côté de nombreux SNP impliqués dans la divergence entre *parviglumis* et *mexicana*. Il est toutefois possible que ces SNP soient réellement impliqués dans l'adaptation car les deux sous-espèces se sont spécialisées écologiquement, et qu'ils aient divergés sous l'action combinée de la sélection et de la limitation des flux de gènes. Nous avons utilisé une approche

gène candidat dans la suite des analyses. Devant le nombre restreint de polymorphismes que nous pouvions génotyper, nous avons choisi de nous limiter aux SNP porteurs des plus forts signaux de sélection.

## **V.2 REFLEXIONS SUR LES METHODES EMPLOYEES**

Nous présentons ici un ensemble d'analyses complémentaires pour étudier l'adaptation locale chez les téosintes. Dans les Chapitre 1 et 2, nous avons employé des méthodes de « reverse ecology » particulièrement intéressantes pour explorer des nouveaux polymorphismes impliqués dans l'adaptation locale (Tiffin & Ross-Ibarra 2014). Dans le chapitre 3, nous présentons un dispositif de génétique d'association qui nous permettra de relier nos SNP candidats à des variations phénotypiques.

Nous avons séquencé le génome complet de six populations dans le chapitre 1. Des puces à ADN permettant de génotyper plusieurs dizaines à centaines de milliers de SNP sont disponibles chez le maïs, mais il nous paraissait important de séquencer des génomes complets. En effet, ces puces, bien que très économiques, sont construites à partir d'un nombre réduit de génotypes de maïs et présentent donc un fort biais d'échantillonnage : le spectre de fréquence obtenu peut ne pas être représentatif de celui d'une population de téosinte. Ceci pose un problème pour les inférences démographiques et les scans génomiques (Albrechtsen *et al.* 2010). En outre, le séquençage de génome complet nous permet d'avoir accès à plus de polymorphismes non géniques (les régions géniques étant majoritaires pour la puce) et surtout à des polymorphismes spécifiques des téosintes non présents chez les maïs. De tels polymorphismes sont particulièrement importants pour l'étude de l'adaptation locale de la téosinte.

Pour pallier aux coûts conséquents du séquençage de génomes complets, nous avons séquencé les populations de téosinte en mélange. Le séquençage en mélange se présente comme une alternative intéressante au séquençage individuel, mais nécessite certaines précautions particulières pour limiter les biais (Schlötterer *et al.* 2014). Zhu *et al.* (2012) préconisent notamment de mélanger un grand nombre d'individus. Cependant ceci n'est pas toujours possibles car certaines populations naturelles sont de petites tailles. Dans le Chapitre 1, nous avons séquencé les populations extrêmes en mélange tout en gardant l'information individuelle au moyen d'étiquettes. Nous avons évalué l'impact de ce type de séquençage sur la détection des polymorphismes en comparant pour les 270 SNP candidats les fréquences alléliques obtenues sur les populations en mélange (Chapitre 1) avec les fréquences obtenues

par la technologie KASPar qui est une technologie très fiable (Chapitres 2 et 3). Pour le génotypage KASPar, les populations extrêmes, incorporées au dispositif de génétique d'association, ont été séquencées sur un nombre bien plus important d'individus (entre 128 et 143 individus provenant d'un même lot de plantes-mères) que les populations intermédiaires (20 individus). Pour les quatre populations extrêmes nous obtenons des corrélations entre 0,82 et 0,87 alors que ces corrélations ne sont que de 0,41 (IP1) et de 0,63 (IM1) pour les populations intermédiaires. Ceci peut être dû d'une part aux différences de profondeur de séquençage entre populations extrêmes et intermédiaires mais surtout à la différence de nombre d'individus pris en compte pour calculer les fréquences alléliques obtenues par KASPar. En effet, lorsqu'on effectue les corrélations dans les populations extrêmes en prenant seulement un descendant de chaque plante mère (20 plantes-mères, donc 20 individus) les valeurs obtenues sont de 0,52 à 0,72. Ceci souligne l'importance de la taille de l'échantillonnage sur l'estimation des fréquences alléliques d'une population. De plus, l'estimation des fréquences alléliques à partir de l'ensemble des individus disponibles est également biaisée en faveur de certains génotypes, car plusieurs individus proviennent d'une même plante-mère dans le dispositif de génétique d'association. Les allèles portés par les plantes-mère sont donc surreprésentés par rapport au reste de la population, et la prise en compte de l'apparentement des individus dans notre échantillonnage pourrait permettre d'augmenter encore les corrélations.

Enfin, nous avons également comparé le spectre de fréquences obtenu avec et sans l'information individuelle chez les populations extrêmes en séquençage complet (Chapitre 1). Les spectres présentent des profils assez similaires avec toutefois des différences surtout au niveau des classes de faibles fréquences. Ceci peut avoir un impact pour étudier les populations lorsque les méthodes employées se basent sur ces catégories de faibles fréquences. Les méthodes employées dans les Chapitres 1 et 2 se focalisent sur l'identification de SNP fortement différenciés entre les populations, elles sont peu sensibles à cette classe de polymorphismes. Nous avons conclu que le séquençage en mélange est une stratégie intéressante pour étudier l'adaptation locale tout en limitant les coûts. A partir de ces données, nous avons découvert des SNP et des régions candidates à l'adaptation locale (Chapitre 1 et 2).

Nous avons voulu valider 270 candidats par des études de clines et par une étude de génétique d'association. Parmi ces 270 SNP, nous avons choisi 93 SNP appartenant aux régions candidates détectées dans le Chapitre 1 qui viennent de recoupage entre plusieurs

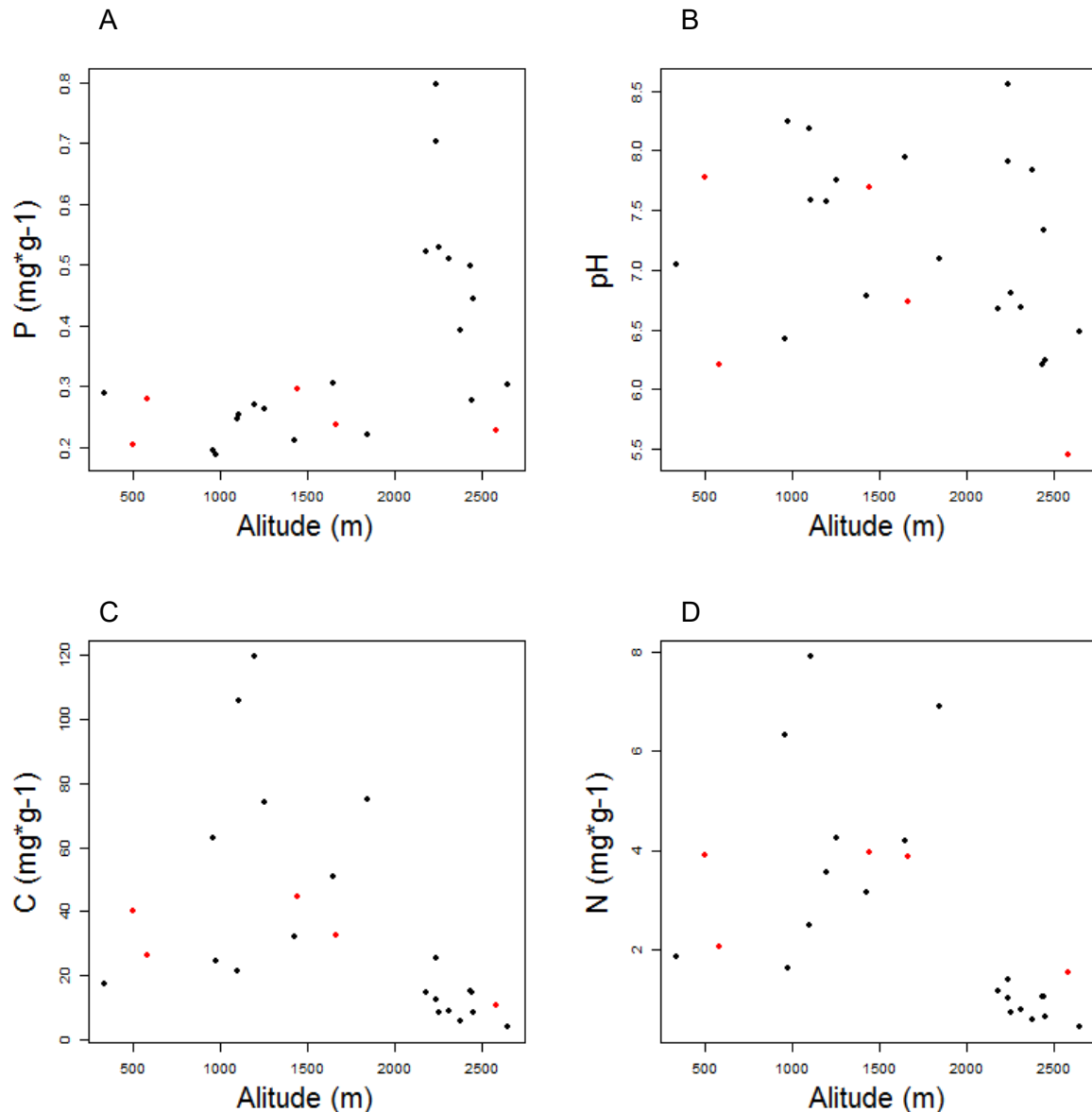
méthodes et sont conditionnés par une corrélation des fréquences alléliques avec des variables environnementales (SNP1), 97 SNP provenant uniquement de scan *Fst* le long du génome et (SNP2), et 80 SNP provenant de régions identifiées dans la littérature (SNP3). Cette combinaison de SNP de différentes origines nous permettra d'une part de confirmer le rôle adaptatif de nos SNP candidats (SNP1 et 2), et en utilisant SNP3 d'étudier le rôle de régions candidates identifiées dans d'autres études ainsi que d'évaluer la variation existante chez des téosintes à des locus que l'on sait très sélectionnés chez le maïs (comme les locus impliqués dans la domestication ou dans l'amélioration). Concernant SNP2, nous avons intégré des SNP provenant de scans génomiques de *Fst*, complémentaires aux analyses *Fst* effectuées dans le Chapitre 1. Ces scans combinent l'information de plusieurs SNP et servent à identifier des régions génomiques impliquées globalement dans l'adaptation locale comme les inversions. De plus, SNP2 a été constitué complètement "à l'aveugle", c'est à dire sans recouper les régions candidates avec la synthèse bibliographique présentée dans le chapitre 2. Concernant SNP3, en intégrant des SNP provenant de gènes impliqués dans la date de floraison, la tolérance à la sécheresse, la croissance de la plante ou encore la résistance à des maladies chez le maïs, l'idée était de tester la variabilité pré-existante dans le pool sauvage et son implication dans les variations phénotypiques. Il a été montré dans des travaux antérieurs que les gènes régulateurs importants chez le maïs (facteurs de transcription) ont des effets phénotypiques chez les téosintes, ce qui suggère que des variants de régulation ont pu être sélectionnés durant la domestication (Weber *et al.* 2007). Il a été aussi montré l'existence de séries alléliques chez les téosintes pour des gènes de domestication et d'amélioration du maïs (Studer & Doebley 2012). Découvrir un signal adaptatif ou une implication phénotypique de locus ayant un effet pour des caractères agronomiques importants chez le maïs tels que la tolérance à la sécheresse ou la composition du grain serait particulièrement intéressant en vue de considérer les téosintes comme un réservoir d'allèles sauvages pouvant être transférés au maïs.

Nous avons génotypé ces 270 SNP au sein de 37 populations réparties en deux gradients (17 et 21 populations pour les gradients 1 et 2 respectivement) et analysé les clines de fréquences avec des résultats pour le moment très préliminaires. Il est pour le moment un peu difficile d'interpréter les clines obtenus car nous avons très peu de recul sur la méthodologie employée. Cependant, la comparaison de l'enrichissement en SNP adaptatifs parmi les trois groupes de SNP (recoupement de méthodes, scan génomique, analyses bibliographiques) nous permettra de discuter de l'intérêt/la puissance de l'une ou l'autre des approches. Par ailleurs, il sera intéressant de réutiliser des analyses testant les corrélations

entre fréquences alléliques et variables environnementales (Analyses BAYENV) sur ce dispositif élargi pour confirmer par exemple que l'on retrouve bien les corrélations obtenues sur l'échantillon de 6 populations (SNP1) et d'explorer celles pour les autres jeux de SNP (SNP2 et 3). Notons qu'en plus des variables climatiques, nous disposons également des mesures de sol pour de nombreuses populations (25 sur 37) qui pourront être utilisées. L'analyse de ces variables édaphiques montre une variation importante de la composition du sol, en particulier concernant la composition en phosphate et richesse en éléments azotés et carbonés (Figure 41) et constitue une piste intéressante au vu des résultats obtenus dans le Chapitre 1.

Pour ce qui concerne la validation par génétique d'association, nous avons fait un pari relativement osé car la plupart des SNP candidats (SNP1 et 2) ont été définis à partir d'approches de *reverse ecology*, pour lesquelles nous n'avons pas de preuves de l'implication des polymorphismes dans l'adaptation présente (mais passée) et pas non plus d'*a priori* phénotypiques. Au vu des régions candidates obtenues qui sont proches de QTL de pilosité ou de gènes impliqués dans des caractères édaphiques, il est dommage que nous n'ayons pas mesuré des caractères phénotypiques qui s'avèrent potentiellement intéressants comme la pilosité des plantes ou bien des caractéristiques de l'appareil racinaire, voire des communautés microbiennes du sol (associées aux racines). La mise en œuvre prochaine des modèles de génétique d'association décrits dans le Chapitre 3 nous dira si notre approche a bien permis d'identifier des SNP adaptatifs et si les SNP détectés sur des critères de sélection passée ont un impact sur la variation phénotypique actuelle. Les polymorphismes neutres obtenus avec les microsatellites pour estimer la structure des populations peuvent être utilisés en témoins négatifs : on s'attend à une absence de lien avec les phenotypes.





**Figure 41 : Composition du sol en fonction de l'altitude.** A Concentration en phosphate, B. pH, C. Concentration en carbone. D Composition en azote. Les populations indiquées en rouge correspondent aux populations L1, L2, IM1, IP1 et H2 séquencées dans le Chapitre 1. Nous n'avons pas les données pour la population H1.

### **V.3 L'ADAPTATION A L'ALTITUDE CHEZ LA TEOSINTE**

Les populations de téosinte poussent dans des environnements extrêmement divers. Le fait que les populations de basse altitude L1 et L2 n'aient pas fleuri à Celaya en 2013, un site de moyenne altitude, est révélateur de l'adaptation locale des *parviglumis* de basse altitude.

Nos résultats du Chapitre 1 et 2 nous ont permis de souligner le rôle adaptatif de deux inversions déjà reportées dans la littérature (Fang *et al.* 2012; Pyhäjärvi *et al.* 2013), l'une située sur le chromosome 1 (Chapitre 1) et l'autre située sur le chromosome 9 (Chapitre

2). Ces inversions présentent un signal adaptatif important, mais il est difficile de déterminer les cibles réelles de la sélection car ces régions présentent un très fort déséquilibre de liaison. Nous avons recoupé nos résultats du Chapitre 1 avec une analyse bibliographique afin de relier ces candidats avec des voies métaboliques ou des caractères fonctionnels. Nous avons répertorié des études présentant des dispositifs de détection de QTL ainsi que des analyses de génétiques d'association (Chapitre 2). Nos régions candidates co-localisent cependant très majoritairement avec des QTL (Chapitre 1) et non pas avec des gènes ou polymorphismes candidats détectés dans des études de génétique d'association. La résolution des détections de QTL est généralement faible, résultant en des régions larges pouvant contenir de nombreux gènes. Ainsi la co-localisation avec nos SNP candidats peut être fortuite. Nous avons cependant localisé des régions candidates dans des régions restreintes/gènes candidats aux QTL *alm2* et *macrohairless1*. Ce résultat ainsi que le fait que les recouvrements concernent la plupart du temps les mêmes catégories de QTL soutiennent l'importance des variables édaphiques, notamment en ce qui concerne la tolérance à l'aluminium ou l'utilisation du phosphate, ressource limitée dans les sols volcaniques où pousse *mexicana*. Ceci révèle l'importance que pourrait avoir cette sous-espèce pour l'amélioration du maïs dans le cadre d'une agriculture raisonnée avec moins d'intrants organiques. Ces recouvrements ne constituent en aucun cas des validations fonctionnelles, mais nous guident pour l'interprétation biologique des résultats.

Les SNP candidats détectés au Chapitre 1 montrent un enrichissement en SNP géniques (dû principalement aux SNP détectés par les *Fst*). Ce résultat va à l'encontre de ceux présentés par Pyhäjärvi *et al.* (2013) qui détectent un enrichissement de SNP non-géniques impliqués dans l'adaptation. Dans leur étude, Pyhäjärvi *et al.* (2013) ont défini les SNP géniques par les régions transcrites des gènes. Lorsque l'on utilise la même définition, notre pourcentage de SNP géniques dans le lot commun aux six populations séquencées dans le Chapitre 1 est de 26%, alors que les SNP utilisés par Pyhäjärvi *et al.* (2013) sont en grande majorité géniques (67%). Les deux compartiments, génique et non-génique, semblent tous deux importants dans l'adaptation locale, et nécessitent d'être considérés dans les prochaines analyses.

Nous avons considéré les deux gradients comme des répétitions biologiques. Savoir si l'adaptation se fait par des chemins différents ou se fait selon les mêmes voies dans différentes situations indépendantes est une question passionnante, à laquelle il est cependant difficile de répondre avec notre matériel. En effet chez les téosintes, il y a eu un ou des

événement(s) de colonisation à l'altitude à une échelle géographique très restreinte avec potentiellement des flux de gènes récurrents (Buckler *et al.* 2006), contrairement à d'autres systèmes où des groupes génétiques ont été soumis à la même pression de sélection à différents endroits géographiques séparés par des barrières géographiques importantes tels que différents continents ou plus restreintes comme des grottes aquatiques isolées (Protas *et al.* 2006). Afin de détecter de la convergence, il est nécessaire de montrer que des phénotypes ou génotypes similaires sont apparus dans des groupes génétiques distincts soumis à la même pression de sélection. C'est le cas par exemple de l'adaptation à l'altitude chez l'homme dans les populations andines et asiatiques (Foll *et al.* 2014). C'est aussi le cas de l'adaptation à l'eau douce des populations d'épinoche *Gasterosteus aculeatus* initialement issues de la population atlantique en de nombreuses localités (Hohenlohe *et al.* 2010). L'histoire démographique des téosintes complique l'étude de la convergence adaptative à l'altitude. En effet, les sous-espèces *parviglumis* et *mexicana* auraient divergés il y a 60 000 ans au niveau de la région de Balsas (région située au milieu du gradient 1) (Fukunaga *et al.* 2005; Ross-Ibarra *et al.* 2009). La sous espèce *parviglumis* aurait colonisé les régions de basses altitudes à l'est et à l'ouest de la région de Balsas, dont la région de Jalisco où se situe une partie du gradient 2, tandis que *mexicana* aurait colonisé les régions de haute altitude puis se serait répandue le long de la ceinture volcanique (Buckler *et al.* 2006; Fukunaga *et al.* 2005). Selon ce modèle, les populations *parviglumis* du gradient 1, dont L1, seraient proches du pool de gènes ancestral, tandis que celles du gradient 2, dont L2, en seraient plus éloignées. Chez les *mexicana*, les populations du gradient 1 seraient plus proches également du pool de gènes ancestral tandis que les populations du gradient 2 dériveraient des populations du gradient 1. Un signal adaptatif commun aux deux gradients peut être expliqué par trois hypothèses évolutives différentes. Dans le scénario 1, l'apparition du polymorphisme adaptatif a lieu entre la divergence des sous-espèces et la colonisation des différents habitats (Figure 42). Dans ce cas, le polymorphisme adaptatif peut se maintenir dans les deux gradients si les pressions environnementales étaient et sont restées similaires dans les populations des deux gradients. Sous ce scénario, les gradients peuvent être considérés comme des répétitions biologiques en ce qui concerne les pressions de sélection mais cela ne permet pas de détecter la convergence. Dans le scénario 2 le polymorphisme adaptatif apparaît après la divergence entre *parviglumis* et *mexicana* et la colonisation par les deux sous-espèces de leurs aires de distributions actuelles (Figure 42). Le même polymorphisme est sélectionné indépendamment dans les deux gradients, c'est un véritable cas d'adaptation convergente. Dans le scénario 3 le polymorphisme adaptatif apparaît dans un des deux gradients, puis apparaît dans l'autre

gradient par migration (Figure 42). Dans ce scénario, il apparaît plus probable que la migration se soit produite chez *mexicana* que chez *parviglumis* car les populations sont plus proches géographiquement. Dans ce scénario, les pressions de sélection doivent être similaires dans les deux gradients pour que le polymorphisme se maintienne dans les deux gradients mais il ne s'agit pas d'un cas de convergence. Tous les scénarios *a priori* sont bien cohérents avec le fait de considérer les gradients comme répétitions biologiques mais seul le scénario 2 permet de conclure à la convergence. Il nous est donc difficile de quantifier l'importance de la convergence dans l'adaptation des téosintes à l'altitude.

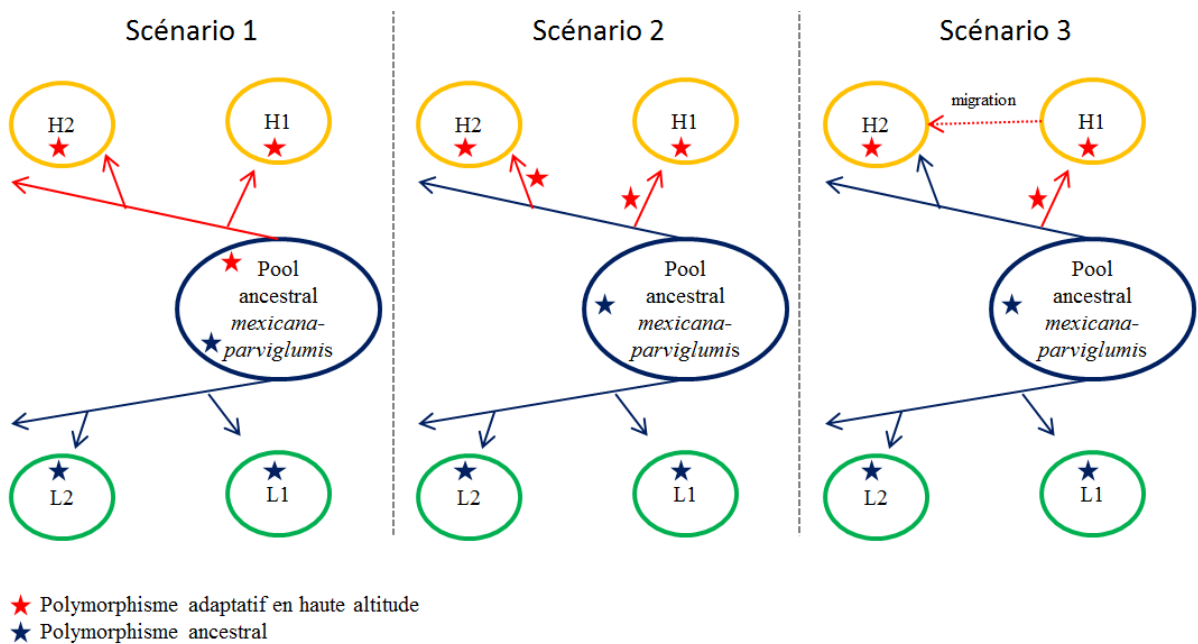


Figure 42 : Trois scénarios possibles de l'origine d'un polymorphisme adaptatif chez *mexicana*.

## V.4 PERSPECTIVES

Pour le moment les perspectives de mon projet de thèse se déclinent en trois axes :

(1) Etude du polymorphisme d'insertions-délétions d'éléments transposables. Notre étude de l'adaptation locale s'est focalisée sur les SNP. Cependant, d'autres polymorphismes peuvent être adaptatifs, particulièrement pour des espèces comme *Zea mays* dont le génome est composé à plus de 85% d'éléments transposables. Les éléments transposables peuvent en effet avoir de forts effets phénotypiques (Vitte *et al.* 2014). Cependant la détection de polymorphismes de présence/absence d'éléments transposables est complexe. Une étude sœur basée sur ce type de polymorphisme est actuellement menée dans le cadre du doctorat de Natalia Martinez.

(2) Etude des mécanismes génétiques de barrières aux flux de gènes entre sous-espèces. Nous avons éliminé des analyses les SNP impliqués dans la différenciation entre *parviglumis* et *mexicana*. L'analyse de ces SNP permettrait de mieux comprendre les bases génétiques de la différenciation entre les deux sous-espèces et d'identifier les régions responsables des barrières reproductives entre les deux compartiments.

(3) Suivi temporel des fréquences alléliques pour des SNP « neutres » et candidats. Dans cette étude nous avons mis en évidence des polymorphismes potentiellement adaptatifs que nous allons bientôt tenter d'associer à des variations phénotypiques. Nous sommes en train de concevoir un projet en partenariat avec l'institut d'écologie de la UNAM dont l'objectif est de suivre l'évolution de fréquence de ces polymorphismes adaptatifs dans des populations de téosintes évoluant dans un environnement différent de leur environnement d'origine. Ce suivi serait réalisé à l'aide d'une puce de génotypage basée sur les polymorphismes adaptatifs que nous avons identifiés (les 270 SNP que nous avons génotypés ainsi que d'autres polymorphismes révélés dans le chapitre 1 pour un total de 15k SNP incluant des SNP « neutres »). Cette thèse fournit des hypothèses sur l'évolution attendue de certains allèles dans des conditions de plus haute ou plus basse altitude par rapport à l'origine des populations. Le suivi des fréquences alléliques au cours du temps (plusieurs années) sur ces populations permettraient de mieux comprendre les mécanismes de l'adaptation locale. De façon similaire, le suivi de populations en conditions naturelles chez *Arabidopsis thaliana* a permis de confirmer le rôle de certains allèles dans l'adaptation locale (Fournier-Level *et al.* 2011). Un autre volet de ce projet serait de tester les valeurs sélectives de certains polymorphismes dans des conditions environnementales fixées. Par exemple, Hancock *et al.* (2011) ont prédit de manière statistique dans des populations d'*Arabidopsis thaliana* la valeur sélective d'allèles détectés dans une population soumise à des conditions environnementales données qui étaient similaires aux populations étudiées de manière statistique. Ceci serait particulièrement intéressant en vue d'introgession de certains allèles de la téosinte vers le maïs.

## **REFERENCES BIBLIOGRAPHIQUES**

- Albrechtsen A, Nielsen FC, Nielsen R (2010) Ascertainment biases in SNP chips affect measures of population divergence. *Molecular Biology and Evolution*. **27**, 2534-2547.
- Buckler ES, Goodman MM, Holtsford TP, Doebley JF, Sanchez J (2006) Phylogeography of the wild subspecies of *Zea mays*. *Maydica* **51**, 123-134.
- Fang Z, Pyhäjärvi T, Weber AL, *et al.* (2012) Megabase-scale inversion polymorphism in the wild ancestor of maize. *Genetics* **191**, 883-894.
- Foll M, Gaggiotti OE, Daub JT, Vatsiou A, Excoffier L (2014) Widespread signals of convergent adaptation to high altitude in Asia and America. *The American Journal of Human Genetics* **95**, 394-407.
- Fournier-Level A, Korte A, Cooper MD, *et al.* (2011) A map of local adaptation in *Arabidopsis thaliana*. *Science* **334**, 86-89.
- Fukunaga K, Hill J, Vigouroux Y, *et al.* (2005) Genetic diversity and population structure of teosinte. *Genetics* **169**, 2241-2254.
- Hancock AM, Brachi B, Faure N, *et al.* (2011) Adaptation to climate across the *Arabidopsis thaliana* genome. *Science* **334**, 83-86.
- Hohenlohe PA, Bassham S, Etter PD, *et al.* (2010) Population genomics of parallel adaptation in threespine sticklebacks using sequenced RAD tags. *Plos Genetics* **6**, e1000862.
- Hufford M, Martinez-Meyer E, Gaut BS, Eguiarte LE, Tenaillon MI (2012a) Inferences from the historical distribution of wild and domesticated maize provide ecological and evolutionary insight. *PLoS ONE* **7**, e47659.
- Moeller DA, Tenaillon MI, Tiffin P (2007) Population structure and its effects on patterns of nucleotide polymorphism in teosinte (*Zea mays* ssp. *parviglumis*). *Genetics* **176**, 1799-1809.
- Protas ME, Hersey C, Kochanek D, *et al.* (2006) Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism. *Nature genetics* **38**, 107-111.
- Pyhäjärvi T, Hufford MB, Mezouk S, Ross-Ibarra J (2013) Complex patterns of local adaptation in teosinte. *Genome Biology and Evolution* **5**, 1594-1609.
- Ross-Ibarra J, Tenaillon MI, Gaut BS (2009) Historical divergence and gene flow in the genus *Zea*. *Genetics* **181**, 1399-1413.
- Schlötterer C, Tobler R, Kofler R, Nolte V (2014) Sequencing pools of individuals [mdash] mining genome-wide polymorphism data without big funding. *Nature Reviews Genetics*.
- Studer AJ, Doebley JF (2012) Evidence for a natural allelic series at the maize domestication locus *teosinte branched1*. *Genetics* **191**, 551-558.
- Tiffin P, Ross-Ibarra J (2014) Advances and limits of using population genetics to understand local adaptation. *Trends in Ecology & Evolution* **29**, 673-680.
- Vitte C, Fustier MA, Alix K, Tenaillon MI (2014) The bright side of transposons in crop evolution. *Briefings in Functional Genomics* **13**, 276-295.
- Weber A, Clark RM, Vaughn L, *et al.* (2007) Major regulatory genes in maize contribute to standing variation in teosinte (*Zea mays* ssp. *parviglumis*). *Genetics* **177**, 2349-2359.
- Wright SI, Gaut BS (2004) Molecular population genetics and the search for adaptive evolution in plants. *Molecular Biology and Evolution* **22**, 506-519.
- Zhu Y, Bergland AO, González J, Petrov DA (2012) Empirical validation of pooled whole genome population re-sequencing in *Drosophila melanogaster*. *PLoS ONE* **7**, e41901.



## **ANNEXES**



**ANNEXE I.1 THE BRIGHT SIDE OF TRANSPOSONS IN CROP EVOLUTION**

# The bright side of transposons in crop evolution

Clémentine Vitte, Margaux-Alison Fustier, Karine Alix and Maud I. Tenaillon

## Abstract

The past decades have revealed an unexpected yet prominent role of so-called 'junk DNA' in the regulation of gene expression, thereby challenging our view of the mechanisms underlying phenotypic evolution. In particular, several mechanisms through which transposable elements (TEs) participate in functional genome diversity have been depicted, bringing to light the 'TEs bright side'. However, the relative contribution of those mechanisms and, more generally, the importance of TE-based polymorphisms on past and present phenotypic variation in crops species remain poorly understood. Here, we review current knowledge on both issues, and discuss how analyses of massively parallel sequencing data combined with statistical methodologies and functional validations will help unravelling the impact of TEs on crop evolution in a near future.

**Keywords:** *crop genomes; next generation sequencing; structural variation; functional validation; phenotypic variation; transposon insertion*

## INTRODUCTION

Transposable elements (TEs) were discovered 65 years ago in maize [1]. This has placed plants at the frontline for investigating the impact of TEs on genome structure and expression. Discovery of TEs and their genomic impact has also been leveraged by the sequencing efforts that have been particularly abundant for economically important plant species: first through cloning and sequencing of large genomic regions containing genes of agronomic interest [2–5], and more recently through whole genome sequencing (reviewed in [6]). While only a few genomes were sequenced in the early 2000s, development of so-called 'next generation sequencing' (NGS) and accompanying assembly methodologies has led to an explosion of the number of crop genomes sequenced, with 16 new crop genomes sequenced since 2012 [6].

While sequenced crop genomes cover distinct phyla, they exhibit common features: all plant genomes are polyploids or display traces of ancient polyploidy [7] and their largest genomic fraction is composed of TEs [8]. Among TEs, long terminal repeat (LTR) retrotransposons are often the most prevalent (Figure 1) because they are both more numerous and larger than DNA transposons. DNA TEs are also more difficult to annotate than LTR retrotransposons: DNA transposon annotation relies on sequence similarity to known transposase sequences, while LTR retrotransposons can also be identified based on detection of structural features (e.g. pair of LTRs). DNA transposons are therefore likely more abundant than depicted by current genome sequence analyses.

Crop genomes also differ in terms of genetic diversity [9], genome size [10], TE content (Figure 1)

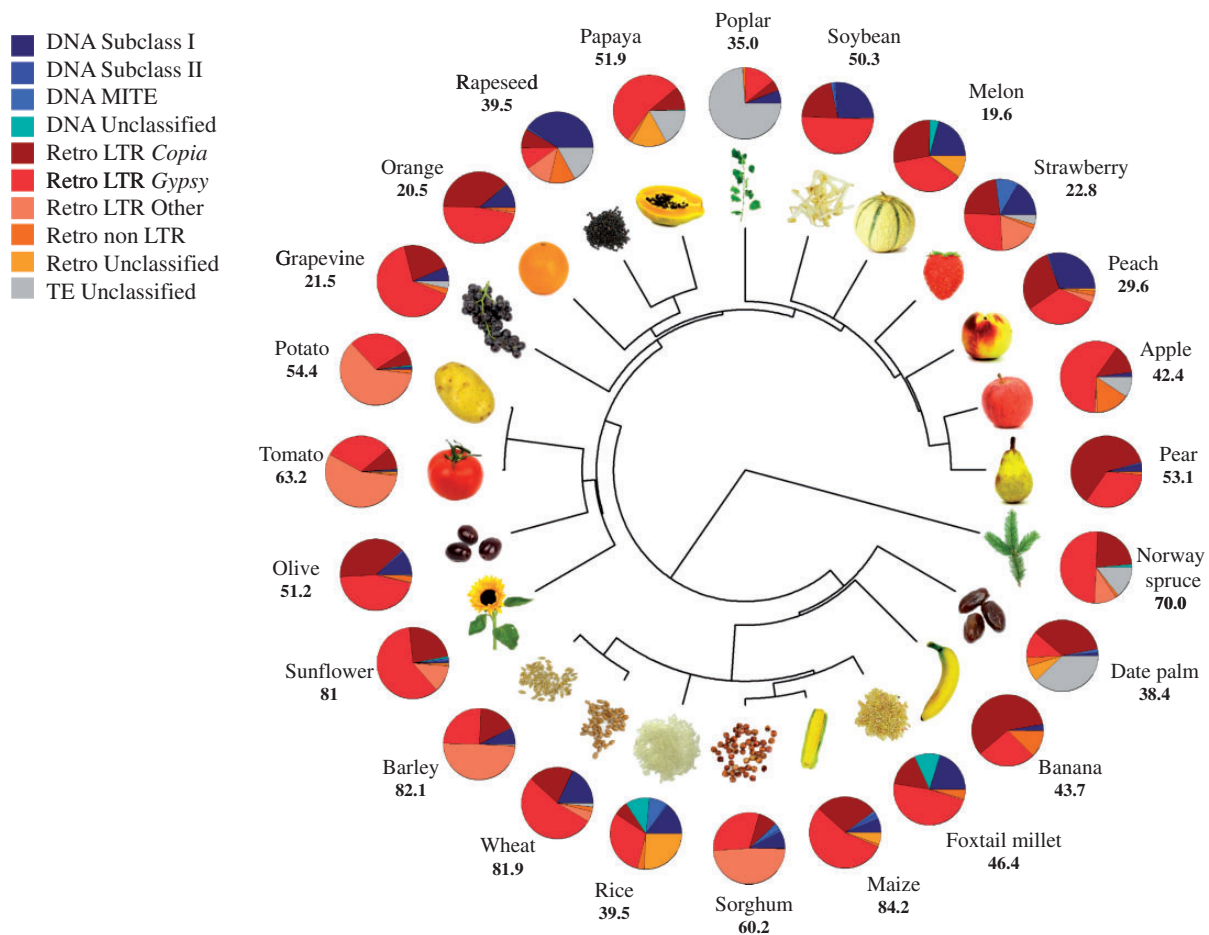
Corresponding author. M.I. Tenaillon, CNRS, UMR de Génétique Végétale, Ferme du Moulon, F-91190 Gif sur Yvette, France. Tel.: +33 1 69 33 23 34; Fax: +33 1 69 33 23 40; E-mail: tenaillon@moulon.inra.fr

**Clémentine Vitte**, PhD, is a CNRS researcher in maize genomics and epigenomics. Her research focuses in deciphering the interplay between genome structure, DNA methylation and gene expression, and how it has contributed to maize adaptation and genome evolution.

**Margaux-Alison Fustier** holds a master degree in Genetics and Biodiversity (University of Montpellier II) and is completing her PhD degree at University of Paris Sud XI on local adaptation of teosintes.

**Karine Alix**, PhD, is lecturer at AgroParisTech. Her research is dedicated to the evolution of *Brassica* species, and notably to the role of transposable elements and the dynamics of small non-coding RNAs in shaping the allopolyploid genome of *Brassica napus*.

**Maud I. Tenaillon** is a CNRS researcher, head of the team 'Genome dynamics and adaptation in crops' (<http://moulon.inra.fr/index.php/equipes/gear>). She investigates the genomic bases of adaptation in maize and its wild relatives.



**Figure 1:** Relative proportion of different TE types within 24 sequenced crop genomes. Phylogenetic relationships among species are shown (divergence time derived from <http://www.timetree.org/> and [13]). Numbers in bold indicate for each crop the percentage of genomic sequence that has been annotated as TE-derived DNA. Pie charts illustrate the proportions of various TE classes (DNA or RNA/retro elements), subclasses, orders. Relative contribution of TE types estimated as percentage of the TE-derived genomic fraction has been taken from genome publication: melon (*C. melo*, [14]), strawberry (*Fragaria vesca*, [15]), peach (*P. persica*, [16]), pear (*Pyrus bretschneideri*, [17]), apple (*Malus domestica*, [18]), norway spruce (*Picea abies*, [19]), date palm (*Phoenix dactylifera*, [20]), banana (*Musa acuminata*, [21]), wheat (*Triticum aestivum* 3B chromosome, [22]), barley (*Hordeum vulgare*, [23]), tomato (*Solanum lycopersicum*, [24]), orange (*Citrus sinensis*, [25]), rapeseed (*Brassica napus*, [26]), papaya (*Carica papaya*, [27]), sunflower (*Helianthus annuus*, [28]), sorghum [29] and foxtail millet (*Setaria italica*, [30]). Data from olive (*Olea europaea*) were kindly provided by A. Cavallini ([31]). Information from the potato genome was extracted from [24], and TE proportions for the soybean, poplar, grapevine, rice and maize genomes were found in [18].

and evolutionary potential [11]. Inspection of TE content of crop genomes in a phylogenetic context reveals no congruence in TE content even in closely related species—except perhaps in the genus *Solanum* (tomato and potato)—and highlights differences in success of TE types (Figure 1). For instance, *Gypsy* LTR retrotransposons have been very successful at invading the maize and grapevine genomes, while non-LTR retrotransposons are prevalent in sorghum, barley, potato and tomato (Figure 1). Of course, these patterns are highly dependent on the quality

of genome assemblies and annotation, the latter depending both on the age of the TEs found in a given genome (recent TEs are more easily detected than older ones) and the methods used for detection. In the next few years, genome assemblies will be improved. Use of a common method to annotate TEs from all crop species will then provide an unbiased picture of the trends observed. It is interesting to note that because TEs differ in their insertional properties (e.g. some are more often closely associated to genes than others [12]), the composition of TEs in a

genome may ultimately affect the potential for functional variation to be impacted by TEs.

When sequencing efforts include more than one genotype, an unsuspected level of structural variation is often found. For instance, comparison of a 1.2 Mb contiguous region of the *indica* and *japonica* subspecies of rice revealed that 13% of the sequence was not shared, the vast majority of the unshared regions being explained by differential TE insertions [32]. But this is particularly striking for maize, where as much as 76% of the genomic sequence between two inbred lines was unshared, mainly caused by TE presence/absence polymorphisms [33]. These variations likely predate maize domestication, but occurred within the past 4 million years. From both inter- and intra-specific analyses, it is now clear that TE differential insertions have shaped crop genomes. But whether this variation is a major contributor of phenotypic changes and, as such, has or could contribute to crop adaptation to human needs is still an open question. In this article, we review current knowledge about beneficial TE insertions in crop species and discuss how combining NGS, statistics and functional validation may be used to paint a more thorough picture of the impact of TE variation on crop phenotypic evolution.

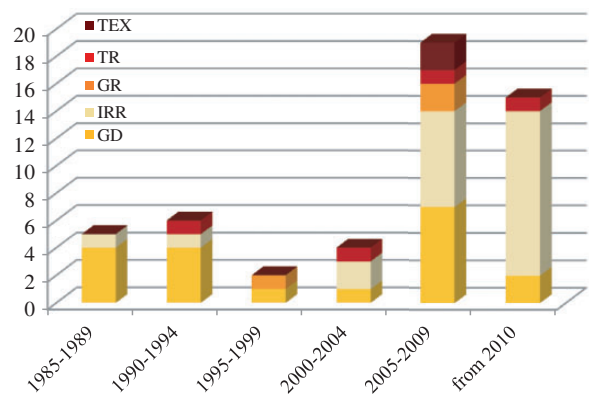
## BENEFICIAL TE INSERTIONS IN CROP GENOMES

TEs have long been recognized for their mutating potential, which generates genetic diversity upon which selection can act. Studies from the 1980s highlighted the adaptive potential of TEs in bacteria [34] and later in *Drosophila* in response to selection for bristle number [35]. In crops, active elements have been reported in maize and rice (reviewed in [36]) and in tobacco following stress activation (reviewed in [37]). While estimating transposition rate is often challenging [36], it is likely orders of magnitude higher than the nucleotide-base mutation rate, thus making TEs powerful agents of evolutionary changes [38].

Here, we list a total of 51 examples of TE-induced phenotypic changes associated with domestication or diversification of cultivated plants (Table 1). While in most cases TE causality has not been experimentally proven, these examples have provided first clues about the underlying mechanisms. We classified them according to the now well-established nomenclature of major TE-triggered mechanisms responsible for

phenotypic variation [39]. We have identified five categories of mechanisms across the selected examples: (i) gene inactivation (or severe dysfunction) caused by TE insertions in coding regions or introns, (ii) differential gene expression caused by TE insertions in or in the vicinity of regulatory regions, (iii) TE-mediated genomic rearrangements resulting in gene insertion, deletion or duplication, (iv) transduction-based regulation of gene expression and (v) a few cases of transposase exaptation responsible for modification of transcription factor abilities. The earliest documented cases date back to the 1980s, but most studies were published between 2005 and 2012, revealing that the effect of TE insertions on plant regulation and evolution is being discovered at a growing pace (Figure 2).

These five mechanisms involve both class I and class II TEs. Nevertheless, the two first categories (i and ii), which relate to insertions of TEs in genes or in their vicinity, are mainly caused by class II transposons (61% of the examples of Table 1). This is in accordance with the frequent association of DNA transposons [40,41] and notably miniature inverted repeats TEs (MITEs) [42–46] with coding sequences or regulatory regions of plant genes. Interestingly, more than 31% of the reviewed cases correspond to gene disruption, 56% of which occur in coding sequences. Such prevalence of coding regions may reflect a detection bias—it seems easier to identify TE insertions that cause a loss of gene function than to detect other types of TE-caused



**Figure 2:** Yearly distribution of 48 published studies depicting TE-triggered mechanisms responsible for phenotypic variation in crops (also listed in Table 1). GD, gene disruption (insertion in exon or intron); IRR, insertion in or close to regulatory regions; GR, genomic rearrangements; TR, transduction; TEX, transposase exaptation.

**Table 1:** TE insertions in crops and associated phenotypes classified by type of functional impact

	Mutation type	Gene	TE class, order, superfamily	Phenotype	Validation	References
Insertions/deletions in coding regions						
<i>Zea mays</i> ssp. <i>mays</i>	Insertion in exon	<i>Alcohol dehydrogenase I (Adhl)</i>	Class I, LTR retrotransposon, <i> copia</i> -like, <i>BsI</i>	Allyl alcohol-resistant pollen grain	Molecular and sequence-based characterization of insertion within exon 9	[112, 113]
<i>Z. mays</i> ssp. <i>mays</i>	Insertion in exon	Granule-bound glycosyl transferase, <i>waxy</i> locus <i>Wx</i>	Class II, subclass I TIR, <i>hAT</i> -like, <i>Ds</i>	Waxy kernels	Molecular characterization of insertion within cDNA, comparison of expression between alleles with/without insertion	[47]
<i>T. aestivum</i>	Insertion in coding sequence	High-molecular weight glutenin peptin ( <i>Glu-I</i> )	Class I, LTR retrotransposon <i> copia</i> -like	Glutenin seed storage	Molecular and sequence-based characterization of the insertion, restriction mapping	[48]
<i>Pisum sativum</i>	Insertion in a exon near the 3' end of the gene	Isoform of starch-branching enzyme (SBEI) ( <i>rugosus</i> locus)	Class II, subclass I TIR, <i>Ac/Ds</i> type	Wrinkled-seed—due to metabolic consequences on starch, lipid and protein biosynthesis in the seed	Molecular characterization of the insertion within cDNAs, linkage analysis of RFLP/wrinkled phenotype	[49]
<i>Oryza sativa</i>	Insertion in exon	Floral organ number1 (FON1)	Class II, subclass I TIR, <i>hAT</i> -like, <i>Ds</i> -like, <i>dTdk0</i>	Alteration of floral number and architecture of primary rachis branches	Genotyping, TD and AFLP analysis	[114]
<i>O. sativa</i>	Insertion in exon	<i>Thumbelina (Th)</i>	Class II, subclass I TIR, <i>hAT</i> -like, <i>nDart1</i>	Mutable dwarfism	TD	[115]
<i>O. sativa</i>	Insertion in exon	Granule-bound starch synthase (GBSS) <i>Wx</i>	Class I, LTR retrotransposon, <i>Dasheng</i> -like	Glutinous rice	Molecular and sequence-based characterization of the insertion, expression analysis using RT-PCR and Northern blot	[116]
<i>Glycine max</i>	Insertion in exon I	<i>GmphyA2</i> (one of paralogs encoding phytochrome A)	Class I, LTR retrotransposon <i>Ty1/ copia</i> -like, <i>SORE-I</i>	Photoperiod insensitivity	Molecular and sequence-based characterization of insertion within exon1 (genomic DNA), comparison of expression between alleles with/without insertion, comparison of NILs phenotype	[117, 118]
<i>Solanum tuberosum</i>	Excision from exon	<i>Flavonoid 3', 5'-hydroxylase</i>	Class II, MITE <i>Stowaway</i> -like	Skin color change obtained from somaclonal variant (from leaf protoplasts culture)	Anthocyanin content, sequencing, Southern blot (F)	[119]
Insertions in introns						
<i>Z. mays</i> ssp. <i>mays</i>	Insertion in intron	<i>Knotted 1</i>	Class II, subclass I TIR, <i>Mutator</i>	Leaf development	Molecular and sequence-based analysis of the insertions, correlation of phenotypic changes with structural variation, analysis of methylation	[120, 121]
<i>S. bicolor</i>	Insertion in the second intron	Y (member of the MYB-transcription factor family)	Class II, subclass I TIR, <i>CACTA</i> -like <i>Candystripe1</i>	Variiegated pericarp color	Molecular characterization of insertion, RFLP- and PCR-based detection of excision in red revertant	[122]
<i>M. domestica</i> (Rae lime, Seedless and Wellington Bloomless)	Insertion in intron 4 (Rae lime) or intron 6 (Spencer Seedless and Wellington Bloomless)	Apple homolog of <i>Arabidopsis pistillata</i> ( <i>MdPI</i> )	Class I, LTR retrotransposon, <i>dem1</i>	Parthenocarpic fruit production	Molecular and sequence-based analysis of the insertion, expression analysis using Northern blot	[123]

(continued)

Table 1: Continued

	Mutation type	Gene	TE class, order, superfamily	Phenotype	Validation	References
<i>G. max</i>	Insertion in intron	Flavanone 3-hydroxylase gene 1 ( <i>Wp</i> locus)	Class II, subclass I TIR, <i>CACTA</i> -like, <i>Tgm-Express!</i>	Coloration flower and protein content of seed	Molecular and sequence-based analysis of the insertion, association of insertion with expression and phenotype	[124]
<i>O. sativa</i>	Insertion in untranslated upstream sequence	Nuclear-coded essential chloroplast protease ( <i>OsCtpP5</i> )	Class II, subclass I TIR, <i>hAT</i> -like, <i>nDart!</i>	Leaf variegation	Structural and sequence-based characterization of germinal revertants	[115]
<i>G. max</i>	Insertion in intron	<i>Dihydroflavonol-4-reductase 2 (Dfr2)</i>	Class II, subclass I TIR, <i>CACTA</i> -like, <i>Tgm9</i>	Variegated flower color	Variegated phenotype, molecular and sequence-based analysis of insertion, analysis of anthocyanins and flavonols content, expression analysis	[125]
<i>Z. mays</i> ssp. <i>mays</i>	Insertion in two introns and one in intron/exon junctions	Granule-bound starch synthase <i>Wx</i>	Class I, LTR retrotransposon	Waxy kernels	Molecular characterization of insertion, analysis of gene expression and alternative splicing	[126]
Insertions (more rarely deletions) in gene regulatory regions						
<i>Z. mays</i> ssp. <i>mays</i>	Insertion 63 bp upstream of TATA box, affecting the steady-state level of mRNA and protein	Bronze 1, involved in production of anthocyanin pigments	Class II, subclass I TIR, <i>Ds</i>	Aleurone pigmentation	Molecular and sequence-based characterization of insertion. Northern blot, and enzyme assays compared between mutants and revertants	[127]
<i>Z. mays</i> ssp. <i>mays</i>	Excision of element from untranslated leader	<i>Alcohol dehydrogenase I (Aldh)</i>	Class II, subclass I TIR, <i>DsI</i>	Differential expression in pollen	Sequence analysis, phenotyping characterization of revertants, quantitative RNase protection assays	[128]
<i>Z. mays</i> ssp. <i>mays</i>	Insertion into the 5' regulatory region	<i>Y1</i> phytoene synthase	Class II, MITE, <i>Tourist</i> -like, <i>Ins2</i>	Endosperm color	Analysis of selection traces, association with phenotype (but not the only polymorphism associated)	[129]
<i>Vitis vinifera</i>	Insertion in the 5' flanking region	<i>VvmybA1</i> , regulator of anthocyanin biosynthesis	Class I, LTR retrotransposon <i>gypsy</i> -like, <i>Gret!</i>	Loss of pigmentation (anthocyanins)	Molecular and sequence-based characterization of insertion, association of TE presence/absence with expression and phenotype	[130]
<i>T. aestivum</i>	Insertion in promoter	<i>TaFT1/VRN3</i>	Class I, LTR retrotransposon	Flowering time	Molecular and sequence-based characterization of insertion, allele frequency in germplasm, association of TE presence/absence with gene expression and phenotype, expression and phenotype analysis of plant transformed with construct carrying insertion (F)	[50]
<i>S. bicolor</i>	Insertion in upstream region	Aluminium tolerance locus <i>AltSB</i>	Class II, MITE <i>Tourist</i> -like	Aluminium resistance	Characterization of the insertion, expression and phenotype analysis of plant transformed with construct carrying insertion (F)	[131]
<i>Z. mays</i> ssp. <i>mays</i>	Insertion ~70 kb upstream a gene ( <i>cis</i> -acting regulatory change)	<i>ZmRap2.7</i> , an <i>Ap2</i> -like transcription factor	Class II, MITE <i>Tourist</i>	Flowering time	Positional cloning and association mapping	[51]
<i>V. vinifera</i>	Insertion in a promoter	Putative regulator of anthocyanin synthesis ( <i>VvMYBA1</i> )	Class I, LTR retrotransposon <i>gypsy</i> -like	Fruit color	Characterization of the insertion, comparison of sequence, transient assays	[132]

(continued)

Table 1: Continued

	Mutation type	Gene	TE class, order, superfamily	Phenotype	Validation	References
<i>O. sativa</i>	Insertion upstream gene	Resistance gene <i>Pit</i>	Class I, LTR retrotransposon copia-like, <i>Renovator</i>	Disease resistance	Characterization of the insertion, construction of chimeric plants, analysis of pathogenicity	[33]
<i>C. melo</i>	Insertion near the promoter provokes its methylation	C2H2 zinc-finger domain transcription factor of the WIP protein subfamily	Class II, subclass I TIR, <i>hAT</i> -like, <i>Gyno-hAT</i>	Sex determination: carpel development, stamen repression	Molecular and sequence-based characterization of insertion, DNA methylation profile, TILLING, association of TE presence/absence with gene expression and phenotype	[53]
<i>Brassica oleracea</i>	Insertion in the upstream regulatory region	Purple ( <i>Pt</i> )	Class II, subclass I TIR, <i>Harbinger</i>	Purple coloration	Association mapping (S), plant transformation (F)	[34]
<i>V. vinifera</i>	Insertion in the promoter	<i>VvTFL1A</i>	Class II, subclass I <i>hAT</i> -like, <i>Hatvinel-nrm</i>	Inflorescence structure	Molecular characterization of the insertion, comparison of expression profile and presence of the insertions	[35]
<i>G. max</i>	Insertion in the promoter	<i>Dihydroflavonol-4-reductase 2 (DFR2)</i>	Class II, subclass I TIR, <i>CACTA</i> -like, <i>Tgm9</i>	Diluted (pale) flower color	Molecular and sequence-based characterization of insertion, analysis of anthocyanins and flavonols content, expression analysis, sequencing	[25]
<i>Triticum turgidum</i> L.	Insertion in 5' UTR, near regulatory sites	<i>Vrn-B1</i> , one of the genes controlling vernalization	Class I, LTR retrotransposon, <i>gypsy</i> -like	Expression of the gene without vernalization	QTL analysis, cloning and sequencing, association of TE presence/absence with gene expression	[36]
<i>Z. mays</i> ssp. <i>mays</i>	Insertion in regulatory region localized about 60 kb of the ORF	TCP family transcriptional regulator <i>teosinte branched1 (tb1)</i>	Class I, LTR retrotransposon copia-like, <i>hopscotch</i>	Increase of apical dominance	Test of selection, haplotype group analysis, phenotypic fine-mapping, transient assays of reporter gene under control of regulatory sequence with and without the insertion	[52]
<i>C. sinensis</i>	Insertion in upstream regulatory region leading to TE-derived new transcription start site	Ruby, a MYB transcriptional activator of anthocyanin production	Class I, LTR retrotransposon copia-like, <i>Tcs1</i>	Cold dependent induction of anthocyanin biosynthesis in fruit	Molecular and sequence-based characterization of insertion, association of insertion with expression and phenotypic, RACE-PCR based characterization of new transcript start site	[37]
<i>C. sinensis</i>	Insertion in upstream regulatory region	Ruby, a MYB transcriptional activator of anthocyanin production	Class I, LTR retrotransposons copia-like, <i>Tcs2</i>	Cold dependent induction of anthocyanin biosynthesis in fruit	Molecular and sequence-based characterization of insertion	[37]
<i>Z. mays</i> ssp. <i>mays</i>	Insertion near a regulatory sequence	Auxin-binding protein I	Class II, MITE, <i>Tourist</i> ; Class II, subclass I, <i>Ac</i> -like, <i>Pilgrim</i> ; Class I, solo-LTR, <i>Battara</i>	Unknown	Expression of reporter gene (luciferase) under control of regulatory sequence with and without the insertion	[38]
<i>O. sativa</i>	Insertion into the 5' UTR of the gene	<i>Chalcone isomerase (OsCHI)</i>	Class I, LTR retrotransposon, <i>Dasheng</i>	Golden hull and internode coloration	Complementation test (F)	[39]
<i>Brassica rapa</i>	Insertion in the ORF	TRANSPARENT TEST8 locus ( $\beta$ TT8)	Class II, subclass II Helitron	Yellow seed coloration	Construction of linkage map, functional complementation (F)	[40]
<i>Z. mays</i> ssp. <i>mays</i>	Tandem insertion at 289 nt 3' of the gene	Cytochrome P450-like	Class II, MITE, <i>Tourist</i> -like, <i>ZEmVI-9</i>	Flowering time	Association mapping (S)	[46]
<i>Z. mays</i> ssp. <i>mays</i>	Insertion in 3' untranslated region	<i>ZmGE2</i> (gene encoding a cytochrome P450 protein)	Class II, subclass I TIR, <i>Mutator</i>	Kernel oil content	Characterization of the insertion, association analysis of expression profile and presence of the insertion	[41]

(continued)

Table I: Continued

	Mutation type	Gene	TE class, order, superfamily	Phenotype	Validation	References
<i>Z. mays</i> ssp. <i>mays</i>	Insertion in promoter	ZmCCT (encoding a CTT domain protein)	Class II, subclass I TIR, CACTA-like	Flowering time	Genome-wide association study, transformation	[142]
Rearrangements <i>Z. mays</i> ssp. <i>mays</i>	TE-mediated chromosomal rearrangement	R-r complex transcription factor regulating the anthocyanin pigment pathway	Class II, subclass I TIR, CACTA-like, <i>doppia</i>	Modification of tissue-specificity of the duplicated genes	Molecular and sequence-based characterization of insertion and rearrangement breakpoints, expression and phenotypic analysis of aleurone layer bombarded with constructs with and without rearranged region, analysis of aleurone mRNA populations <i>in vivo</i>	[149]
<i>Triticum</i> and <i>Aegilops</i> species	Illegitimate DNA recombination events	Hardness ( <i>Ha</i> ) locus	Class I retrotransposons (different families)	Variation in grain hardness	Sequence-based characterization of the corresponding locus	[54]
<i>S. lycopersicum</i>	TE-mediated insertion of a 24.7 kb rearranged fragment inducing gene duplication and expression modification from new <i>cis</i> -acting elements	SUN locus ( <i>QD12</i> gene)	Class I LTR retrotransposon <i>copial</i> -like, <i>Rider</i>	Elongated fruit shape	Molecular and sequence-based characterization of insertion and rearrangement breakpoints, analysis of fruit shape index and gene expression in transformants (i) carrying deleted versions of the rearrangement, (ii) overexpressing the gene and (iii) underexpressing the gene through RNAi	[150]
Transduction <i>Z. mays</i> ssp. <i>mays</i>	Transduction, likely mediated through RNA-mediated transposition	Plasma membrane proton ATPase gene, <i>Mh1d</i>	Class I, LTR retrotransposon, <i>BsI</i>	Unknown	Molecular and sequence-based comparison of TE and complete gene	[143]
<i>Z. mays</i> ssp. <i>mays</i>	Possible emergence of a chimeric new gene from transduction of three genes fragment	Fragments of plasma membrane proton ATPase gene, Xylan endohydrolase gene, beta-1,3-glucanase gene	Class I, LTR retrotransposon, <i>BsI</i>	Unknown	Spatio-temporal expression (transcription and translation) of hybrid ORF combining the retrotransposon gag domain and the three gene fragments, reconstruction of sequence evolution, test of selection	[144, 145]
<i>Z. mays</i> ssp. <i>mays</i>	Fusion of the coding and regulatory sequences of two linked paralogous genes	Mybhomologous transcriptional regulator ( <i>pl</i> )	Class II, subclass I TIR, <i>hAT</i> -like, <i>Ac</i>	Kernel color and cob glume	Molecular characterization of the transposition, comparison of expression analysis and phenotype	[146]
Transposase exaptation <i>H. vulgare</i> , <i>O. sativa</i> , <i>T. aestivum</i>	Transposase exaptation	Gary <i>hAT</i> -transposase like	Class II, subclass I TIR, <i>hAT</i> -like, <i>gary</i>	Unknown	Comparison of gary genes and transposase at the structural and sequence levels, phylogeny	[147]
<i>Z. mays</i> ssp. <i>mays</i>	Transposase exaptation and duplication of gene	CYP72A27-Zm gene duplicated from the monooxygenase ( <i>P450</i> ) gene	Class II, subclass II Helitron	Unknown	Characterization of the insertion at the sequence level, reconstruction of sequence evolution	[148]

(continued)



Table 1: Continued

	Mutation type	Gene	TE class, order, superfamily	Phenotype	Validation	References
Unstable <i>Z. mays</i> ssp. <i>mays</i>	Insertion in intron	Knotted ( <i>Knl1</i> )	Class II, subclass I TIR, <i>Ds2</i>	Protrusions or knots along the lateral veins of the first leaf	Molecular and sequence-based characterization of insertion, cosegregation of phenotype with insertion, analysis of penetrance	[15]
Unclassified <i>Z. mays</i> ssp. <i>mays</i>	Insertion in 5' UTR inducing a duplication of TATA box	Alcohol dehydrogenase ( <i>Adhl</i> )	Class II, subclass I TIR, <i>Mutator</i>	Conditional anaerobic lethal phenotype	Molecular characterization of insertion and TATA box duplication, restriction pattern distinguishing this allele from seven other <i>Adhl</i> alleles, genetic instability of mutation (from single-cell aleurone layer stained for ADH activity)	[152, 153]
<i>S. italica</i> ssp. <i>italica</i>	Insertion at five sites (intron and exon)	Granule-bound starch synthase I	Class I (exons and intron 12) and Class II (intron 1)	Waxy endosperm (little or no starch amylose)	Sequence analysis, cosegregation of phenotype with insertion	[154]
<i>O. sativa</i>	Cytosine methylation of flanking sequences	Several	Class I, LTR retrotransposon, <i>Dasheng</i>	Unknown	Transposon methylation display, RNA analysis, correlative data	[155]

mutations such as TE-mediated local genome rearrangements requiring sequencing of large genomic regions and for which the demonstration of the causative effect remains challenging. This is particularly clear from the observation that older studies in Table 1 (eight papers published between 1985 and 1994) report the characterization of TE insertions resulting in direct gene disruption. Several of these studies revealed TE insertions in genes underlying agronomic traits such as seed content of amylose (maize, [47]), high-molecular weight glutenin (wheat, [48]) or starch (pea, [49]), which were cloned following fine mapping or genetic linkage approaches.

The predominant TE-mediated mechanism involved in crop phenotypic variation is related to modifications in the regulatory regions of plant genes (45%). This mechanism has been discovered more recently than gene disruption (most studies postdate 2006), which is likely due to parallel improvement of methodologies that allow positional cloning, association mapping in complex traits and the analysis of gene mutations *via* tilling or plant transformation. With this technological development, more complex traits such as flowering time have been analyzed (wheat, [50]; maize, [46,51]) and highlighted the more subtle impact of TEs on quantitative variation *via* modulation of gene expression. Improvement of sequencing also provided the means to identify regulatory sites located tens of kilobases upstream to the target gene, as exemplified by the textbook example of 'teosinte branched 1' (*tb1*), where insertion of a LTR retrotransposon *Hopscotch* element ~60 kb upstream of the gene [52] induces gene overexpression, leading to apical dominance in maize. Such studies have also brought considerable insights into the epigenetic impact of TE insertions in crops. For instance, transition from male to female flowers in melon is caused by DNA methylation spreading of the class II *Gyno-hAT* transposon to the promoter of the *CmWIP1* transcription factor gene [53].

The three other categories of TE-mediated mechanisms (iii, iv and v) have been reported more rarely. The complexity of genome rearrangements mediated by TEs has nevertheless been well illustrated through comparative genomics. For instance, comparative analysis of the *Hardness* locus sequenced for seven *Triticum* and *Aegilops* species demonstrated the role of various retroelements in gene loss and the further association of deleted haplotypes to trait quality in polyploid wheats [54].

With the recent boom of NGS-based technologies allowing the identification of TE polymorphisms (see below) together with the transcriptome and epigenome characterization, the number of TE-mediated phenotypic changes is expected to greatly increase in the coming years. Hence, NGS-based technologies will soon offer a less biased view of the relative contribution of the different types of mechanisms through which TEs have shaped crop evolution.

## USING NGS TO DETECT TE INSERTION POLYMORPHISMS

Most of the examples presented in Table 1 were found through the analysis of specific traits and/or elements. Whole genome sequencing allows for detection of potentially all TEs of a given genome—but for most species, the assembled sequence does not fully access the heterochromatic part of the genome, and estimation of TE content is likely underestimated. Nevertheless, sequencing costs remain too high to apply *de novo* assembly to detect TE polymorphisms at the population/germplasm level. Moreover, complete genome sequences based on *de novo* assembly of massive parallel data alone are often not of enough high quality to accurately detect TE presence/absence polymorphisms. Alternatively, analysis of sequence reads from massive parallel sequencing data provides unprecedented opportunities to investigate without *a priori* assumptions the propensity to which TE insertions can underlie adaptive variation. Over the past decade, several methodologies have been developed to detect TE insertion polymorphisms in complex eukaryotic species. In this section, we give an overview of these methodologies, their use and limitations and how they could be improved.

### Sequencing transposon display

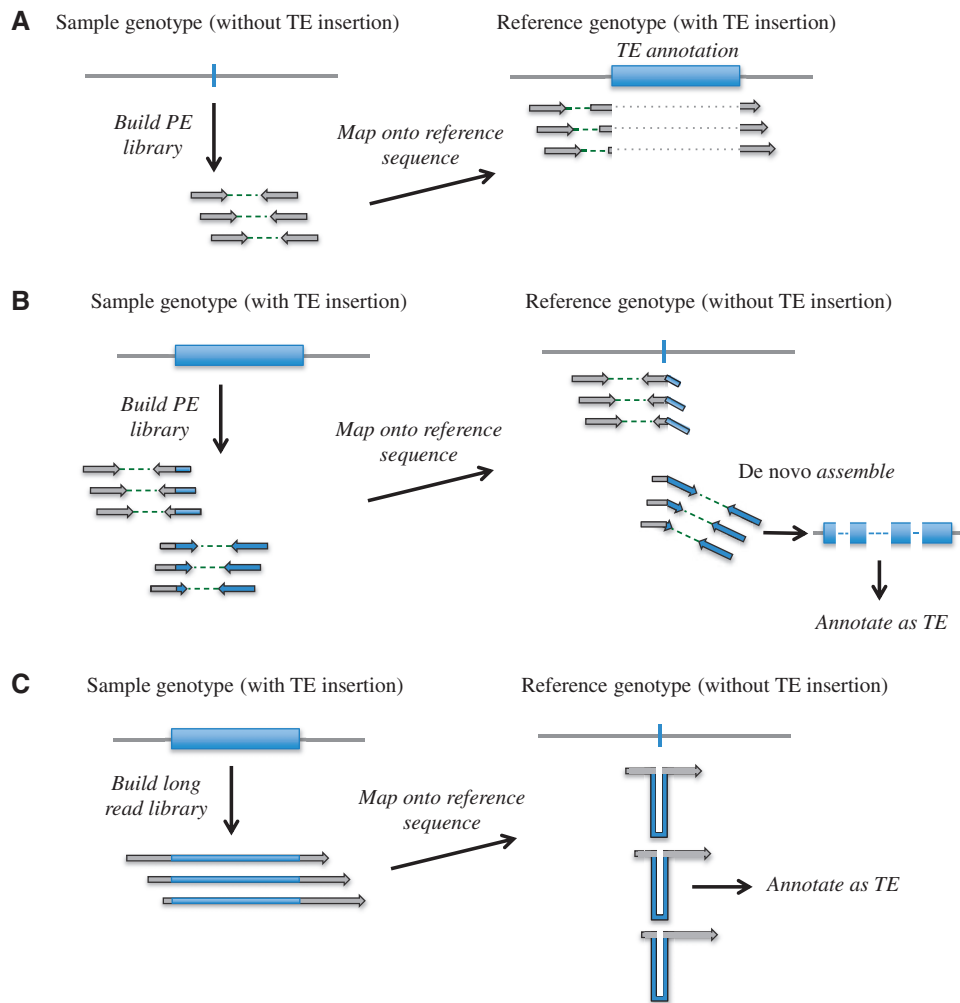
Polymerase chain reaction (PCR)-based methods following genome digestion with restriction enzymes and amplification of TE insertion flanking regions are known as sequence-specific amplification polymorphism (S-SAP, [55]) or transposon display (TD, [56]). They have long been used to detect insertion polymorphisms in plants [57–60]. Several improvements of the original TD technique that allow distinguishing polymorphisms of the restriction site from TE-based polymorphisms include vectorette, splinkerette and digestion–ligation–amplification (DLA)-454. They provide fast, extensive and reliable

parallel characterization of TE insertions in many individuals. For instance, *mPing*-specific vectorette on 24 rice siblings of a self-cross allowed the identification of ~1600 *mPing* insertion sites [61]. Similarly, DLA-454 applied on active *Mu* transposon maize lines led to discover 21 new *Mu* elements and to characterize the distribution of 40 000 non-redundant *Mu* insertion sites along maize chromosomes [62]. While parallel sequencing of differently tagged amplicons targeting various TEs is feasible, the obvious drawback of TD-derived techniques is that they specifically target the set of TEs chosen for PCR amplification (e.g. *mPing*, *Mu*). They therefore do not offer a full characterization of the TE insertional landscape. Availability of a reference genome sequence is useful to localize the TE insertions discovered. Quality of this reference sequence will not highly impact detection of the TE polymorphism itself. Rather, it will impact connection of the insertion with other genomic features, such as nearby genes. To some extent, direct comparison of TD derived sequence reads can also be achieved to detect presence/absence polymorphisms.

### Mapping of single reads on a reference genome

One way to avoid targeting a set of chosen TE families is to directly analyze sequence reads from genomic DNA, therefore, bypassing the PCR amplification step. Over the past few years, the growing interest for structural variation has led to the development of methodologies allowing for detection of presence–absence and copy-number variants from high-throughput sequencing data (mostly Illumina). These methodologies provide a deep characterization of genome wide TE-based polymorphisms, but are highly dependent on the quality of the reference sequence.

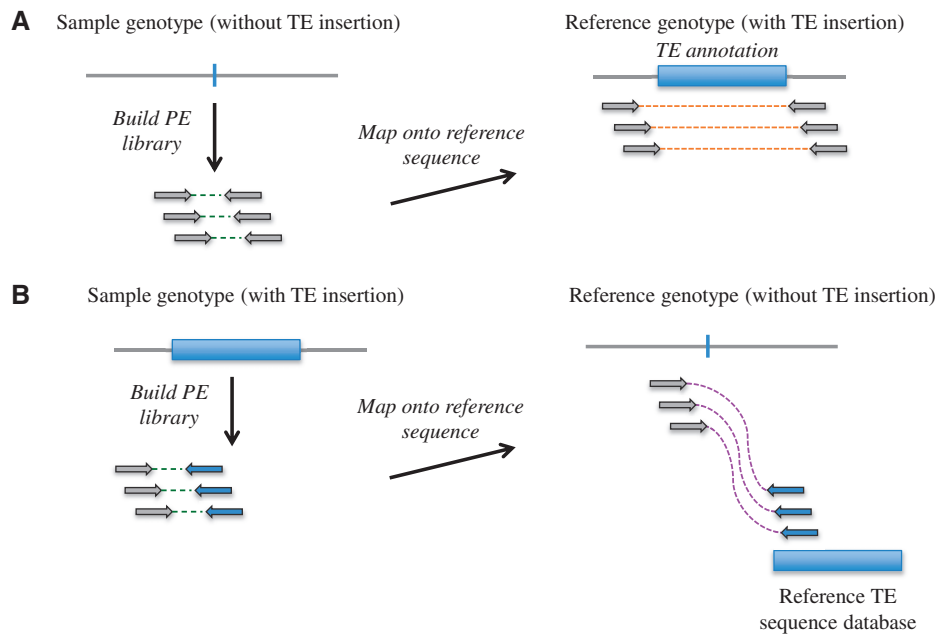
The identification of TE insertion breakpoints with base-pair precision can be achieved by detecting a ‘split mapping’ signature of reads on the reference genome. One split may correspond to an insertion breakpoint, thus separating the sequencing read into one part that matches the genomic region flanking the TE insertion site, and another part that is anchored in the TE sequence. After mapping reads onto the reference genome, uniquely mapping reads are then analyzed. This allows finding reads matching low copy regions flanking TE insertions. A read spanning a region deleted in the sample genome will be mapped twice onto the reference genome (i.e. on



**Figure 3:** Split-read-based methodologies to detect TE presence/absence polymorphisms from NGS data. TE sequence is represented as a rectangle, and empty site of insertion is shown as a vertical bar. **(A)** Detection of a TE insertion that is absent from the sample genome using ‘split read’ signature. A read spanning a deletion breakpoint will map as a split read in two locations (corresponding to the two sides flanking the insertion) in the reference genome. Long reads help detection as they reduce spurious mapping of the split reads. Alternatively, as shown here, use of paired reads can increase proper anchoring of the split reads. Annotation as a TE sequence of the DNA fragment located between the two signatures of a pair points to a TE presence/absence polymorphism. **(B)** Detection of a new TE insertion in the sample genome using ‘split-hanging reads’. Pairs where one read is uniquely mapped, and the other is spanning one of the insertion breakpoints reveal one boundary of a sequence absent from the reference genome. *De novo* assembly of these hanging reads with reads from pairs where one read is hanging and the other is multiply mapping can be used to walk inside the inserted fragment sequence (construction of a scaffold, depicted as discontinuous blocks). Annotation of the assembled fragment as a TE allows characterization of a new TE insertion in the sample genome. **(C)** Detection of a new TE insertion in the sample genome with the use of long reads. Reads for which the middle section does not align to the reference genome while both 5’ and 3’ ends do point to a sequence absent from the reference genome. Further annotation of the middle section may reveal a TE insertion.

both sides of the insertion), but at each genomic location, only part of the read will be mapped (i.e. each read will be split) (Figure 3A). Long reads (such as 454/Roche or Pacific Bioscience—PacBio) are therefore more adapted, as they more reliably capture partial mappings. Alternatively, as proposed by [63], use

of paired reads can provide additional anchoring of the split reads when these are short (Figure 3A). To partially control for false positive, paired-end mapping (PEM) strategies (see below) can be used in complement to detect pairs mapping with larger-than-expected insert size.



**Figure 4:** PEM-based methodologies to detect TE presence/absence polymorphisms from NGS data. TE sequence is represented as a rectangle, and empty site of insertion is shown as a vertical bar. **(A)** Detection of a TE insertion that is absent from the sample genome. Paired-end reads are generated at a distance that is approximately known and homogeneous (dotted lines). Paired reads are mapped onto a reference genome, and distance between reads is computed. Pairs of reads mapping at a distance that is substantially different from the average expected library length (long dotted lines) suggests the presence of a DNA segment in the reference genome that is absent in the sample genome. Annotation of this fragment as a TE sequence then points to a TE presence/absence polymorphism. **(B)** Detection of a new TE insertion in the sample genome. Pairs of reads in which one read maps on a unique sequence in the reference genome sequence and the other maps in a TE sequence are detected. Here, a reference TE database is used to search for TE-mapping reads, but search for reads mapping at multiple positions annotated as TEs in the reference genome is also possible.

Detection of split reads that map only once but with partial length mapping (or ‘hanging reads’) can be used to detect insertions in the sample genome, but additional clustering work is required to reconstruct the inserted segment (Figure 3B). As proposed by [64], *de novo* assembly of hanging reads detected as one cluster can be performed to reconstruct the insertion, but will help only if the inserted segment is smaller than the library insert size. For longer TEs, *de novo* assembly of unmapped reads is also required (Figure 3B).

Finally, if reads are long enough to cover the entire TE length, split read signature detection can be used to directly detect insertions in the sample genome. Because the size of the insertion needs to be smaller than read length, Roche/454 sequencing is suitable for detection of new MITE or SINE insertions, but discovery of insertions of larger elements such as autonomous DNA transposons or LTR retrotransposons requires the use of technologies providing longer reads, such as PacBio (Figure 3C).

### Mapping of paired reads on a reference genome

PEM (from paired-end or mate pair libraries) can be used to detect structural variants. Paired ends are generated from the sample genome. During library preparation, the distance between the two paired ends is size fractionated, hence the expected distribution of distances between them is known. These paired reads are then mapped onto a reference genome. Pairs of reads for which the distance on the reference genome substantially differs from the expected distribution indicate the presence of an insertion/deletion between the sample and the reference genomes. This can be used to detect TEs that are present in the reference and absent from the sample genome (Figure 4A). They can also be used to detect insertions in the sample genome, but in this case (i) the size of the insertion needs to be smaller than the library insert size and (ii) it does not provide any information on the inserted DNA fragment, therefore, hampering to detect whether it is a TE or not.

Alternatively, PEM can be used to detect new TE insertions in the sample genome by searching for pairs in which one read maps unambiguously onto the reference sequence and the other maps to a TE sequence (Figure 4B). This TE sequence can be either multiple genomic regions annotated as TEs, or more simply, come from a TE sequence database that has been collected from genomic sequences of the species. In plants, suitability of the PEM to detect neo-insertions from active TEs was first demonstrated in *Arabidopsis thaliana* and was used to identify TEs activated through the alteration of silencing pathways [65]. It was then successfully applied to detect 34 TE insertions in a rice mutant line regenerated from callus culture [66].

Of course, the efficiency of TE polymorphism detection greatly depends on the quality of the reference genome sequence and that of the TE annotation. This is why its use in crops has been limited to rice. With a median contig length over 100 kb and an assembly covering more than 85% of the genome (reviewed in [6]), other crop species such as sorghum (*Sorghum bicolor*), cucumber (*Cucumis melo*) or peach (*Prunus persica*) are likely to soon provide the level of information required to pursue such approaches. Another parameter that should be considered is the degree to which the reference genome is representative for the whole species. While PEM may be fruitful for species with low structural diversity, such as grapevine [67], they are likely more challenging for outcrossing species with high and complex structural diversity such as maize [33,68]. Note that PEM does not provide precise detection of the TE insertion site (variant breakpoint), and collection of such information therefore requires further investigation once the TE polymorphisms are detected.

### Depth of coverage

Under the assumption that sequencing uniformly covers the sample genome, the number of reads mapping to a genomic location of the reference genome follows a Poisson distribution and is expected to be proportional to the number of times the region is present within the genome. Methods similar to those applied to comparative genome hybridization array data (segmentation, change-point analysis) can be used to find signal breakpoints [69, 70]. However, because strength of the depth of coverage (DOC) signature increases with variant size, it is mainly used to detect large variants. Note also that DOC is efficient for detecting

duplicates or triplicates, but becomes more challenging when the variants are in high copy numbers in both the sample and the reference genomes, which is often the case for TEs. In addition, it provides a poor characterization of the breakpoints. For these multiple reasons, DOC is usually not the preferred approach to detect TE insertions but can be used in combination other approaches, such as PEM [71,72].

### Current limitations and future directions

Several of the methods presented above have been implemented into algorithms that are publicly available. Those include CNVnator, CNVseq, SegSeq, SVdetect, PEMer, VariationHunter, MoDIL, Pindel BreakDancer, HYDRA and SVDetect (reviewed in [73–75]). These softwares may be used *a priori* for analyzing one genotype in different conditions or population-based samples. However, our ability to detect TE polymorphism will ultimately depend on the level of structural diversity between the genotype(s) under survey and the species reference genome. Two recent tools have also been specifically designed to simultaneously discover and characterize TE polymorphisms in large populations: T-lex [76], which detects whether a TE present in the reference genome sequence is present or absent in other individuals, and RelocaTE [77], which has been developed to detect new insertion events from movement of active TEs.

Even with the development of these new tools, we are still facing several challenges inherent to the use of NGS. First, correct mapping on the reference genome sequence is affected by sequencing errors. Second, PEM relies entirely on insert size estimates, which are both sensitive to the presence of chimeric clones and accuracy of library construction. Indeed, while libraries are generally considered homogeneous in size, this is usually questionable. Third, DOC usage is affected by uneven sampling of the genome and may therefore not be robust enough for an accurate prediction of duplications. Because these different methods are sensitive to diverse biases, combining evidence from different types of signature to detect the same event allows for more sensitive detection, better characterization of breakpoints, together with the need of a lesser coverage.

One limitation of all these methodologies is that they are applicable only to uniquely mapped reads (reviewed in [74]) and hence, limited to the detection of TE polymorphisms in the non-repeated

genomic content of the reference and sample genomes. While this is problematic to understand TE-mediated species diversity, it likely provides most of the information important to discover insertions with adaptive value. The main limitation so far is thus the availability of a good genome sequence and a good TE annotation. While the dramatic increase in read size that will arise in coming years will bring solution to assembly issues, efforts need to be made to develop tools allowing for parallel high-level annotation of TEs in multiples genomes.

## METHODS AND CHALLENGES FOR THE DETECTION OF SELECTED TE INSERTIONS

Genome-wide characterization of TE-based polymorphisms using NGS technologies does not predict their adaptive contribution. In fact, TE insertions that have been selected for (such as those presented in Table 1) likely represent a negligible fraction of genomic TE content. To capture this fraction, TE-based population genetic studies and association studies can be undertaken. On one hand, TE-based population genetic studies can help revealing past selection of adaptive insertions, the ones that have been selected for during the domestication and/or breeding history of crops (Table 1). On the other hand, TE-based association mapping can assist both characterization of additional adaptive variation in germplasm collections and elucidation of TE contribution to the genetic architecture of quantitative traits.

Current crop literature, however, lacks such studies. So far, only a single study (in maize) reports the use of polymorphic MITE insertions to screen a population, in this case a sample of 26 maize landraces, as well as an association panel of 367 maize inbred lines. Results of this analysis reveal a significant association between male flowering time and a tandem MITE insertion located in the 3' end of a cytochrome P450-like gene [46].

In principle, TE-based markers could be used as any other markers in population genetic studies to investigate the genetic bases of adaptation. However, as demonstrated in several model species both in animals and plants [78-82], TE variants are specific in at least one respect: they are often deleterious and therefore maintained at low population frequencies as a result of purifying selection. As shown in *Arabidopsis* species, the strength of selection varies

among TE families [82] and strongly depends on parameters influencing the effective population size such as demography [83] as well as the mating system [78,81,82]. Because neutrality tests aiming at detecting positive selection are often based on frequency spectrum, deviation of TE variants toward low population frequencies deserves specific attention.

First, the strength of selection acting against TE insertions for a given TE family can be estimated [84]. The most recent methodological development [85] incorporates ascertainment bias and demographic inferences. It could be used to estimate the strength of purifying selection independently for multiple TE families in several crop genomes for which demographic scenarios for domestication and diversification have been established—see, for instance, maize [86], rice [87], sunflower [88] and apple [89]. Relying on the hypothesis that TE insertions are unlikely to drift to high frequencies when they belong to families subjected to strong purifying selection, it is then possible to identify putatively selected insertions. This procedure has been proposed and employed successfully in *Drosophila melanogaster* [90,91]. But, in theory, estimates of selection coefficients could also serve in coalescent simulations that explicitly model background selection [92] to approximate null distributions of various summary statistics and even ultimately to determine the strength of selection acting on individual TE insertions (although such framework has yet to be developed).

Alternatively to these model-based approaches, TE-insertion genome scans based on empirical distribution of statistics summarizing population data can be utilized to identify insertions that have contributed to past adaptation, with the known pitfall that by definition empirical distributions always contain extreme values in their tails [93]. But copies of a given TE family are spread in the genome (and most likely not in linkage), and therefore statistics relying on an expected 'spatial' pattern of polymorphisms as a function of linkage to the selected allele (where multiple contiguous markers deviate from the neutral expectation) are not applicable. Perhaps, the only summary statistic that could be used because it does not necessitate information at multiple loci but rather information across populations is the population-average  $F_{st}$ .  $F_{st}$  estimates the sample variance in allele frequencies among populations and therefore their degree of differentiation and may be used to

provide evidence for selected insertions [94]. The idea behind the use of estimates of population differentiation is that recent adaptations in populations often reflect the peculiarities of local environments. But these *Fst* measures suffer from several known drawbacks: they are highly variable [95], the detection power depends on the polymorphism frequency at the onset of positive selection [96] and non-independence between populations increases the false positive rate [97].

Finally, correlation of population allele frequencies with environmental variables, e.g. along an environmental cline, offers a valuable tool to investigate how crops have adapted to particular selection pressures (e.g. biotic or abiotic). Correlation methods offer more power than differentiation-based methods but suffer from a higher rate of false positive [98]. Providing a detailed environmental characterization [99], genotype–environment correlation methods allow detection of candidate polymorphisms that have been selected for and contribute to variation in allele frequencies among populations. It also gives insights into the nature of the selective biotic or abiotic ‘agent’ involved. Along this line, Gonzalez *et al.* [100] have used a regression model to test the association between the frequency of TE candidate insertions, and geographic/climatic variables (latitude, maximum and minimum temperature, rainfall) in four Australian populations of *D. melanogaster*. Their results revealed a proportion of significant associations greater than that expected at 5% false discovery rate, suggesting the contribution of TE insertions to climate adaptation. However, treating neighboring populations as independent observations is problematic because populations may share a recent common population history. Coop *et al.* [101] have proposed a Bayesian approach to overcome these difficulties. They provided a null model in which derived population allele frequencies evolve from an ancestral population allele frequency, and estimated the covariance of allele frequencies across derived populations using a Monte Carlo Markov chain. This covariance matrix is used to correct for relatedness among populations. The null model is compared with the alternative model that encompasses a linear relationship between allele frequencies and an environmental variable. Loci in the tail of the resulting Bayes factors’ empirical distribution, i.e. outlier loci, are considered as candidates for adaptation. More recent developments include the estimation of genotype–environment correlations

and random effects linked to population history simultaneously and hence without *a priori* on a ‘neutral’ set of markers [102].

Finally, association mapping studies directly evaluate the link between genotypes and phenotypes measured in the field in large panels of distantly related individuals [103]. It therefore offers a valuable tool to characterize adaptive variation segregating within species—including polymorphisms generated by TEs. It has been applied to crops with great success in mapping genes involved in flowering time [51], photoperiod response [104] and aluminium tolerance in maize [105]; in flowering time, spike length and stem diameter in pearl millet [106]; in grain color, size and quality in rice [107]; in resistance to late blight in potato [108] and in sclerotinia head rot in sunflower [109]. Just as for *Fst* scans or associations with environmental variables, however, the consistency of a pattern/association among markers in tight linkage is usually taken as evidence for true positive as opposed to false positive and facilitates statistical detection. With the lack of linkage among TE insertions, such evidence will remain difficult to obtain except if information at other markers surrounding the insertion is available. In addition, the frequency of TE variants is expected to be lower than other types of variants, stressing the need for screening larger dataset [110] and for adjusting statistical association mapping methods [111].

## FUNCTIONAL VALIDATION

Coupling NGS detection with the use of population genetic/statistical tools allows for finding candidate TE insertions whether they stand as outliers in accordance with a past selective history or whether they are significantly associated to phenotypic variation. Among these candidate insertions, however, some are false positive. Hence, while these strategies represent a major first step toward finding TE adaptive insertions, further functional validation is necessary.

As exemplified by cases presented in Table 1, several methods can be applied. In many cases, fine mapping may be used for cloning the candidate TE insertion. While this allows characterizing the type of TE involved and gives first clues about the possible mechanism underlying the phenotypic change, it does not provide any direct evidence of causality, except in particular cases where TE insertion occurs within an exon and thus disrupts or highly

modifies the corresponding protein sequence. Even in such cases, characterization of the complete cDNA and/or protein remains necessary, as it ensures that the TE sequence is not spliced out. An intermediate level of validation, which is often conducted, resides in comparing the phenotype (at the whole plant level or at the gene expression level) of individuals that do carry or do not the candidate TE insertion. This is usually efficient although individuals often harbor, in addition to the insertion itself, other genomic differences making it hard to rule out possible effect of other *cis* or *trans* factors on the phenotype observed. If however the effect of the insertion is consistent across different genetic backgrounds then the most plausible explanation would be the causality of the TE insertion. Methylation studies can also be performed and used as an additional link between TE presence and gene expression change.

When possible, the best validation resides in analyzing the phenotypic effect of a series of transgenic plants or plants originating from culture of protoplasts injected with constructs that differ only by the presence/absence of the candidate TE. However, in most cases, transformation analyses rule out long distance *cis* effects as well as *trans* effects, but does not exclude possible phenotypic effect due to SNPs or other small indels located in the close vicinity of the candidate TE.

## CONCLUSION

While TEs have long been considered as junk DNA, numerous examples of TE insertions generating new phenotypes of interest for human needs can be uncovered in crops. Technological advances have driven the discovery of TE-triggered phenotypic changes and the underlying mechanisms. Among them, the most recent studies have revealed a prominent role of TE insertions in gene regulatory regions at the origin of expression changes. Full characterization of the functional impact of specific TE insertions is costly, long and therefore rarely undertaken. NGS technologies can nevertheless provide a number of correlates between TE presence/absence and molecular variation (e.g. DNA methylation, transcription or translation) that will contribute to elucidate mechanisms through which TEs can impact phenotypes. The growing number of crop complete genome sequences will be a gold mine in which to dig up TE-mediated polymorphisms. Statistical developments also offer new opportunities

to test the association of these TE polymorphisms with phenotype and/or environmental variation and will ultimately contribute to unravel TE contribution to crop evolution.

### Key points

- There are several mechanisms through which TEs create phenotypic changes: gene disruption, regulation of gene expression, rearrangements, transduction and transposase exaptation. While gene disruption has been historically the easiest to detect, the flow of omics data will provide unprecedented material to estimate their relative contribution to crop evolution.
- The use of NGS data to detect TE-based polymorphisms is a challenging task. Four methods are currently available: sequencing of family specific TD reactions, mapping of paired reads on a reference genome, detecting split read signatures and detecting variations in DOC. Recent advances suggest combining several of these methods promises to lead to more reliable detection.
- Population genetics offer valuable tools to identify TE insertions that have potentially contributed to crop evolution. In theory, both selection coefficients for a given TE family and demographic parameters for the considered crop could be incorporated in coalescent-based models. Such models would be useful to detect signatures of past selection.
- While association mapping provides first hints about the impact of transposon insertions on phenotypic variation, true functional validation is necessary to establish a causal relationship. This critical step has revealed several striking examples of TE-driven changes in crops.
- Up to now, relatively few cases of TE-induced DNA methylation-based expression changes have been characterized. Deciphering the intimate interplay between TEs, DNA methylation and nearby gene expression will provide new insights into how TEs can shape crop phenotypic diversity.

### Acknowledgements

The authors are grateful to Sylvain Fustier for help in arranging Figure 1 photographs, Aleksandra Walczak, Johann Joets and Yves Vigouroux for valuable inputs on specific sections of the manuscript. They thank Josep Casabuberta and Brandon Gaut for precious comments that helped improving the manuscript as well as two anonymous reviewers.

### FUNDING

The PhD fellowship of M.A.F. is supported by the Agence Nationale de la Recherche, project grant ANR-12-ADAP-002 to M.I.T.

### References

1. McClintock B. Mutable loci in maize. *Carnegie Institution of Washington Yearbook* 1951;**50**:174–81.
2. Llaca V, Messing J. Amplicons of maize zein genes are conserved within genic but expanded and constricted in intergenic regions. *Plant J* 1998;**15**:211–20.



3. Mao L, Begum D, Goff SA, et al. Sequence and analysis of the tomato JOINTLESS locus. *Plant Physiol* 2001;**126**: 1331–40.
4. SanMiguel P, Tikhonov A, Jin YK, et al. Nested retrotransposons in the intergenic regions of the maize genome. *Science* 1996;**274**:765–8.
5. Wei F, Gobelman-Werner K, Morroll SM, et al. The Mla (powdery mildew) resistance cluster is associated with three NBS-LRR gene families and suppressed recombination within a 240-kb DNA interval on chromosome 5S (1HS) of barley. *Genetics* 1999;**153**:1929–48.
6. Michael TP, Jackson SA. The first 50 plant genomes. *Plant Genome* 2013;**6**: doi:10.3835/plantgenome2013.03.0001in.
7. Jiao Y, Wickett NJ, Ayyampalayam S, et al. Ancestral polyploidy in seed plants and angiosperms. *Nature* 2011; **473**:97–100.
8. Tenaillon MI, Hollister JD, Gaut BS. A triptych of the evolution of plant transposable elements. *Trends Plant Sci* 2010;**15**:417–78.
9. Glemin S, Bataillon T. A comparative view of the evolution of grasses under domestication. *New Phytol* 2009;**183**: 273–90.
10. Muñoz Diez C, Vitte C, Ross-Ibarra J, et al. Using nextgen sequencing to investigate genome size variation and transposable element content. In: Grandbastien M-A, Casacuberta J (eds). *Plant Transposons (Topics in Current Genetics)*, Vol. 24. Germany: Springer-Verlag, 2012:41–58.
11. Vitte C, Bennetzen JL. Analysis of retrotransposon structural diversity uncovers properties and propensities in angiosperm genome evolution. *Proc Natl Acad Sci USA* 2006;**103**: 17638–43.
12. Vitte C, Peterson T, Bennetzen JL. Transposable elements and the evolving structure of the maize genome. In: Prioul JL (ed). *Advances in Maize*, Vol. 3. London: The Society for Experimental Biology, 2011;111–20.
13. Zhang SD, Soltis DE, Yang Y, et al. Multi-gene analysis provides a well-supported phylogeny of Rosales. *Mol Phylogenet Evol* 2011;**60**:21–8.
14. García-Mas J, Benjak A, Sanseverino W, et al. The genome of melon (*Cucumis melo* L.). *Proc Natl Acad Sci USA* 2012;**109**: 11872–7.
15. Shulaev V, Sargent DJ, Crowhurst RN, et al. The genome of woodland strawberry (*Fragaria vesca*). *Nat Genet* 2011;**43**: 109–16.
16. The International Peach Genome Initiative, Verde I, Abbott AG, et al. The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. *Nat Genet* 2013;**45**:487–94.
17. Wu J, Wang Z, Shi Z, et al. The genome of the pear (*Pyrus bretschneideri* Rehd.). *Genome Res* 2013;**23**:396–408.
18. Velasco R, Zharkikh A, Affourtit J, et al. The genome of the domesticated apple (*Malus × domestica* Borkh.). *Nat Genet* 2010;**42**:833–9.
19. Nystedt B, Street NR, Wetterbom A, et al. The Norway spruce genome sequence and conifer genome evolution. *Nature* 2013;**8**:e72439.
20. Al-Mssallem IS, Hu S, Zhang X, et al. Genome sequence of the date palm *Phoenix dactylifera* L. *Nat Commun* 2013;**4**:2274.
21. D'Hont A, Denoeud F, Aury J-M, et al. The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. *Nature* 2012;**488**:213–7.
22. Choulet F, Wicker T, Rustenholz C, et al. Megabase level sequencing reveals contrasted organization and evolution patterns of the wheat gene and transposable element spaces. *Plant Cell* 2010;**22**:1686–701.
23. The International Barley Genome Sequencing Consortium. A physical, genetic and functional sequence assembly of the barley genome. *Nature* 2012;**491**:711–6.
24. The Tomato Genome Consortium. The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* 2012;**485**:635–41.
25. Xu Q, Chen L-L, Ruan X, et al. The draft genome of sweet orange (*Citrus sinensis*). *Nat Genet* 2013;**45**:59–66.
26. The Brassica rapa Genome Sequencing Project C, Wang X, Wang H, et al. The genome of the meso-polyploid crop species *Brassica rapa*. *Nat Genet* 2011;**43**:1035–9.
27. Ming R, Hou S, Feng Y, et al. The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus). *Nature* 2008;**452**:991–6.
28. Natali L, Cossu RM, Barghini E, et al. The repetitive component of the sunflower genome as shown by different procedures for assembling next generation sequencing reads. *BMC Genomics* 2013;**14**:686.
29. Paterson AH, Bowers JE, Bruggmann R, et al. The *Sorghum bicolor* genome and the diversification of grasses. *Nature* 2009;**457**:551–6.
30. Zhang G, Liu X, Quan Z, et al. Genome sequence of foxtail millet (*Setaria italica*) provides insights into grass evolution and biofuel potential. *Nat Biotechnol* 2012;**30**: 549–54.
31. Barghini E, Cavallini A, Cossu RM, et al. A survey of the repetitive component of the olive (*Olea europaea* L.) genome. *Poster at the Plant Genome Evolution Conference*. Amsterdam, 2013.
32. Ma JX, Bennetzen JL. Rapid recent growth and divergence of rice nuclear genomes. *Proc Natl Acad Sci USA* 2004;**101**: 12404–10.
33. Wang Q, Dooner HK. Remarkable variation in maize genome structure inferred from haplotype diversity at the bz locus. *Proc Natl Acad Sci USA* 2006;**103**:17644–9.
34. Chao L, Vargas C, Spear BB, et al. Transposable elements as mutator genes in evolution. *Nature* 1983;**303**:633–5.
35. Torkamanzehi A, Moran C, Nicholas FW. P-element transposition contributes substantial new variation for a quantitative trait in *Drosophila melanogaster*. *Genetics* 1992; **131**:73–8.
36. Huang CRL, Burns KH, Boeke JD. Active transposition in genomes. *Annu Rev Genet* 2012;**46**:651–75.
37. Grandbastien MA. Activation of plant retrotransposons under stress conditions. *Trends Plant Sci* 1998;**3**:181–7.
38. Casacuberta E, Gonzalez J. The impact of transposable elements in environmental adaptation. *Mol Ecol* 2013;**22**: 1503–17.
39. Lisch D. How important are transposons for plant evolution? *Nat Rev Genet* 2013;**14**:49–61.
40. Alix K, Ryder CD, Moore J, et al. The genomic organization of retrotransposons in *Brassica oleracea*. *Plant Mol Biol* 2005;**59**:839–51.

41. Walbot V, Petrov DA. Gene galaxies in the maize genome. *Proc Natl Acad Sci USA* 2001;**98**:8163–4.
42. Benjak A, Boue S, Forneck A, *et al.* Recent amplification and impact of MITEs on the genome of grapevine (*Vitis vinifera* L.). *Genome Biol Evol* 2009;**1**:75–84.
43. Bureau TE, Ronald PC, Wessler SR. A computer-based systematic survey reveals the predominance of small inverted-repeat elements in wild-type rice genes. *Proc Natl Acad Sci USA* 1996;**93**:8524–9.
44. Bureau TE, Wessler SR. Stowaway: a new family of inverted repeat elements associated with the genes of both monocotyledonous and dicotyledonous plants. *Plant Cell* 1994;**6**:907–16.
45. Sarilar V, Marmagne A, Brabant P, *et al.* BraSto, a Stowaway MITE from Brassica: recently active copies preferentially accumulate in the gene space. *Plant Mol Biol* 2011;**77**:59–75.
46. Zerjal T, Rousselet A, Mhiri C, *et al.* Maize genetic diversity and association mapping using transposable element insertion polymorphisms. *Theor Appl Genet* 2012;**124**:1521–37.
47. Wessler SR, Baran G, Varagona M, *et al.* Excision of Ds produces waxy proteins with a range of enzymatic activities. *EMBOJ* 1986;**5**:2427–32.
48. Harberd NP, Flavell RB, Thompson RD. Identification of a transposon-like insertion in a Glu-1 allele of wheat. *Mol Gen Genet* 1987;**209**:326–32.
49. Bhattacharyya MK, Smith AM, Ellis THN, *et al.* The wrinkled-seed character of pea described by Mendel is caused by a transposon-like insertion in a gene encoding starch-branching enzyme. *Cell* 1990;**60**:115–22.
50. Yan L, Fu D, Li C, *et al.* The wheat and barley vernalization gene VRN3 is an orthologue of FT. *Proc Natl Acad Sci USA* 2006;**103**:19581–6.
51. Salvi S, Sponza G, Morgante M, *et al.* Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. *Proc Natl Acad Sci USA* 2007;**104**:11376–81.
52. Studer A, Zhao Q, Ross-Ibarra J, *et al.* Identification of a functional transposon insertion in the maize domestication gene tb1. *Nat Genet* 2011;**43**:1160–3.
53. Martin A, Troadec C, Boualem A, *et al.* A transposon-induced epigenetic change leads to sex determination in melon. *Nature* 2009;**461**:1135–8.
54. Chantret N, Salse J, Sabot F, *et al.* Molecular basis of evolutionary events that shaped the hardness locus in diploid and polyploid wheat species (*Triticum* and *Aegilops*). *Plant Cell* 2005;**17**:1033–45.
55. Waugh R, McLean K, Flavell AJ, *et al.* Genetic distribution of Bare-1-like retrotransposable elements in the barley genome revealed by sequence-specific amplification polymorphisms (S-SAP). *Mol Gen Genet* 1997;**253**:687–94.
56. Van den Broeck D, Maes T, Sauer M, *et al.* Transposon display identifies individual transposable elements in high copy number lines. *Plant J* 1998;**13**:121–9.
57. Parisod C, Alix K, Just J, *et al.* Impact of transposable elements on the organization and function of allopolyploid genomes. *New Phytol* 2010;**186**:37–45.
58. Pearce SR, Knox M, Ellis TH, *et al.* Pea Ty1-copia group retrotransposons: transpositional activity and use as markers to study genetic diversity in *Pisum*. *Mol Gen Genet* 2000;**263**:898–907.
59. Petit M, Guidat C, Daniel J, *et al.* Mobilization of retrotransposons in synthetic allotetraploid tobacco. *New Phytol* 2010;**186**:135–47.
60. Sarilar V, Palacios PM, Rousselet A, *et al.* Allopolyploidy has a moderate impact on restructuring at three contrasting transposable element insertion sites in resynthesized *Brassica napus* allotetraploids. *New Phytol* 2013;**198**:593–604.
61. Naito K, Zhang F, Tsukiyama T, *et al.* Unexpected consequences of a sudden and massive transposon amplification on rice gene expression. *Nature* 2009;**461**:1130–4.
62. Liu S, Yeh CT, Ji T, *et al.* Mu transposon insertion sites and meiotic recombination events co-localize with epigenetic marks for open chromatin across the maize genome. *PLoS Genet* 2009;**5**:e1000733.
63. Ye K, Schulz MH, Long Q, *et al.* Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* 2009;**25**:2865–71.
64. Kidd JM, Cooper GM, Donahue WF, *et al.* Mapping and sequencing of structural variation from eight human genomes. *Nature* 2008;**453**:56–64.
65. Mirouze M, Reinders J, Bucher E, *et al.* Selective epigenetic control of retrotransposition in *Arabidopsis*. *Nature* 2009;**461**:427–30.
66. Sabot F, Picault N, El-Baidouri M, *et al.* Transpositional landscape of the rice genome revealed by paired-end mapping of high-throughput re-sequencing data. *Plant J* 2011;**66**:241–6.
67. Jaillon O, Aury JM, Noel B, *et al.* The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 2007;**449**:U463–5.
68. Swanson-Wagner RA, Eichten SR, Kumari S, *et al.* Pervasive gene content variation and copy number variation in maize and its undomesticated progenitor. *Genome Res* 2010;**20**:1689–99.
69. Campbell PJ, Stephens PJ, Pleasance ED, *et al.* Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. *Nat Genet* 2008;**40**:722–9.
70. Chiang DY, Getz G, Jaffe DB, *et al.* High-resolution mapping of copy-number alterations with massively parallel sequencing. *Nat Methods* 2009;**6**:99–103.
71. Cridland JM, Thornton KR. Validation of rearrangement break points identified by paired-end sequencing in natural populations of *Drosophila melanogaster*. *Genome Biol Evol* 2010;**2**:83–101.
72. Qi J, Zhao F. inGAP-sv: a novel scheme to identify and visualize structural variation from paired end mapping data. *Nucleic Acids Res* 2011;**39**:W567–75.
73. Elbaidouri M, Panaud O. Genome-wide analysis of transposition using next generation sequencing technologies. In: Grandbastien M-A, Casacuberta J (eds). *Plant Transposons (Topics in Current Genetics)*, Vol. Vol. 24. Germany: Springer-Verlag, 2012:59–70.
74. Medvedev P, Stanciu M, Brudno M. Computational methods for discovering structural variation with next-generation sequencing. *Nat Methods* 2009;**6**:S13–20.
75. Teo SM, Pawitan Y, Ku CS, *et al.* Statistical challenges associated with detecting copy number variations with next-generation sequencing. *Bioinformatics* 2012;**28**:2711–8.

76. Fiston-Lavier AS, Carrigan M, Petrov DA, et al. T-lex: a program for fast and accurate assessment of transposable element presence using next-generation sequencing data. *Nucleic Acids Res* 2011;**39**:e36.
- 5 77. Robb SM, Lu L, Valencia E, et al. The use of RelocaTE and unassembled short reads to produce high-resolution snapshots of transposable element generated diversity in rice. *G3 (Bethesda)* 2013;**3**:949–57.
78. Petrov DA, Fiston-Lavier AS, Lipatov M, et al. Population genomics of transposable elements in *Drosophila melanogaster*. *Mol Biol Evol* 2011;**28**:1633–44.
- 10 79. Charlesworth B, Langley CH. The population-genetics of *Drosophila* transposable elements. *Annu Rev Genet* 1989;**23**:251–87.
- 15 80. Charlesworth B, Sniegowski P, Stephan W. The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* 1994;**371**:215–20.
81. Wright SI, Le QH, Schoen DJ, et al. Population dynamics of an Ac-like transposable element in self- and cross-pollinating *Arabidopsis*. *Genetics* 2001;**158**:1279–88.
- 20 82. Lockton S, Gaut BS. The evolution of transposable elements in natural populations of self-fertilizing *Arabidopsis thaliana* and its outcrossing relative *Arabidopsis lyrata*. *BMC Evol Biol* 2010;**10**:10.
- 25 83. Lockton S, Ross-Ibarra J, Gaut BS. Demography and weak selection drive patterns of transposable element diversity in natural populations of *Arabidopsis lyrata*. *Proc Natl Acad Sci USA* 2008;**105**:13965–70.
84. Petrov DA, Aminetzach YT, Davis JC, et al. Size matters: non-LTR retrotransposable elements and ectopic recombination in *Drosophila*. *Mol Biol Evol* 2003;**20**:880–92.
- 30 85. Gonzalez J, Macpherson JM, Messer PW, et al. Inferring the strength of selection in *Drosophila* under complex demographic models. *Mol Biol Evol* 2009;**26**:513–26.
- 35 86. Wright SI, Bi IV, Schroeder SG, et al. The effects of artificial selection on the maize genome. *Science* 2005;**308**:1310–4.
87. Molina J, Sikora M, Garud N, et al. Molecular evidence for a single evolutionary origin of domesticated rice. *Proc Natl Acad Sci USA* 2011;**108**:8351–6.
- 40 88. Strasburg JL, Rieseberg LH. Molecular demographic history of the annual sunflowers *Helianthus annuus* and *H. petiolaris*—large effective population sizes and rates of long-term gene flow. *Evolution* 2008;**62**:1936–50.
- 45 89. Cornille A, Gladieux P, Smulders MJM, et al. New insight into the history of domesticated apple: secondary contribution of the European wild apple to the genome of cultivated varieties. *PLoS Genet* 2012;**8**:e1002703.
90. Gonzalez J, Macpherson JM, Petrov DA. A recent adaptive transposable element insertion near highly conserved developmental loci in *Drosophila melanogaster*. *Mol Biol Evol* 2009;**26**:1949–61.
- 50 91. Gonzalez J, Lenkov K, Lipatov M, et al. High rate of recent transposable element-induced adaptation in *Drosophila melanogaster*. *PLoS Biol* 2008;**6**:2109–29.
- 55 92. Walczak AM, Nicolaisen LE, Plotkin JB, et al. The structure of genealogies in the presence of purifying selection: a fitness-class coalescent. *Genetics* 2012;**190**:U753–83.
93. Tenaillon MI, Tiffin PL. The quest for adaptive evolution: a theoretical challenge in a maze of data. *Curr Opin Plant Biol* 2008;**11**:110–5.
- 60 94. Weir BS, Hill WG. Estimating F-statistics. *Annu Rev Genet* 2002;**36**:721–50.
95. Weir BS, Cardon LR, Anderson AD, et al. Measures of human population structure show heterogeneity among genomic regions. *Genome Res* 2005;**15**:1468–76.
- 65 96. Oleksyk TK, Smith MW, O'Brien SJ. Genome-wide scans for footprints of natural selection. *Philos Trans R Soc B Biol Sci* 2010;**365**:185–205.
97. Excoffier L, Hofer T, Foll M. Detecting loci under selection in a hierarchically structured population. *Heredity* 2009;**103**:285–98.
- 70 98. de Mita S, Thuillet A-C, Gay L, et al. Detecting selection along environmental gradients: analysis of eight methods and their effectiveness for outbreeding and selfing populations. *Mol Ecol* 2013;**22**:1383–99.
- 75 99. Porter JH, Hanson PC, Lin CC. Staying afloat in the sensor data deluge. *Trends Ecol Evol* 2012;**27**:121–9.
100. Gonzalez J, Karasov TL, Messer PW, et al. Genome-wide patterns of adaptation to temperate environments associated with transposable elements in *Drosophila*. *PLoS Genet* 2010;**6**:e1000905.
- 80 101. Coop G, Witonsky D, Di Rienzo A, et al. Using environmental correlations to identify loci underlying local adaptation. *Genetics* 2010;**185**:1411–23.
- 85 102. Frichot E, Schoville SD, Bouchard G, et al. Testing for associations between loci and environmental gradients using latent factor mixed models. *Mol Biol Evol* 2013;**30**:1687–99.
103. Bergelson J, Roux F. Towards identifying genes underlying ecologically relevant traits in *Arabidopsis thaliana*. *Nat Rev Genet* 2010;**11**:867–79.
- 90 104. Hung HY, Shannon LM, Tian F, et al. ZmCCT and the genetic basis of day-length adaptation underlying the post-domestication spread of maize. *Proc Natl Acad Sci USA* 2012;**109**:E1913–21.
- 95 105. Krill AM, Kirst M, Kochian LV, et al. Association and linkage analysis of aluminum tolerance genes in maize. *PLoS One* 2010;**5**:e9958.
106. Saidou AA, Mariac C, Luong V, et al. Association studies identify natural variation at PHYC linked to flowering time and morphological variation in pearl millet. *Genetics* 2009;**182**:899–910.
107. Huang XH, Wei XH, Sang T, et al. Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat Genet* 2010;**42**:U961–76.
- 105 108. Pajeroska-Mukhtar K, Stich B, Achenbach U, et al. Single nucleotide polymorphisms in the allene oxide synthase 2 gene are associated with field resistance to late blight in populations of tetraploid potato cultivars. *Genetics* 2009;**181**:1115–27.
- 110 109. Fusari CM, Di Rienzo JA, Troglia C, et al. Association mapping in sunflower for sclerotinia head rot resistance. *BMC Plant Biol* 2012;**12**:93.
110. Morrell PL, Buckler ES, Ross-Ibarra J. Crop genomics: advances and applications. *Nat Rev Genet* 2012;**13**:85–96.
- 115 111. Wallace JG, Larsson SJ, Buckler ES. Entering the second century of maize quantitative genetics. *Heredity* 2013;**9**:e1003246.
- 120 112. Johns MA, Mottinger J, Freeling M. A low copy number, copia-like transposon in maize. *EMBO J* 1985;**4**:1093–101.

113. Jin YK, Bennetzen JL. Structure and coding properties of Bs1, a maize retrovirus-like transposon. *Proc Natl Acad Sci USA* 1989;**86**:6235–9.
114. Moon S, Jung K-H, Lee D-E, *et al.* Identification of active transposon dTok, a member of the hAT family, in rice. *Plant Cell Phys* 2006;**47**:1473–83.
115. Tsugane K, Maekawa M, Takagi K, *et al.* An active DNA transposon nDart causing leaf variegation and mutable dwarfism and its related elements in rice. *Plant J* 2006;**45**: 46–57.
116. Hori Y, Fujimoto R, Sato Y, *et al.* A novel wx mutation caused by insertion of a retrotransposon-like sequence in a glutinous cultivar of rice (*Oryza sativa*). *Theor Appl Genet* 2007;**115**:217–24.
117. Liu B, Kanazawa A, Matsumura H, *et al.* Genetic redundancy in soybean photoresponses associated with duplication of the phytochrome A gene. *Genetics* 2008;**180**: 995–1007.
118. Kanazawa A, Liu B, Kong F, *et al.* Adaptive evolution involving gene duplication and insertion of a novel Ty1/copia-like retrotransposon in soybean. *J Mol Evol* 2009;**69**: 164–75.
119. Momose M, Abe Y, Ozeki Y. Miniature inverted-repeat transposable elements of stowaway are active in potato. *Genetics* 2010;**186**:59–66.
120. Lowe B, Mathern J, Hake S. Active Mutator elements suppress the knotted phenotype and increase recombination at the Kn1-O tandem duplication. *Genetics* 1992; **132**:813–22.
121. Greene B, Walko R, Hake S. Mutator insertions in an intron of the maize knotted1 gene result in dominant suppressible mutations. *Genetics* 1994;**138**:1275–85.
122. Chopra S, Brendel V, Zhang J, *et al.* Molecular characterization of a mutable pigmentation phenotype and isolation of the first active transposable element from *Sorghum bicolor*. *Proc Natl Acad Sci USA* 1999;**96**:15330–5.
123. Yao J-L, Dong Y-H, Morris BAM. Parthenocarpic apple fruit production conferred by transposon insertion mutations in a MADS-box transcription factor. *Proc Natl Acad Sci USA* 2001;**98**:1306–11.
124. Zabala G, Vodkin LO. The wp mutation of glycine max carries a gene-fragment-rich transposon of the CACTA superfamily. *Plant Cell Online* 2005;**17**:2619–32.
125. Xu M, Brar HK, Grosic S, *et al.* Excision of an active CACTA-like transposable element from DFR2 causes variegated flowers in Soybean [*Glycine max* (L.) Merr.]. *Genetics* 2010;**184**:53–63.
126. Varagona MJ, Purugganan M, Wessler SR. Alternative splicing induced by insertion of retrotransposons into the maize waxy gene. *Plant Cell* 1992;**4**:811–20.
127. Schiefelbein JW, Furtek DB, Dooner HK, *et al.* Two mutations in a maize bronze-1 allele caused by transposable elements of the Ac-Ds family alter the quantity and quality of the gene product. *Genetics* 1988;**120**:767–77.
128. Dawe RK, Lachmansingh AR, Freeling M. Transposon-mediated mutations in the untranslated leader of maize Adh1 that increase and decrease pollen-specific gene expression. *Plant Cell* 1993;**5**:311–9.
129. Palaisa KA, Morgante M, Williams M, *et al.* Contrasting effects of selection on sequence diversity and linkage disequilibrium at two phytoene synthase loci. *Plant Cell* 2003; **15**:1795–806.
130. Kobayashi S, Goto-Yamamoto N, Hirochika H. Retrotransposon-induced mutations in grape skin color. *Science* 2004;**304**:982.
131. Magalhaes JV, Liu J, Guimarães CT, *et al.* A gene in the multidrug and toxic compound extrusion (MATE) family confers aluminum tolerance in sorghum. *Nat Genet* 2007; **39**:1156–61.
132. Walker AR, Lee E, Bogs J, *et al.* White grapes arose through the mutation of two similar and adjacent regulatory genes. *Plant J* 2007;**49**:772–85.
133. Hayashi K, Yoshida H. Refunctionalization of the ancient rice blast disease resistance gene Pit by the recruitment of a retrotransposon as a promoter. *Plant J* 2009;**57**:413–25.
134. Chiu L-W, Zhou X, Burke S, *et al.* The purple cauliflower arises from activation of a MYB transcription factor. *Plant Physiol* 2010;**154**:1470–80.
135. Fernandez L, Torregrosa L, Segura V, *et al.* Transposon-induced gene activation as a mechanism generating cluster shape somatic variation in grapevine. *Plant J* 2010;**61**: 545–57.
136. Chu CG, Tan CT, Yu GT, *et al.* A novel retrotransposon inserted in the dominant Vrn-B1 allele confers spring growth habit in tetraploid wheat (*Triticum turgidum* L.). *G3* 2011;**1**:637–45.
137. Butelli E, Licciardello C, Zhang Y, *et al.* Retrotransposons control fruit-specific, cold-dependent accumulation of anthocyanins in blood oranges. *Plant Cell* 2012;**24**:1242–55.
138. Elrouby N, Bureau TE. Modulation of AUXIN-BINDING PROTEIN 1 gene expression in maize and the teosintes by transposon insertions in its promoter. *Mol Genet Genomics* 2012;**287**:143–53.
139. Hong L, Qian Q, Tang D, *et al.* A mutation in the rice chalcone isomerase gene causes the golden hull and internode 1 phenotype. *Planta* 2012;**236**:141–51.
140. Li X, Chen L, Hong M, *et al.* A large insertion in bHLH transcription factor BrTT8 resulting in yellow seed coat in *Brassica rapa*. *PLoS ONE* 2012;**7**:e44145.
141. Zhang P, Allen WB, Nagasawa N, *et al.* A transposable element insertion within ZmGE2 gene is associated with increase in embryo to endosperm ratio in maize. *Theor Appl Genet* 2012;**125**:1463–71.
142. Yang Q, Li Z, Li W, *et al.* CACTA-like transposable element in ZmCCT attenuated photoperiod sensitivity and accelerated the postdomestication spread of maize. *Proc Natl Acad Sci USA* 2013;**110**:16969–74.
143. Jin YK, Bennetzen JL. Integration and nonrandom mutation of a plasma membrane proton ATPase gene fragment within the Bs1 retroelement of maize. *Plant Cell* 1994;**6**: 1177–86.
144. Elrouby N, Bureau TE. A novel hybrid open reading frame formed by multiple cellular gene transductions by a plant long terminal repeat retroelement. *J Biol Chem* 2001; **276**:41963–8.
145. Elrouby N, Bureau TE. Bs1, a new chimeric gene formed by retrotransposon-mediated exon shuffling in maize. *Plant Physiol* 2010;**153**:1413–24.
146. Zhang J, Zhang F, Peterson T. Transposition of reversed Ac element ends generates novel chimeric genes in maize. *PLoS Genet* 2006;**2**:e164.
147. Muehlbauer GJ, Bhau BS, Syed NH, *et al.* A hAT superfamily transposase recruited by the cereal grass genome. *Mol Genet Genomics* 2006;**275**:553–63.

148. Jameson N, Georgelis N, Fouladbash E, *et al.* Helitron mediated amplification of cytochrome P450 monooxygenase gene in maize. *Plant Mol Biol* 2008;**67**: 295–304.
- 5 149. Walker EL, Robbins TP, Bureau TE, *et al.* Transposon-mediated chromosomal rearrangements and gene duplications in the formation of the maize R-r complex. *EMBOJ* 1995;**14**:2350–63.
- 10 150. Xiao H, Jiang N, Schaffner E, *et al.* A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. *Science* 2008;**319**:1527–30.
151. Hake S, Vollbrecht E, Freeling M. Cloning knotted, the dominant morphological mutant in maize using Ds2 as a transposon tag. *EMBOJ* 1989;**8**:15–22.
152. Chen C-H, Oishi KK, Kloeckener-Gruissem B, *et al.* Organ-specific expression of maize Adh1 is altered after a Mu transposon insertion. *Genetics* 1987;**116**:469–77. 15
153. Kloeckener-Gruissem B, Freeling M. Transposon-induced promoter scrambling: a mechanism for the evolution of new alleles. *Proc Natl Acad Sci USA* 1995;**92**:1836–40. 20
154. Kawase M, Fukunaga K, Kato K. Diverse origins of waxy foxtail millet crops in East and Southeast Asia mediated by multiple transposable element insertions. *Mol Genet Genomics* 2005;**274**:131–40.
155. Kashkush K, Khasdan V. Large-scale survey of cytosine methylation of retrotransposons and the impact of readout transcription from long terminal repeats on expression of adjacent rice genes. *Genetics* 2007;**177**:1975–85. 25

## ANNEXE I.2 DESCRIPTION DES 37 POPULATIONS

Tableau ANNEXE I.2.a Localisation des populations de téosintes échantillonnées.

Population	Nomenclature	Sous-espèce	Longitude	Latitude	Altitude (m)	Municipalité	Etat
<b>Gradient 1</b>							
L1	CM04	<i>parviglumis</i>	-99,28547	16,98114	581	Teconoapa	Guerrero
2	CM03	<i>parviglumis</i>	-99,54147	17,17192	343	Juan R Escudero	Guerrero
3	CM05	<i>parviglumis</i>	-99,47764	17,39175	1201	Chilpancingo	Guerrero
4	CM06	<i>parviglumis</i>	-99,45072	17,42189	1251	Chilpancingo	Guerrero
5	CM07	<i>parviglumis</i>	-99,36831	17,45992	1107	Mochitlán	Guerrero
6	CM08	<i>parviglumis</i>	-99,21800	18,23769	1101	Huitzuco	Guerrero
7	CM01	<i>parviglumis</i>	-99,84108	18,34983	1649	Teloloapan	Guerrero
IP1	CM02	<i>parviglumis</i>	-99,90842	18,41100	1439	Teloloapan	Guerrero
IM1	CM09	<i>mexicana</i>	-99,07028	18,97411	1669	Tepoztlán	Morelos
10	SMH581	<i>parviglumis</i>	-100,42144	19,32739	1383	Zitacuaro	Michoacán
H1	CM10	<i>mexicana</i>	-99,62706	19,40747	2581	Otzolotepec	México
33	Texcoco	<i>mexicana</i>	-98,91461	19,50253	2234	Texcoco	México
34	Cocotitlan	<i>mexicana</i>	-98,86614	19,22625	2252	Cocotitlán	México
35	Tenancingo	<i>mexicana</i>	-98,18542	19,15947	2306	Tenancingo	Tlaxcala
36	San Nicolas	<i>mexicana</i>	-97,55278	19,17425	2375	Sn Nicolas Buenos Aires	Puebla
37	Calpan	<i>mexicana</i>	-98,48744	19,08389	2447	Calpan	Puebla
<b>Gradient 2</b>							
12	SMH577	<i>parviglumis</i>	-104,87186	14,73247	572	Villa de Purificación	Jalisco
L2	CM13	<i>parviglumis</i>	-104,80489	19,71544	504	Villa Purificación	Jalisco
14	CM12	<i>parviglumis</i>	-104,40789	20,62731	1426	Guauchinango	Jalisco
15	SMH578	<i>parviglumis</i>	-104,05825	19,53536	1369	Tolimán	Jalisco
16	SMH564	<i>parviglumis</i>	-104,17261	19,90997	1407	Ejutla	Jalisco

17	SMH565	<i>parviglumis</i>	-104,17675	19,89614	1317	Ejutla	Jalisco
18	SMH566	<i>parviglumis</i>	-104,15717	19,90267	1140	Ejutla	Jalisco
19	SMH567	<i>parviglumis</i>	-104,11686	19,91297	976	Ejutla	Jalisco
20	MIT15	<i>parviglumis</i>	-104,11233	19,91289	976	Ejutla	Jalisco
21	SMH569	<i>parviglumis</i>	-104,09772	19,92183	963	Ejutla	Jalisco
22	SMH570	<i>parviglumis</i>	-104,00772	19,93389	944	Ejutla	Jalisco
23	MIT14	<i>parviglumis</i>	-103,99581	19,94506	958	Ejutla	Jalisco
24	CM11	<i>mexicana</i>	-102,06844	20,13939	1846	Churintzio	Michoacán
25	SMH579	<i>mexicana</i>	-101,43417	20,13381	2002	Puruandiro	Michoacán
26	SMH576	<i>mexicana</i>	-101,37308	20,16072	1856	Yuriria	Guanajuato
27	SMH580	<i>mexicana</i>	-101,23542	19,96461	1844	Huandacareo	Michoacán
28	SMH575	<i>mexicana</i>	-101,08811	20,05319	1849	Santa Ana Maya	Michoacán
29	SMH573	<i>mexicana</i>	-100,95992	19,98281	1878	Acambaro	Guanajuato
H2	SMH571	<i>mexicana</i>	-100,88489	19,99172	1861	Acambaro	Guanajuato
31	SMH572	<i>mexicana</i>	-100,84967	19,95819	1854	Acambaro	Guanajuato
32	SMH582	<i>mexicana</i>	-102,08983	20,71956	2176	Jesús María	Jalisco

Tableau ANNEXE I.2.b Variables environnementales des populations de téosintes échantillonnées.

Population	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	P12	P13	P14	P15	P16	P17	P18	P19
<b>Gradient 1</b>																			
L1	262	142	78	818	354	172	182	261	263	274	252	1625	416	2	116	1060	9	365	16
2	277	143	76	981	372	185	187	276	280	291	266	1250	303	1	107	762	5	281	61
3	232	138	72	1221	326	135	191	234	234	248	216	1132	251	2	110	742	15	253	22
4	226	140	71	1248	322	127	195	229	229	242	209	1078	254	2	109	708	16	243	22
5	232	142	71	1328	330	131	199	235	222	250	215	1029	232	2	108	670	18	235	22
6	247	160	66	1882	369	129	240	253	224	271	222	988	218	2	107	619	13	300	18
7	218	130	66	1509	324	129	195	213	215	240	201	1131	233	3	106	696	14	84	27
IP1	224	137	66	1615	336	131	205	219	236	248	207	1035	228	2	108	647	15	65	24
IM1	199	158	75	1299	304	96	208	204	191	216	182	967	207	3	106	596	12	271	23
10	201	149	69	1426	311	98	213	205	205	221	183	924	198	3	98	555	18	240	43
H1	141	151	67	1875	245	21	224	156	114	161	114	853	171	9	87	489	37	253	37
33	156	173	70	1871	271	24	247	171	129	176	129	659	139	6	90	384	22	206	22
34	156	172	70	1822	270	25	245	170	129	176	129	699	148	6	90	410	22	217	22
35	155	156	69	1847	261	37	224	168	128	175	128	891	166	6	88	477	24	290	24
36	154	166	72	1730	264	36	228	167	129	172	129	545	101	4	78	264	20	215	20
37	147	161	70	1716	255	28	227	156	121	165	121	990	193	7	94	565	23	297	23
<b>Gradient 2</b>																			
12	241	144	68	1669	339	129	210	254	226	259	218	1606	412	3	115	1137	23	562	62
L2	242	145	68	1646	342	130	212	254	229	260	219	1630	416	3	115	1152	22	576	61
14	192	158	65	2199	311	70	241	213	180	215	160	882	237	5	105	569	27	422	56
15	189	137	66	1433	288	83	205	197	180	205	168	1105	251	5	101	707	21	438	59
16	211	146	66	1565	319	101	218	218	203	229	189	802	180	6	96	478	23	171	51
17	220	148	66	1697	330	108	222	228	211	238	195	819	183	8	96	492	25	173	50
18	225	149	66	1817	337	113	224	234	217	246	199	847	191	7	99	518	24	172	50



19	232	152	65	1950	346	115	231	242	222	253	203	891	202	6	99	550	24	372	50
20	232	152	65	1950	346	115	231	242	222	253	203	891	202	6	99	550	24	372	50
21	233	153	65	1942	349	116	233	244	225	255	205	896	203	6	100	553	24	374	49
22	233	153	64	1989	349	113	236	244	224	254	203	848	189	6	98	512	24	359	50
23	232	153	65	1961	348	113	235	243	224	253	203	845	187	6	97	507	24	179	50
24	185	166	63	2532	311	49	262	198	180	214	148	925	235	6	106	601	23	201	29
25	186	163	64	2335	310	56	254	202	186	214	153	811	205	6	102	510	26	188	35
26	192	164	64	2403	317	61	256	202	193	221	159	767	190	6	102	485	26	171	33
27	184	162	64	2275	306	56	250	194	185	211	152	720	169	6	99	450	25	172	27
28	184	162	64	2341	307	57	250	195	186	212	151	698	167	6	101	441	24	164	26
29	181	162	65	2305	304	55	249	193	181	207	149	694	169	6	105	448	23	161	24
H2	177	158	64	2262	298	54	244	188	177	204	145	715	172	6	104	456	24	169	27
31	179	160	65	2259	300	55	245	191	177	205	147	718	172	6	105	460	22	167	27
32	164	152	63	2290	284	44	240	181	161	191	131	907	241	5	111	610	22	217	29

T1 :Température annuelle moyenne

T2 : Gamme de Température journalière moyenne (Moyenne des mois (max temp-min temp))

T3 : Isothermalité (T2/T7)\*100

T4 : Saisonnalité de la température (écart-type\*100)

T5 : Température maximale du mois le plus chaud

T6 : Température minimale du mois le plus froid

T7 : Gamme de variation annuelle des température (T5-T6)

T8 : Température moyenne du trimestre le plus humide

T9 : Température moyenne du trimestre le plus sec

T10 : Température moyenne du trimestre le plus chaud

T11 : Température moyenne du trimestre le plus froid

P12 : Précipitations annuelles

P13 : Précipitation du mois le plus humide

P14 : Précipitation du mois le plus sec

P15 : Saisonnalité de la précipitation (coefficient de variation)

P16 : Précipitation du trimestre le plus humide

P17 : Précipitation du trimestre le plus sec

P18 : Précipitation du trimestre le plus chaud,

P19 : Précipitation du trimestre le plus froid.

## ANNEXE III.1 SYNTHÈSE BIBLIOGRAPHIQUE

**Tableau 1 : Synthèse bibliographique de locus/gènes candidats à l'adaptation à l'altitude.** Lorsqu'il s'agissait d'un gène, son code a été indiqué dans la colonne correspondante, et sinon c'est le mot « marker » et le/les nom(s) du(des) marqueur(s) associé(s) à la variation phénotypique qui ont été indiqués. Nous avons également indiqué les coordonnées de l'inversion 1 reportée dans l'étude citée comme étant enrichie pour des candidats à l'adaptation locale chez les téosintes.

Gene code or marker	Gene or marker Name	Chro <sub>1</sub>	start position <sub>1</sub>	stop position <sub>1</sub>	phenotypic trait	discovery method <sub>2</sub>	reference	SNP_ID <sub>3</sub>
<b>Heavy metal tolerance</b>								
GRMZM2G097457	pyruvate orthophosphate dikinase2 (pdk2)	8	107003965	107014861	Aluminium tolerance	QTL mapping	(Ninamango-Cárdenas, Guimarães et al. 2003)	-
marker	bnlg198	2	198482067	199166751	Aluminium tolerance	QTL mapping	(Ninamango-Cárdenas, Guimarães et al. 2003)	-
marker	bnlg161b	6	1	499542	Aluminium tolerance	QTL mapping	(Ninamango-Cárdenas, Guimarães et al. 2003)	-
marker	bnlg238	6	1	499542	Aluminium tolerance	QTL mapping	(Ninamango-Cárdenas, Guimarães et al. 2003)	-
marker	phi126	6	1	499542	Aluminium tolerance	QTL mapping	(Ninamango-Cárdenas, Guimarães et al. 2003)	-
marker	nc013	6	148768110	149676378	Aluminium tolerance	QTL mapping	(Ninamango-Cárdenas, Guimarães et al. 2003)	-
marker	bnlg1828	8	169774387	169893763	Aluminium tolerance	QTL mapping	(Ninamango-Cárdenas, Guimarães et al. 2003)	-
marker	nc003	2	164805123	171884854	Aluminium tolerance	QTL mapping	(Ninamango-Cárdenas, Guimarães et al. 2003)	MS_7
marker	NPI235	6	7065267	15096930	Aluminium tolerance	molecular mapping	(Sibov, Gaspar et al. 1999)	-
marker	umc85	6	8273465	8274025	Aluminium tolerance	molecular mapping	(Sibov, Gaspar et al. 1999)	-
marker	CSU70	6	8652623	9588508	Aluminium tolerance	molecular mapping	(Sibov, Gaspar et al. 1999)	-
marker	umc59b	6	96493617	97044869	Aluminium tolerance	molecular mapping	(Sibov, Gaspar et al. 1999)	-
marker	NPI223	6	105480908	105883111	Aluminium tolerance	molecular mapping	(Sibov, Gaspar et al. 1999)	-

marker	umc130	10	10599309	13694033	Aluminium tolerance	molecular mapping	(Sibov, Gaspar et al. 1999)	-
marker	CSU86	10	121252530	122924003	Aluminium tolerance	molecular mapping	(Sibov, Gaspar et al. 1999)	-
marker	NPI232	10	127514254	131092297	Aluminium tolerance	molecular mapping	(Sibov, Gaspar et al. 1999)	-
marker	alm2 -(aluminum tolerance2)	6	88248022	131092297	Aluminium tolerance	molecular mapping	(Sibov, Gaspar et al. 1999)	SB_51 SB_52 Fst_64 Fst_65
marker	alm1 -(aluminum tolerance1)	10	7065267	15096930	Aluminium tolerance	molecular mapping	(Sibov, Gaspar et al. 1999)	-
marker	umc64a	10	88248022	93081741	Aluminium tolerance, drought tolerance, nitrogen tolerance	molecular mapping QTL mapping	(Sibov, Gaspar et al. 1999) (Malosetti, Ribaut et al. 2008)	-
marker	umc59a	6	88014372	89004837	Aluminium tolerance, low-Phosphate tolerance	QTL mapping association mapping	(Sibov, Gaspar et al. 1999) Reiters et al 1991	SB_51
GRMZM5G870170	mate1	6	5866756	5871777	aluminium tolerance	molecular mapping gene expression	Maron et al 2013	-

### Flowering-related genes

GRMZM2G122871	elfg1 - elongation factor gamma1	6	104066600	104071298	anthesis-silking interval	QTL mapping	(Malosetti, Ribaut et al. 2008)	-
marker	umc316-umc107	1	246787256	263079095	anthesis-silking interval	QTL mapping	(Malosetti, Ribaut et al. 2008)	SB_12 SB_13 SB_14 Fst_11
marker	umc322-bn18.01	3	143222128	176248449	anthesis-silking interval	QTL mapping	(Malosetti, Ribaut et al. 2008)	Fst_31
marker	umc30-umc150a	8	162631540	163374227	anthesis-silking interval	QTL mapping	(Malosetti, Ribaut et al. 2008)	-
GRMZM2G003514	<i>zea agamous5</i>	4	156083969	156091115	days for silking	association mapping	(Yu, Li et al. 2012)	-
GRMZM2G102218	drooping leaf ortholog2 - tdsgR22D08	9	146693340	146698631	days for silking	association mapping	(Yu, Li et al. 2012)	-

GRMZM2G421033	tdsgR39F07b tdsgR24B11 dlo2 AP2-EREBP- transcription factor 156 - ereb156	1 0	145279859	145281275	days for silking, response to freezing	association mapping	(Yu, Li et al. 2012)	-
marker	umc156-umc318	4	143187277	158493466	days to male flowering	QTL mapping	(Malosetti, Ribaut et al. 2008)	-
marker	umc91-umc35	7	168150360	169049744	days to male flowering	QTL mapping	(Malosetti, Ribaut et al. 2008)	-
marker	umc317-npi285	1 0	3725625	5273620	days to male flowering	QTL mapping	(Malosetti, Ribaut et al. 2008)	-
GRMZM2G102356		2	40588194	40589174	days to male flowering, number of ears per plant	QTL mapping	(Malosetti, Ribaut et al. 2008)	-
GRMZM2G027995	eif4A	6	87119198	87123018	flowering time	QTL mapping	Durand, Bouchet et al. 2012)	SB_50
GRMZM2G032339	ZMM4	1	277151350	277232124	flowering time		Hung et al 2012 (Danilevskaya, Meng et al. 2008)	-
GRMZM2G144744	dwarf8	1	266094769	266097836	flowering time	association mapping	(Camus-Kulandaivelu, Veyrieras et al. 2006) (Thornsberry, Goodman et al. 2001)	-
GRMZM2G171822	bif2 barren inflorescence2	1	173846756	173848499	flowering time	QTL mapping, association mapping, NAM	Pressoir, Brown et al. 2009) (Buckler, Holland et al. 2009)	-
GRMZM5G861678	elongated mesocoty11 elm1	8	158759002	158763168	Flowering time	functional	(Sawers, Linley et al. 2002)	-
marker	vgt1 -(vegetative to generative transition1)	8	132515815	132877193	flowering time	QTL mapping, Association mapping, functional	(Salvi, Sponza et al. 2007) (Buckler, Holland et al. 2009)	-
GRMZM2G381691	ZmCCt	1 0	94248710	94251264	day-length adaption	NAM	HUNG et al 2012 (Hung, Shannon et al. 2012)	-
GRMZM2G009406	dof2 - DNA binding with one finger2	2	188526058	188527888	flowering time/photoperiod	NAM	(Hung, Shannon et al. 2012)	-

GRMZM2G010290	dof23 - C2C2-Dof-transcription factor 23	1 0	137204209	137206460	sensitivity flowering time/photoperiod sensitivity	NAM	(Hung, Shannon et al. 2012)	-
GRMZM2G011357	Id1 indeterminate growth1	1	239608944	239612267	flowering time/photoperiod sensitivity	NAM	(Hung, Shannon et al. 2012)	-
GRMZM2G013398	col6 - C2C2-CO-like- transcription factor 6	9	38572946	38578563	flowering time/photoperiod sensitivity	NAM	(Hung, Shannon et al. 2012)	-
GRMZM2G016756	bhlh60	5	5101600	5105442	flowering time/photoperiod sensitivity	NAM	(Hung, Shannon et al. 2012)	SB_36
GRMZM2G019686	FLP1	3	155090890	155091742	flowering time/photoperiod sensitivity	NAM	(Hung, Shannon et al. 2012)	-
GRMZM2G038783	col13 - C2C2-CO-like- transcription factor 13	9	108428288	108430108	flowering time/photoperiod sensitivity	NAM	(Hung, Shannon et al. 2012)	-
GRMZM2G042198	col14 - C2C2-CO-like- transcription factor 14	4	162854483	162860718	flowering time/photoperiod sensitivity	NAM	(Hung, Shannon et al. 2012)	-
GRMZM2G049549	crystal proteinIA(b)2	2	43557836	43561417	flowering time/photoperiod sensitivity	NAM	(Hung, Shannon et al. 2012)	-
GRMZM2G058745	hex3	8	124652438	124657923	flowering time/photoperiod sensitivity	NAM	(Hung, Shannon et al. 2012)	SB_65
GRMZM2G062052	pebp19	1 0	126915271	126919128	flowering time/photoperiod sensitivity	NAM	(Hung, Shannon et al. 2012)	-
GRMZM2G062541	bhlh117	1 0	87690145	87692179	flowering time/photoperiod sensitivity	NAM	(Hung, Shannon et al. 2012)	-
GRMZM2G065374	bhlh47 - bHLH- transcription factor 47	1	283599877	283604922	flowering time/photoperiod	NAM	(Hung, Shannon et al. 2012)	-

GRMZM2G067921	dlf1 delayed flowering1	7	175538274	175539760	sensitivity flowering time/photoperiod sensitivity	NAM	(Hung, Shannon et al. 2012)	-
GRMZM2G075562	col5 - C2C2-CO-like- transcription factor 5	5	204170200	204175065	flowering time/photoperiod sensitivity	NAM	(Hung, Shannon et al. 2012)	-
GRMZM2G076602	ereb212 - AP2-EREBP- transcription factor 212	1 0	145350317	145352828	flowering time/photoperiod sensitivity	NAM	(Hung, Shannon et al. 2012)	-
GRMZM2G089850	dof16 - C2C2-Dof- transcription factor 16	7	130360893	130362159	flowering time/photoperiod sensitivity	NAM	(Hung, Shannon et al. 2012)	-
GRMZM2G098813	zfl1 zea floricaula/leafy1	1 0	140851443	140854467	flowering time/photoperiod sensitivity	NAM	(Hung, Shannon et al. 2012) (Weber, Briggs et al. 2008)	-
GRMZM2G115960	bhlh68 - bHLH- transcription factor 68)	3	48804985	48808667	flowering time/photoperiod sensitivity	NAM	(Hung, Shannon et al. 2012)	-
GRMZM2G124532	(phyB1 - phytochromeB1)	1	50025637	50036980	flowering time/photoperiod sensitivity	NAM	(Hung, Shannon et al. 2012)	SB_1
GRMZM2G129247	bzip101 - bZIP- transcription factor 101	6	150080461	150082967	flowering time/photoperiod sensitivity	NAM	(Hung, Shannon et al. 2012)	-
GRMZM2G129889	phyC2	5	7121947	7126963	flowering time/photoperiod sensitivity	NAM	(Hung, Shannon et al. 2012)	-
GRMZM2G131897	dof18 - C2C2-Dof- transcription factor 18	1	164628494	164629568	flowering time/photoperiod sensitivity	NAM	(Hung, Shannon et al. 2012)	-
GRMZM2G132868	GBF4 - bzip116 - bZIP- transcription factor 116	8	100710548	100714109	flowering time/photoperiod sensitivity	NAM	(Hung, Shannon et al. 2012)	-
GRMZM2G144172	dof33 - C2C2-Dof- transcription factor 33	5	2336204	2338578	flowering time/photoperiod	NAM	(Hung, Shannon et al. 2012)	-

GRMZM2G158809	pebp18 - phosphatidylethanolamin e-binding protein18)	2	193189598	193195103	sensitivity flowering time/photoperiod sensitivity	NAM	(Hung, Shannon et al. 2012)	SB_19
GRMZM2G161913	RNA polymerase- associated protein CTR9 homolog	2	16450273	16472826	flowering time/photoperiod sensitivity	NAM	(Hung, Shannon et al. 2012)	-
GRMZM2G162749	dof21 C2C2-Dof- transcription factor 21	1	14850871	14853754	flowering time/photoperiod sensitivity	NAM	(Hung, Shannon et al. 2012)	-
GRMZM2G165042	bhlh43 - bHLH- transcription factor 43	1	252374241	252377817	flowering time/photoperiod sensitivity	NAM	(Hung, Shannon et al. 2012)	-
GRMZM2G165090	bhlh158 - bHLH- transcription factor 158	1	168428180	168432944	flowering time/photoperiod sensitivity	NAM	(Hung, Shannon et al. 2012)	-
GRMZM2G167560	col9	4	165314097	165318410	flowering time/photoperiod sensitivity	NAM	(Hung, Shannon et al. 2012)	-
GRMZM2G171373	hex6 - hexokinase6	3	201789514	201799556	flowering time/photoperiod sensitivity	NAM	(Hung, Shannon et al. 2012)	SB_25
GRMZM2G178795	SRR1	4	159211115	159212289	flowering time/photoperiod sensitivity	NAM	(Hung, Shannon et al. 2012)	-
GRMZM2G179264	zcn8, vgt2, zcn8 phosphatidylethanolamin e-binding protein8	8	123501085	123502873	flowering time/photoperiod sensitivity	NAM	(Hung, Shannon et al. 2012)	-
GRMZM2G311316	ca5p5	1 0	57539877	57543280	flowering time/photoperiod sensitivity	NAM	(Hung, Shannon et al. 2012)	-
GRMZM2G366873	DFL1	8	121797016	121799823	flowering time/photoperiod sensitivity	NAM	(Hung, Shannon et al. 2012)	-
GRMZM2G387528	pif3 - phytochrome interacting factor3	8	935667	938714	flowering time/photoperiod	NAM	(Hung, Shannon et al. 2012)	-



GRMZM2G405368	conz1 constans1	9	34628528	34633066	sensitivity flowering time/photoperiod sensitivity	NAM	(Hung, Shannon et al. 2012)	-
GRMZM2G415077	mybr45	1	300744156	300744814	flowering time/photoperiod sensitivity	NAM	(Hung, Shannon et al. 2012)	-
GRMZM2G432801	hex2	6	160237817	160245075	flowering time/photoperiod sensitivity	NAM	(Hung, Shannon et al. 2012)	-
GRMZM2G440949	ca5p10 - CCAAT-HAP5- transcription factor 510	6	155661607	155667746	flowering time/photoperiod sensitivity	NAM	(Hung, Shannon et al. 2012)	-
GRMZM2G459645	LSH1	4	142384197	142386425	flowering time/photoperiod sensitivity	NAM	(Hung, Shannon et al. 2012)	-
GRMZM2G700665	rap2 - rap2.7 orthologue	8	132044001	132047428	flowering time/photoperiod sensitivity	NAM	(Hung, Shannon et al. 2012)	-
marker	atdof1.8	8	132515815	132877193	flowering time/photoperiod sensitivity	NAM	(Hung, Shannon et al. 2012)	-

### Pigmentation

GRMZM2G005066	c1 - colored aleurone1	9	9740803	9741876	anthocyanin content	molecular evolution study	(Hanson, Gaut et al. 1996)	-
GRMZM5G822829	r1 - colored1	10	138462252	138471072	anthocyanin content	molecular evolution study	(Hanson, Gaut et al. 1996)	-
GRMZM2G012966	lyce1 (lycopene epsilon cyclase1)	8	138882594	138889812	carotenoid composition	association mapping, functionnal	(Harjes, Rocheford et al. 2008)	-
GRMZM2G062559	-	2	44471623	44474212	carotenoid composition	association mapping	(Owens, Lipka et al. 2014)	-
GRMZM2G106250	-	8	171704519	171706395	carotenoid composition	association mapping	(Owens, Lipka et al. 2014)	-
GRMZM2G127139	zep1, (zeaxanthin epoxidase1)	2	44440299	44449237	carotenoid composition	association mapping	(Owens, Lipka et al. 2014)	-

GRMZM2G143202	lut1, cyp14 (cytochrome P450 14)	1	86838334	86848726	carotenoid composition	association mapping	(Owens, Lipka et al. 2014)	-
GRMZM2G152135	crtRB1 ,(hyd3 - hydroxylase3)	10	136057100	136060219	carotenoid composition	association mapping, QTL mapping, expression analyses	Owens et al. 2014, Yan et al. 2010	-
GRMZM2G493395	dxs2 (deoxy xylulose synthase 2)	7	14077852	14081075	carotenoid composition	association mapping	(Owens, Lipka et al. 2014)	-
GRMZM5G837869	lut5, cyp13 (cytochrome P450 13)	5	215827224	215831730	carotenoid composition	association mapping	(Owens, Lipka et al. 2014)	SB_48
marker	vp9 (viviparous9)	7	14539980,00	17368515,00	carotenoid composition	QTL	(Wong, Lambert et al. 2004)	-
GRMZM2G300348	y1 yellow endosperm1	6	82017148	82021007	coloration of endosperm, grain yield, kernel number	molecular evolution, QTL mapping, association mapping	(Palaisa, Morgante et al. 2003) (Coque and Gallais 2006) (Fu, Cheng et al. 2013) (Wong, Lambert et al. 2004)	-
GRMZM2G172795	b1 colored plant1	2	18417370	18421827	flavonoid content, macrohairs density	QTL mapping	(Lauter, Gustus et al. 2004)	-
marker	umc63-csu25a	3	216256506	232072064	pigment coverage and intensity	QTL mapping	(Lauter, Gustus et al. 2004)	-
marker	umc126b-umc66a	4	144917962	162885280	pigment intensity, macrohairs	QTL mapping	(Lauter, Gustus et al. 2004)	-
marker	a3 -(anthocyanin3)	3	214225932	214967721	flavonoid content, macrohairs density	QTL mapping	(Lauter, Gustus et al. 2004)	-
GRMZM2G422750	c2 - colorless2	4	192580450	192583847	pigment intensity, insect resistance	QTL mapping	Lauter et al 2004 (Meyer, Snook et al. 2007)	-
<b>Plant architecture</b>								
GRMZM2G005624	grassy tillers1	1	23241091	23244476	Apical dominance	functional	(Whipple, Kebrom et al. 2011)	-
AC233950.1_FG002	tb1	1	265745979	265747712	Apical dominance /	QTL mapping,	(Briggs, McMullen et al.	-

					inflorescence trait / tassel branch number	functional, association mapping	2007) (Studer and Doebley 2011) (Weber, Briggs et al. 2008) (Doebley, Stec et al. 1995) (Clark, Nussbaum Wagler et al. 2006)	
AC217051.3_FG006	zen1, pebp5 phosphatidylethanol amine-binding protein	10	114029804	114030887	length of lateral branch	Association mapping	Weber et al 2007	-
GRMZM2G036340	d1 dwarf plant1	3	10451980	10454261	Plant height	NAM	(Peiffer, Romay et al. 2014)	-
marker	umc119a-umc33a	1	195068415	209513848	plant height	QTL mapping	(Malosetti, Ribaut et al. 2008)	SB_9 SB_10 SB_11
marker	bn15.46-bn15.71	4	17851597	118237095	plant height	QTL mapping	(Malosetti, Ribaut et al. 2008)	MS_16 SB_28 SB_29 SB_30 SB_31 SB_32 SB_33 Fst_39 Fst_40 Fst_41 Fst_42 Fst_43
marker	umc105a-bn13.06	9	18607113	20660303	plant height	QTL mapping	(Malosetti, Ribaut et al. 2008)	Fst_81
GRMZM2G300133	td1 thick tassel dwarf1	5	61670412	61673965	tassel branch number	association mapping	(Weber, Briggs et al. 2008)	-
GRMZM2G180190	zfl2 Zea floricaula leafy2	2	12641797	12644804	Tiller number	association mapping	(Weber, Briggs et al. 2008)	-

<b>Inflorescence architecture</b>	-	-	-	-	-	-	-	-
GRMZM2G003927	ra1 - ramosa1	7	110331505	110332253	branching tassel and ear	molecular evolution, association mapping	(Sigmon and Vollbrecht 2010)	-
GRMZM2G148693	zap1 zea apetala homolog1	2	235844921	235854324	ear length, fruitcase length	association mapping	(Weber, Briggs et al. 2008)	-
GRMZM2G026223	zag1 zag6 agamous-like61	1	4862047	4877625	ear shattering, ear length, leaf number, tassel branch number	association mapping	(Weber, Briggs et al. 2008)	-
AC233943.1_FG002	ra2 ramosa2	3	12882491	12883276	inflorescence architecture, ear structure, Percentage of yoked fruitcase	association mapping	(Weber, Briggs et al. 2008)	-
GRMZM5G862109	tasselseed6 ids1	1	292889740	292893983	lateral inflorescence branch number	association mapping	(Weber, Briggs et al. 2008)	-
GRMZM2G101511	tga1 teosinte glume architecture1	4	44508235	44512898	inflorescence trait	gene mapping, molecular evolution, functional	wang et al 2005 (Briggs, McMullen et al. 2007)	-
GRMZM2G160687	zag2 Zea AGAMOUS homolog2	3	137228194	137233945	inflorescence trait	QTL mapping	(Briggs, McMullen et al. 2007)	-
GRMZM5G803935	miR172 ts4 tasselseed4	3	144884032	144884741	sex determination, floral meristem growth	QTL mapping	(Briggs, McMullen et al. 2007)	-
GRMZM2G397518	ba, barren stalk 1	3	183084299	183085268	Inflorescence trait	QTL mapping, functional, association mapping	(Briggs, McMullen et al. 2007) (Gallavotti, Zhao et al. 2004)	-
GRMZM2G085113	te1 terminal ear1	3	165174146	165178071	inflorescence trait, femal ear length	QTL mapping, association mapping	(Weber, Briggs et al. 2008) (Briggs, McMullen et al. 2007), Weber et al 2007	-
GRMZM2G060216	lg2 liguleless2	3	176800215	176808889	inflorescence trait, leaf morphology	QTL mapping, association mapping	(Briggs, McMullen et al. 2007)	-

GRMZM2G022489	miR156 cg1 corngrass1	3	7773334	7774907	leaf initiation and floral architecture	functionnal	(Yu, Li et al. 2012) Tian, Bradbury et al. 2011)	-
marker	bnl9.44	8	90317173	95150703	Number of branch in lateral inflorescence	QTL mapping	(Doebley and Stec 1993)	-
marker	umc46	6	143461102	144482145	Number of branch in lateral inflorescence, Number of ears on lateral branches	QTL mapping	(Doebley and Stec 1993)	-
marker	umc42a	4	65702021	75548657	Number of cupules, hardness of outer glumes	QTL mapping	(Doebley and Stec 1993)	-
marker	umc110a	7	143169908	143389581	Number of cupules, hardness of outer glumes	QTL mapping	(Doebley and Stec 1993)	-
marker	umc60	3	181079164	181079889	Number of cupules, hardness of outer glumes	QTL mapping	(Doebley and Stec 1993)	-
marker	umc95	9	126933130	126933789	Number of ears on lateral branches	QTL mapping	(Doebley and Stec 1993)	-
marker	umc5a	2	186410808	187894598	Percentage pf male spikelet in primary inflorescences	QTL mapping	(Doebley and Stec 1993)	-
marker	umc104b	5	205440973	217012402	Percentage pf male spikelet in primary inflorescences	QTL mapping	(Doebley and Stec 1993)	SB_48
marker	bnl5.71-umc156	4	135333950	143061373	number of ears per plant	QTL mapping	(Malosetti, Ribaut et al. 2008)	-
marker	umc105a	9	16979346	18607113	number of ears per plant	QTL mapping	(Malosetti, Ribaut et al. 2008)	-
marker	umc38-umc95	9	124251796	126933130	number of ears per plant	QTL mapping	(Malosetti, Ribaut et al. 2008)	SB_71 Fst_85
marker	np1114-umc10	3	45519051	95402898	number of ears per plant, plant height	QTL mapping	(Malosetti, Ribaut et al. 2008)	MS_11 MS_12 MS_13 SB_21

marker	umc149a-bnl8.39	7	143552212	163011802	number of ears per plant, plant height	QTL mapping	(Malosetti, Ribaut et al. 2008)	MS_39 Fst_77 Fst_78
marker	umc46- umc132a(chk)	6	144482145	161024525	Number of ears, drought tolerance	QTL mapping	(Agrama and Moussa 1996)	Fst_66 Fst_67
marker	umc132a(chk)	6	161024525	161618104	Number of ears, drought tolerance, grain yield, plant height, number of ears per plant	QTL mapping	(Agrama and Moussa 1996) (Malosetti, Ribaut et al. 2008)	-
marker	qTBN1	1	194300000	250900000	tassel branch number	QTL mapping	chen et al 2014	-
marker	qTBN3	3	180800000	186300000	tassel branch number	QTL mapping	chen et al 2014	-
marker	qTBN4	4	183900000	190000000	tassel branch number	QTL mapping	chen et al 2014	-
marker	qTBN5	5	177200000	178000000	tassel branch number	QTL mapping	chen et al 2014	-
marker	qTBN7	7	37000000	38600000	tassel branch number	QTL mapping	chen et al 2014	-
marker	qTBN8	8	0	7600000	tassel branch number	QTL mapping	chen et al 2014	-
marker	qTBN9	9	44400000	74900000	tassel branch number	QTL mapping	chen et al 2014	-
marker	qKRN5	5	173500000	178300000	kernel row number	QTL mapping	chen et al 2014	-
marker	qEL4	4	182200000	188100000	ear length	QTL mapping	chen et al 2014	-
marker	qEL5	5	186900000	194000000	ear length	QTL mapping	chen et al 2014	-
GRMZM2G316366	mads19 (MADS- transcription factor 19)	5	177222141	177223931	gene candidate for tassel architecture	QTL mapping	chen et al 2014	-
GRMZM2G492156	mads78 (MADS- transcription factor 78)	5	177229408	177232968	gene candidate for tassel architecture	QTL mapping	chen et al 2014	-
GRMZM5G805008	bpd1 (BTB/POZ domain1)	7	37726780	37727688	gene candidate for tassel architecture	QTL mapping	chen et al 2014	-
GRMZM2G022489	corngrass	3	7773334	7774907	branch tassel			-
<b>Heat tolerance</b>								
GRMZM2G360681	hsp101 (heat-shock	6	160621048	160624717	kernel development and	functional	Nieto-Sotelo et al 2002	-

marker	protein 101) umc132 -npi280	6	157833157	161024525	germination and in response to heat stress kernel development and germination and in response to heat stress	gene mapping	Nieto-Sotelo et al 2002	-
<b>Cold tolerance</b>								
GRMZM2G036908	lip15 low temperature-induced protein15	6	24608442	24612226	chilling tolerance	functional	(Kusano, Berberich et al. 1995)	-
GRMZM2G079348	cat3 catalase3	4	240105016	240108983	chilling tolerance	functional	(Anderson, Prasad et al. 1994)	-
GRMZM2G115518	cdpk1	5	60801754	60808528	chilling tolerance	functional	(Berberich and Kusano 1997)	-
GRMZM2G333916	met1 DNA methyl transferase1	7	11127929	11134326	chilling tolerance	functional	(Steward, Kusano et al. 2000)	-
<b>Drought tolerance</b>								
GRMZM2G097457	pdk2 (pyruvate, orthophosphate dikinase2)	8	107003965	107014861	drought tolerance	association mapping	(Setter, Yan et al. 2010)	-
GRMZM2G110153	zmm16 (Zea mays MADS16)	3	171427820	171430412	drought tolerance	association mapping	(Setter, Yan et al. 2010)	-
marker	dbf1 DRE-binding protein 1	1	217260344	219195665	drought tolerance	functional	kizis pages 2002	-
marker	umc60-umc96	3	181079889	223879129	flowering time, number of ears, drought tolerance	QTL mapping	(Agrama and Moussa 1996)	SB_24 SB_25 SB_26 Fst_33 Fst_34
marker	bnl5.59a-umc58	1	183904444	194726580	flowering time, number of ears, drought tolerance	QTL mapping	(Agrama and Moussa 1996)	-

marker	umc140a	1	257875680	261722575	flowering time, number of ears, drought tolerance	QTL mapping	(Agrama and Moussa 1996)	Fst_11
marker	umc96	3	223879129	224227888	flowering time, number of ears, drought tolerance	QTL mapping	(Agrama and Moussa 1996)	-
marker	umc65a-umc62	6	104450983	164802199	flowering time, number of ears, drought tolerance	QTL mapping	(Agrama and Moussa 1996)	SB_52 Fst_64 Fst_65 Fst_66 Fst_67
marker	umc76a	1	27398858	27918358	Grain yield, drought tolerance	QTL mapping	(Agrama and Moussa 1996)	-
marker	bnl5.37b-umc96	3	168368471	223879129	Grain yield, drought tolerance	QTL mapping	(Agrama and Moussa 1996)	SB_24 SB_25 SB_26 Fst_32 Fst_33 Fst_34
marker	bnl6.25a-umc43	5	3024418	43661111	Grain yield, drought tolerance	QTL mapping	(Agrama and Moussa 1996)	SB_36 SB_37 SB_38 SB_39 Fst_47 Fst_48
marker	bnl5.24	5	205440973	217012402	Grain yield, drought tolerance	QTL mapping	(Agrama and Moussa 1996)	SB_48
marker	umc124a(chk)- umc93a	8	22380791	141956453	Grain yield, drought tolerance	QTL mapping	(Agrama and Moussa 1996)	MS_71 MS_72 MS_73 MS_74



								MS_75
								MS_76
								MS_77
								MS_78
								MS_79
								MS_80
								MS_81
								MS_82
								MS_83
								MS_84
								MS_85
								MS_86
								MS_87
								SB_57
								SB_58
								SB_59
								SB_60
								SB_61
								SB_62
								SB_63
								SB_64
								SB_65
								SB_66
								SB_67
								Fst_79
marker	umc120a	8	39465328	63258558	Grain yield, drought tolerance	QTL mapping	(Agrama and Moussa 1996)	MS_82
								MS_83
								MS_84
								MS_85
								MS_86
								SB_57

								SB_58
								SB_59
								SB_60
								SB_61
								SB_62
								Fst_79
AC205274.3_FG001	TIDP2793 / prp1 pathogenesis-related protein1)	8	52299255	52299758	water deficit stress	QTL mapping	Welcker, Sadok et al. 2011)	-
AC209208.3_FG002	pip1b plasma membrane intrinsic protein1	5	193306422	193309271	water deficit stress	QTL mapping	Welcker, Sadok et al. 2011)	-
marker	bnlg1347a	1	283187143	287386173	water deficit stress	QTL mapping	Welcker, Sadok et al. 2011)	-
marker	umc1422	2	13716641	14406715	water deficit stress	QTL mapping	Welcker, Sadok et al. 2011)	-
marker	bnlg1721	2	205842785	207194203	water deficit stress	QTL mapping	Welcker, Sadok et al. 2011)	-
marker	pza02792	5	18445046	19750433	water deficit stress	QTL mapping	Welcker, Sadok et al. 2011)	-
marker	umc1315	5	42302040	42303953	water deficit stress	QTL mapping	Welcker, Sadok et al. 2011)	-
marker	pza00270	5	142281997	151727279	water deficit stress	QTL mapping	Welcker, Sadok et al. 2011)	MS_37
								SB_47
								Fst_56
marker	umc1857	6	110292855	110293172	water deficit stress	QTL mapping	Welcker, Sadok et al. 2011)	-
marker	gpm913a	7	9777114	10620195	water deficit stress	QTL mapping	Welcker, Sadok et al. 2011)	-
marker	umc98b	7	118007174	120870241	water deficit stress	QTL mapping	Welcker, Sadok et al. 2011)	-
marker	gpm583	8	39115777	74024212	water deficit stress	QTL mapping	Welcker, Sadok et al. 2011)	MS_82
								MS_83
								MS_84
								MS_85
								MS_86
								MS_87
								SB_57
								SB_58
								SB_59

SB\_60  
 SB\_61  
 SB\_62  
 SB\_63  
 SB\_64  
 Fst\_79

marker	bnlg2181	8	133560163	133978830	water deficit stress	QTL mapping	Welcker, Sadok et al. 2011)	-
marker	uaz128	8	166498697	167142956	water deficit stress	QTL mapping	Welcker, Sadok et al. 2011)	-
marker	csu634	9	136553558	136553847	water deficit stress	QTL mapping	Welcker, Sadok et al. 2011)	-
marker	IDP7215	9	149778153	150843632	water deficit stress	QTL mapping	Welcker, Sadok et al. 2011)	-
marker	psr119c	10	2457855	3725625	water deficit stress	QTL mapping	Welcker, Sadok et al. 2011)	-
marker	gpm702b, umc117	8	162498870	162631540	water deficit stress, low-P tolerance, Number of ears on lateral branche	QTL mapping	Welcker, Sadok et al. 2011) (Reiter, Coors et al. 1991) ,(Doebley and Stec 1993)	-

**Disease resistance and herbivory tolerance**

GRMZM2G086773	hm2 Helminthosporium carbonum susceptibility2	9	103234471	103236102	ear rot disease resistance	evolution molecular	(Zhang, Peek et al. 2002)	-
GRMZM5G881887	hm1 Helminthosporium carbonum susceptibility1	1	199754467	199756329	ear rot disease resistance	evolution molecular	(Zhang, Peek et al. 2002)	-
marker	rhml resistance to Helminthosporium maydis1	6	2776439	3087809	fungus disease resistance		(Simmons, Grant et al. 2001)	-
marker	ht1 Helminthosporium turcicum resistance1	2	206643048	207194203	fungus disease resistance	linkage analyses	(Ogliari, Guimarães et al. 2005) (McMullen and Simcox 1995) (Bentolila, Guitton et al.	-

GRMZM2G012821		4	7616846	7618466	Fusarium ear rot disease resistance	association mapping	1991) (Zila, Ogut et al. 2014)	-
GRMZM2G035665		9	19532252	19535196	Fusarium ear rot resistance	association mapping	(Zila, Ogut et al. 2014)	-
GRMZM2G060659		5	64769600	64773966	Fusarium ear rot resistance	association mapping	(Zila, Ogut et al. 2014)	-
GRMZM2G106752		4	124927190	124931744	Fusarium ear rot resistance	association mapping	(Zila, Ogut et al. 2014)	-
GRMZM2G372364		4	7565889	7566669	Fusarium ear rot resistance	association mapping	(Zila, Ogut et al. 2014)	-
GRMZM2G419836		4	9351020	9354236	Fusarium ear rot resistance	association mapping	(Zila, Ogut et al. 2014)	-
GRMZM2G156632	wip1 (wound induced protein1)	8	12 273514	12 274222	defense against insect herbivore	evolution molecular	Moeller and Tiffin 2008	-
GRMZM2G023325	bx12 benzoxazinone synthesis12	1	66500692	66501706	insect resistance	QTL mapping, fonctionnal	(Meihls, Handrick et al. 2013)	-
GRMZM2G073465	ccp3 cysteine protease3	10	145479598	145483678	insect resistance	fonctionnal	(Tzin, Lindsay et al. 2015) (Pechan, Jiang et al. 1999) (van der Linde, Hemetsberger et al. 2012)	-
GRMZM2G026930	a1 anthocyaninless1	3	216304734	216306568	insect resistance	association mapping	(Szalma, Buckler Iv et al. 2005)	-
GRMZM2G063756	bx5 benzoxazinone synthesis5	4	3108397	3110973	insect resistance	molecular evolution	(Nomura, Ishihara et al. 2003)	-
GRMZM2G167549	bx3 benzoxazinone synthesis3	4	2998634	3001404	insect resistance	molecular evolution	(Nomura, Ishihara et al. 2003)	-
GRMZM2G172491	bx4 benzoxazinone synthesis4	4	3046687	3049508	insect resistance	molecular evolution	(Nomura, Ishihara et al. 2003)	-
GRMZM2G063756	bx5 benzoxazinone synthesis5	4	3108397	3110973	insect resistance	molecular evolution	(Nomura, Ishihara et al. 2003)	-
GRMZM2G085381	bx1 benzoxazinless1	4	3256234	3258478	insect resistance	molecular evolution, NAM	(Nomura, Ishihara et al. 2003)	-
GRMZM2G085661	bx2 benzoxazinone	4	3260685	3262840	insect resistance	molecular evolution	(Nomura, Ishihara et al.	-

	synthesis2						2003)	
GRMZM2G150256	mir2 maize insect resistance2	6	88903392	88908316	insect resistance	functionnal	(Butrón, Chen et al. 2010)	-
GRMZM2G150276	mir1 maize insect resistance1	6	88909372	88911308	insect resistance	functionnal, molecular evolution	(Jiang, Siregar et al. 1995, Pechan, Jiang et al. 1999)	-
GRMZM2G151227	whp1 white pollen1	2	223888706	223892691	insect resistance	QTL mapping, association mapping	(Meyer, Snook et al. 2007)	-
GRMZM2G166281	mir3 maize insect resistance3	2	5269563	5273858	insect resistance	functionnal	(Szalma, Buckler Iv et al. 2005)	-
GRMZM2G311036	bx10 benzoxazinone synthesis10	1	66304872	66306978	insect resistance	QTL mapping, functionnal	(van der Linde, Hemetsberger et al. 2012)	-
GRMZM2G336824	bx11 benzoxazinone synthesis11	1	66387644	66390084	insect resistance	QTL mapping, functionnal	(Meihls, Handrick et al. 2013)	-
GRMZM2G441753	bx7 benzoxazinone synthesis7	4	18538004	18626350	insect resistance	molecular evolution	(Tzin, Lindsay et al. 2015)	-
GRMZM6G617209	bx6 benzoxazinone synthesis6	4	4746444	5119868	insect resistance	molecular evolution	(Frey, Huber et al. 2003)	SB_28
marker	mir3a(thp)	9	8704533	9587949	insect resistance	molecular evolution	(Frey, Huber et al. 2003)	-
GRMZM2G057027	p2 pericarp color2	1	48092121	48097632	insect resistance, flavonoid content, leave coloration	molecular evolution, QTL maping	(Pechan, Jiang et al. 1999)	-
GRMZM2G084799	p1 pericarp color1	1	48117497	48128047	insect resistance, flavonoid content, leave coloration	molecular evolution, QTL maping	Zhang, Chopra et al. 2000)	-
AC214524.3_FG002	-	1	77966985	77969035	northern leaf blight disease resistance	NAM	(Meyer, Snook et al. 2007)	-
GRMZM2G006212	-	10	136615140	136617256	northern leaf blight disease resistance	NAM	Casati and Walbot 2005	-
GRMZM2G012224	-	1	290521787	290523917	northern leaf blight	NAM	Zhang et al., 2000,	-
							(Poland, Bradbury et al. 2011)	-
							(Poland, Bradbury et al. 2011)	-
							(Poland, Bradbury et al.	-

					disease resistance		2011)	
GRMZM2G014089	-	8	152715030	152719617	northern leaf blight disease resistance	NAM	(Poland, Bradbury et al. 2011)	-
GRMZM2G016443	-	1	16436083	16436786	northern leaf blight disease resistance	NAM	(Poland, Bradbury et al. 2011)	-
GRMZM2G019317	-	5	204589912	204594285	northern leaf blight disease resistance	NAM	(Poland, Bradbury et al. 2011)	-
GRMZM2G022822	-	1	94423347	94425952	northern leaf blight disease resistance	NAM	(Poland, Bradbury et al. 2011)	-
GRMZM2G043600	bzip22 (bZIP-transcription factor 22)	7	140706875	140710245	northern leaf blight disease resistance	NAM	(Poland, Bradbury et al. 2011)	
GRMZM2G051974	mlo9 (barley mlo defense gene homolog9)	2	1270643	1276180	northern leaf blight disease resistance	NAM	(Poland, Bradbury et al. 2011)	-
GRMZM2G056903	-	5	191625428	191627936	northern leaf blight disease resistance	NAM	(Poland, Bradbury et al. 2011)	-
GRMZM2G057386	ereb107 (AP2-EREBP-transcription factor 107)	5	191375628	191377153	northern leaf blight disease resistance	NAM	(Poland, Bradbury et al. 2011)	-
GRMZM2G059740	-	2	4512059	4514108	northern leaf blight disease resistance	NAM	(Poland, Bradbury et al. 2011)	-
GRMZM2G076423	-	1	12185737	12189286	northern leaf blight disease resistance	NAM	(Poland, Bradbury et al. 2011)	-
GRMZM2G080041	-	2	9441323	9445539	northern leaf blight disease resistance	NAM	(Poland, Bradbury et al. 2011)	-
GRMZM2G080652	-	1	183908459	183912375	northern leaf blight disease resistance	NAM	(Poland, Bradbury et al. 2011)	-
GRMZM2G083759	-	2	181972542	181978925	northern leaf blight disease resistance	NAM	(Poland, Bradbury et al. 2011)	-
GRMZM2G087201	-	2	3859120	3863795	northern leaf blight disease resistance	NAM	(Poland, Bradbury et al. 2011)	-
GRMZM2G090981	-	5	205230215	205235779	northern leaf blight disease resistance	NAM	(Poland, Bradbury et al. 2011)	-

GRMZM2G107562	-	7	159130478	159136959	northern leaf blight disease resistance	NAM	(Poland, Bradbury et al. 2011)	-
GRMZM2G114190	-	1	21702392	21710565	northern leaf blight disease resistance	NAM	(Poland, Bradbury et al. 2011)	-
GRMZM2G117335	-	6	138503486	138505007	northern leaf blight disease resistance	NAM	(Poland, Bradbury et al. 2011)	-
GRMZM2G127990	-	9	12618868	12622153	northern leaf blight disease resistance	NAM	(Poland, Bradbury et al. 2011)	-
GRMZM2G129083	-	5	13726531	13727525	northern leaf blight disease resistance	NAM	(Poland, Bradbury et al. 2011)	-
GRMZM2G129140	-	5	13723734	13724402	northern leaf blight disease resistance	NAM	(Poland, Bradbury et al. 2011)	-
GRMZM2G145019	-	8	154290192	154292856	northern leaf blight disease resistance	NAM	(Poland, Bradbury et al. 2011)	-
GRMZM2G153075	-	6	147900966	147905245	northern leaf blight disease resistance	NAM	(Poland, Bradbury et al. 2011)	-
GRMZM2G163724	-	2	163676823	163680091	northern leaf blight disease resistance	NAM	(Poland, Bradbury et al. 2011)	-
GRMZM2G169458	-	5	190228137	190232283	northern leaf blight disease resistance	NAM	(Poland, Bradbury et al. 2011)	-
GRMZM2G169558	-	5	190236078	190237670	northern leaf blight disease resistance	NAM	(Poland, Bradbury et al. 2011)	-
GRMZM2G173771	-	4	178590006	178591058	northern leaf blight disease resistance	NAM	(Poland, Bradbury et al. 2011)	-
GRMZM2G177046	bzip26 (bZIP-transcription factor 26)	1	281333729	281334681	northern leaf blight disease resistance	NAM	(Poland, Bradbury et al. 2011)	-
GRMZM2G348956	-	2	1277292	1281626	northern leaf blight disease resistance	NAM	(Poland, Bradbury et al. 2011)	-
GRMZM2G349665	-	6	115900740	115904237	northern leaf blight disease resistance	NAM	(Poland, Bradbury et al. 2011)	-
GRMZM2G400938	-	3	3421920	3424963	northern leaf blight disease resistance	NAM	(Poland, Bradbury et al. 2011)	-
GRMZM2G464985	-	7	130878675	130882596	northern leaf blight disease resistance	NAM	(Poland, Bradbury et al. 2011)	-

GRMZM2G469224	-	1	265591328	265593341	northern leaf blight disease resistance	NAM	(Poland, Bradbury et al. 2011)	-
GRMZM2G164612	htn1 Helminthosporium turcicum resistanceN1	8	152102260	152107347	northern leaf blight disease resistance	gene mapping	Ogliari et al., 2005 (Simcox and Bennetzen 1993)	-
marker	ht2 Helminthosporium turcicum resistance2	8	135858398	136042717	northern leaf blight disease resistance	QTL mapping	(Ogliari, Guimarães et al. 2005  (Yin, Wang et al. 2003) cité dans Poland et al. 2011 (Chung, Jamann et al. 2010)	-
AC230011.2_FG002	rp3 resistance to Puccinia sorghi3	3	113896010	113902317	resistance to common rust disease	QTL mapping	(McMullen and Simcox 1995)	-
marker	rp4 resistance to Puccinia sorghi4	4	3589042	4130278	resistance to common rust disease	QTL mapping	(McMullen and Simcox 1995)	-
marker	rpp9 resistance to Puccinia polysora and Puccinia sorghi9	10	4484179	4736989	resistance to Southern rust disease	QTL mapping	(McMullen and Simcox 1995)	-
AC152495.1_FG002	rp1 resistance to Puccinia sorghi1	10	3280568	3284464	resistance to common rust disease	QTL mapping	(McMullen and Simcox 1995)	-
marker	scmv2 -(resistance to sugarcane mosaic virus)	3	126576367	168366251	resistance to sugarcane mosaic virus1 disease	QTL mapping	Xia et al 1999 Quint et al 2002	Fst_31
marker	scmv1 -(resistance to sugarcane mosaic virus1)	6	263968	81880425	resistance to sugarcane mosaic virus1 disease	QTL mapping	(Xia, Melchinger et al. 1999) Quint, Mihaljevic et al. 2002)	MS_38 SB_49 Fst_59 Fst_60 Fst_61 Fst_62 Fst_63
GRMZM2G424053	trh1 thioredoxin h homolog1	6	157770836	157775279	candidate for smcv1	QTL mapping	Tao et al 2013	-
GRMZM2G116204	abp1 - auxin binding protein1	3	133888888	133893638	candidate for smcv2	QTL mapping	Leng et al 2015	-



marker	QTL interval	5	2136100	2501321	response to herivory and abiotic stress	NAM	(Yan, Lipka et al. 2014)	-
marker	QTL interval	7	1291946	2066409	response to herivory and abiotic stress	NAM	(Yan, Lipka et al. 2014)	-
AC186406.4_FG006	mdm1 maize dwarf mosaic virus resistance1	6	14951256	14953843	virus resistance	molecular mapping,QTL mapping	(McMullen and Louie 1989) (Xia, Melchinger et al. 1999)	-
marker	msv1 maize streak virus tolerance1	1	64940737	67474234	virus resistance	QTL mapping, gene mapping	(Welz, Xia et al. 1999) (Kyetere, Ming et al. 1999)	-
marker	wsm2 wheat streak mosaic virus resistance2	3	127609718	128902878	virus resistance	molecular mapping,QTL mapping	(McMullen, Jones et al. 1994) (Xia, Melchinger et al. 1999)	-
marker	wsm1 wheat streak mosaic virus resistance1	6	7065267	28440201	virus resistance	molecular mapping,QTL mapping	(McMullen, Jones et al. 1994) (Xia, Melchinger et al. 1999)	-
marker	wsm3 wheat streak mosaic virus resistance3	10	132121673	132739546	virus resistance	molecular mapping,QTL mapping	(McMullen, Jones et al. 1994) (Xia, Melchinger et al. 1999)	-
marker	mv1 resistance to maize mosaic virus I ("corn stripe")1)	3	136086344	137548200	virus resistance	molecular mapping,	(Ming, Brewbaker et al. 1997)	-
<b>Yield</b>								
marker	umc167a-csu61b	1	154138515	180709507	grain yield	QTL mapping	(Malosetti, Ribaut et al. 2008)	MS_1 MS_2 MS_3 SB_5 SB_6 SB_7

marker	umc313-umc50	3	7830037	56824250	grain yield	QTL mapping	(Malosetti, Ribaut et al. 2008)	Fst_6 MS_9 MS_10 SB_20 Fst_24 Fst_25 Fst_26
marker	umc53a-umc6a	2	4190624	14781166	grain yield, plant height	QTL mapping	(Malosetti, Ribaut et al. 2008)	Fst_13
marker	umc330-npi593	4	204706420	235644775	grain yield, days to male flowering	QTL mapping	(Malosetti, Ribaut et al. 2008)	MS_25 SB_34
<b>Kernel composition</b>								
GRMZM2G015534	o2 - opaque endosperm2	7	10793341	10796233	kernel composition	NAM, molecular analysis	(Hirsch, Flint-Garcia et al. 2014)	-
GRMZM2G024993	wx1 (waxy1)	9	23256308	23260236	kernel composition	association mapping and functional	(Wilson, Whitt et al. 2004)	-
GRMZM2G124524	wri1 (WRI1 transcription factor1)	2	149336313	149342184	kernel composition	functional	(Pouvreau, Baud et al. 2011)	-
GRMZM2G146283	pbfl (prolamin-box binding factor1)	2	153510052	153513609	kernel composition	functional	(Vicente-Carbajosa, Moose et al. 1997)	-
GRMZM2G169089	DGAT1-2, ln1 (linoleic acid1)	6	104859135	104866019	kernel composition	NAM	(Cook, McMullen et al. 2012)	-
GRMZM2G379656	glk45 - G2-like-transcription factor 45	1	234417882	234419889	kernel composition	NAM, molecular analysis	(Hirsch, Flint-Garcia et al. 2014)	-
GRMZM2G397687	(fl2 - floury2)	4	21312310	21313379	kernel composition	NAM, molecular analysis	(Hirsch, Flint-Garcia et al. 2014)	-
GRMZM2G568636	nnr1 (nitrate reductase(NADH)1)	4	56308168	56313225	kernel composition	NAM	(Cook, McMullen et al. 2012)	-
GRMZM2G138060	su1 sugary1	4	41369510	41378299	kernel content	association mapping, molecular evolution	(Weber, Briggs et al. 2008) (Tracy, Whitt et al. 2006)	SB_30

GRMZM2G068506	bt2 - brittle endosperm2	4	58954361	58960521	starch content	molecular evolution and functional	(Whitt, Wilson et al. 2002)	-
GRMZM2G032628	ae1	5	168451664	168468750	starch content	association and functional	(Wilson, Whitt et al. 2004)	-
GRMZM2G089713	sh1	9	11496011	11502772	starch content	association mapping and functional	(Wilson, Whitt et al. 2004)	-
GRMZM2G429899	sh2 - shrunken2	3	216414684	216424048	starch content	functional	(Bhave, Lawrence et al. 1990)	SB_26
GRMZM2G009785	vte1, sxd1 (sucrose export defective1)	5	133501928	133518495	tochromanol antioxydants, vitamine E	association mapping	(Lipka, Gore et al. 2013)	-
GRMZM2G031952	-	5	190734008	190738904	tochromanol antioxydants, vitamine E	association mapping	(Lipka, Gore et al. 2013)	-
GRMZM2G035213	vte4 (vitamin E synthesis4)	5	200367029	200370851	tochromanol antioxydants, vitamine E	association mapping	(Lipka, Gore et al. 2013)	-
GRMZM2G058943	-	5	198206906	198208665	tochromanol antioxydants, vitamine E	association mapping	(Lipka, Gore et al. 2013)	-
GRMZM2G105494	-	5	133330912	133339332	tochromanol antioxydants, vitamine E	association mapping	(Lipka, Gore et al. 2013)	-
GRMZM2G108996	-	5	200107274	200115056	tochromanol antioxydants, vitamine E	association mapping	(Lipka, Gore et al. 2013)	-
GRMZM2G161641	-	5	204062211	204065565	tochromanol antioxydants, vitamine E	association mapping	(Lipka, Gore et al. 2013)	-
GRMZM2G167431	-	5	200379998	200382416	tochromanol antioxydants, vitamine E	association mapping	(Lipka, Gore et al. 2013)	-
GRMZM2G173358	hggt1 (homogentisate geranylgeranyl transferase1)	9	92483549	92487268	tochromanol antioxydants, vitamine E	association mapping	(Lipka, Gore et al. 2013)	-
GRMZM2G325001	-	5	200429608	200433476	tochromanol antioxydants, vitamine E	association mapping	(Lipka, Gore et al. 2013)	-
GRMZM2G325019	otp51 (organelle transcript processing51)	5	200434762	200438299	tochromanol antioxydants, vitamine E	association mapping	(Lipka, Gore et al. 2013)	-
GRMZM2G325038	-	5	200438375	200439307	tochromanol antioxydants, vitamine E	association mapping	(Lipka, Gore et al. 2013)	-

GRMZM5G823157	wrky46 (WRKY-transcription factor 46)	5	199646796	199650458	tochromanol antioxydants, vitamine E	association mapping	(Lipka, Gore et al. 2013)	-
GRMZM5G833760	-	9	92696086	92716980	tochromanol antioxydants, vitamine E	association mapping	(Lipka, Gore et al. 2013)	-
marker	pdk1 (pyruvate, orthophosphate dikinase1)	6	145987055	145997616	lysine content and the protein-versus-starch balance	association mapping	Manicacci, D, et al. 2009	-
<b>Leaf architecture</b>								
GRMZM2G047129	-	3	213800453	213804523	leaf length	Expression	(Yu, Li et al. 2012)	-
GRMZM2G129777	(ereb192 - AP2-EREBP-transcription factor 192)	3	22387754	22389768	leaf length	association mapping	(Yu, Li et al. 2012)	-
GRMZM2G020187	rgd1 - ragged seedling1	6	17165362	17171793	leaf length and leaf width	association mapping	(Yu, Li et al. 2012)	-
GRMZM2G028041	rs1 - rough sheath1	7	3664698	3670891	leaf morphology	association mapping	(Yu, Li et al. 2012)	-
GRMZM2G083812	-	5	214166462	214170119	leaf width	Expression	(Yu, Li et al. 2012)	-
GRMZM2G417843	pco102265b	1	299197998	299202494	leaf width	Expression	(Yu, Li et al. 2012)	-
AC217051.3_FG006	zen1, pebp5 phosphatidylethanol amine-binding protein	10	114029804	114030887	length of lateral branch	Association mapping	Weber et al 2007	-
GRMZM2G145041	ncsu4(cca1) circadian clock associated1 ortholog	5	64117509	64120689	leaf number	association mapping	(Weber, Briggs et al. 2008)	-
GRMZM2G107101	gil gigantea1	8	21098977	21107262	leaf number, tassel branch number	association mapping	(Weber, Briggs et al. 2008)	-
marker	umc107b-bnl4,36	5	10207280	78362776	macrohairs	QTL mapping	(Lauter, Gustus et al. 2004)	-
marker	umc192-umc95	9	98531838	126933789	macrohairs	QTL mapping	(Lauter, Gustus et al. 2004)	-
marker	umc64-umc259a	10	93081741	127514254	macrohairs	QTL mapping	(Lauter, Gustus et al. 2004)	-
GRMZM2G136887	smh1 - single myb	8	140378673	140386322	macrohairs	QTL mapping	(Lauter, Gustus et al. 2004)	-

	histone1							
marker	mhl1 (macrohairless1)	9	95742689	121585452	macrohairs density	QTL mapping	Moose, Lauter and Carlson 2004	-
<b>Nutrient uptake</b>								
marker	umc42b	3	86754041	118147511	Low phosphorus tolerance	QTL mapping	(Reiter, Coors et al. 1991)	MS_14 MS_15 SB_22 SB_23 Fst_27 Fst_28 Fst_29
marker	umc19	4	173459597	173459948	low-phosphorus tolerance	QTL mapping	(Reiter, Coors et al. 1991)	-
marker	umc59f	6	96493617	97044869	low-phosphorus tolerance	QTL mapping	(Reiter, Coors et al. 1991)	-
marker	umc46	6	143461102	144482145	low-phosphorus tolerance	QTL mapping	(Reiter, Coors et al. 1991)	-
GRMZM2G146720	-	4	186038693	186040522	Male flowering time	Association mapping, molecular evolution	(Weber, Zhao et al. 2009) (Wright, Bi et al. 2005)	-
marker	bnlg389	5	205440973	217012402	Nitrogen use efficiency	QTL mapping	(Coque and Gallais 2006)	SB_48
marker	bnlg2295	1	80183152	80752921	Nitrogen use efficiency	QTL mapping	(Bertin and Gallais 2001) (Coque and Gallais 2006)	-
marker	umc1035	1	195430362	195858518	Nitrogen use efficiency	QTL mapping	(Bertin and Gallais 2001) (Coque and Gallais 2006)	SB_9
marker	umc1335	1	197251378	197404358	Nitrogen use efficiency	QTL mapping	(Bertin and Gallais 2001) (Coque and Gallais 2006)	SB_10
marker	bnlg1175	2	42409105	42409822	Nitrogen use efficiency	QTL mapping	(Bertin and Gallais 2001) (Coque and Gallais 2006)	-
marker	bnlg426	6	27385922	28248280	Nitrogen use efficiency	QTL mapping	Bertin and Gallais (2001) (Coque and Gallais 2006)	-
GRMZM2G046804	gpc1 glyceraldehyde-3- phosphate dehydrogenase1	4	36880092	36884877	Nitrogen use efficiency, water logging tolerance, plant heigth	QTL mapping	Bertin and Gallais (2001), (Coque and Gallais 2006) Qui et al.	-
marker	bnlg2291	4	162885280	166915798	Nitrogen use efficiency	QTL mapping	(Coque and Gallais 2006)	-

marker	bnlg2248	2	22753277	28400591	Nitrogen use efficiency	QTL mapping	(Coque and Gallais 2006)	-
marker	umc1079-bnlg371	2	152207394	163405828	Nitrogen use efficiency	QTL mapping	(Coque and Gallais 2006)	-
marker	umc1580a-bnlg1175	2	40894003	42409822	Nitrogen use efficiency	QTL mapping	(Coque and Gallais 2006)	SB_16
marker	umc1653	6	165841538	166791275	Nitrogen use efficiency	QTL mapping	(Coque and Gallais 2006)	-
marker	umc1060-dupssr10	5	136341892	142392213	Nitrogen use efficiency	QTL mapping	(Coque and Gallais 2006)	Fst_55
marker	bnlg1601-mm0022	3	129198395	143222128	Nitrogen use efficiency	QTL mapping	(Coque and Gallais 2006)	-
marker	bnlg1401	9	18561278	18562043	Nitrogen use efficiency	QTL mapping	(Coque and Gallais 2006)	-
marker	bnlg2082	8	34593113	60338399	nitrogen use efficiency	QTL mapping	(Coque and Gallais 2006)	MS_82
								MS_83
								MS_84
								MS_85
								MS_86
								SB_57
								SB_58
								SB_59
								SB_60
								SB_61
								SB_62
								Fst_79
marker	bnlg666	8	133560163	133978830	nitrogen use efficiency	QTL mapping	(Coque and Gallais 2006)	-
marker	bnlg1940	2	219772895	221618693	nitrogen use efficiency	QTL mapping	(Coque and Gallais 2006)	-
marker	umc1366	9	146205555	146206480	nitrogen use efficiency	QTL mapping	(Coque and Gallais 2006)	-
marker	dupssr23	3	167744313	167747632	Nitrogen use efficiency	QTL mapping	(Coque and Gallais 2006)	-
marker	mmc0481-bnlg278	5	183069354	192286935	Nitrogen use efficiency	QTL mapping	(Coque and Gallais 2006)	Fst_57
marker	ssu2 ribulose biphosphate carboxylase small subunit2	2	142794871	142796060	Nitrogen use efficiency	QTL mapping	(Coque and Gallais 2006)	-
marker	frk2 -(fructokinase2)	6	1	499542	Nitrogen use efficiency	QTL mapping	(Coque and Gallais 2006)	-
marker	DUPSSR26- BNLG2295	1	80766478	80768230	Nitrogen use efficiency	QTL mapping	(Coque and Gallais 2006)	-
GRMZM2G046804	gpc1	4	36880092	36884877	Nitrogen use efficiency,	QTL mapping	Bertin and Gallais (2001)	-

	glyceraldehyde-3-phosphate dehydrogenase1				water logging tolerance, plant height		(Coque and Gallais 2006) Qui et al.	
GRMZM2G412760	-	3	20173885	20177370	root moprhology, low p tolerance	MetaAnalysis QTL	(Azevedo, Cheavegatti-Gianotto et al. 2015)	-
GRMZM2G448672	-	3	206920286	206928325	root moprhology, low p tolerance	MetaAnalysis QTL	(Azevedo, Cheavegatti-Gianotto et al. 2015)	-
AC193632.2_FG002	-	4	39794587	39796928	root moprhology, low p tolerance	MetaAnalysis QTL	(Azevedo, Cheavegatti-Gianotto et al. 2015)	-
GRMZM2G172396	-	8	13264177	13270435	root moprhology, low p tolerance	MetaAnalysis QTL	(Azevedo, Cheavegatti-Gianotto et al. 2015)	-
GRMZM2G451147	-	8	152043185	152046899	root moprhology, low p tolerance	MetaAnalysis QTL	(Azevedo, Cheavegatti-Gianotto et al. 2015)	-
GRMZM2G164612	-	8	152102260	152107347	root moprhology, low p tolerance	MetaAnalysis QTL	(Azevedo, Cheavegatti-Gianotto et al. 2015)	-
marker	PZA00766_1	8	116800000	152000000	root moprhology, low p tolerance	MetaAnalysis QTL	(Azevedo, Cheavegatti-Gianotto et al. 2015)	-
marker	umc1073	1	27500000	44500000	root moprhology, low p tolerance	MetaAnalysis QTL	(Azevedo, Cheavegatti-Gianotto et al. 2015)	-
marker	<i>ZmPSTOL4.05</i>	4	33900000	59400000	root moprhology, low p tolerance	MetaAnalysis QTL	(Azevedo, Cheavegatti-Gianotto et al. 2015)	-
marker	PZA01690_7	7	98100000	123600000	root moprhology, low p tolerance	MetaAnalysis QTL	(Azevedo, Cheavegatti-Gianotto et al. 2015)	-
marker	PHM2770_19	10	62100000	77500000	root moprhology, low p tolerance	MetaAnalysis QTL	(Azevedo, Cheavegatti-Gianotto et al. 2015)	-
marker	PHM12693_8	1	203700000	240600000	root moprhology, low p tolerance	MetaAnalysis QTL	(Azevedo, Cheavegatti-Gianotto et al. 2015)	-
marker	PZA02212_1	3	165800000	216000000	root moprhology, low p tolerance	MetaAnalysis QTL	(Azevedo, Cheavegatti-Gianotto et al. 2015)	-
marker	cQTL1-1	1	11610000	14130000	ear diameter, row number,		Zhang et al 2014	-

marker	cQTL1-2	1	33160000	35740000	ear length kernel weight, grain yield kernel weight, kernel number per row, grain yield, ration ear-height to plant height	Zhang et al 2014	-
marker	cQTL1-3	1	205790000	210360000	relative fibrous root number, relative tio soil root dry weight, phosphorus efficiency, plant height	Zhang et al 2014	-
marker	cQTL2-1	2	28690000	30520000	floraison time, chlorophyll level ration ear-height to plant height, kernel weight, seminal root number, lateral root number plasticity, seminal root number	Zhang et al 2014	-
marker	cQTL2-2	2	48000000	48590000	kernel number per row, relative fibrous root number, plant height	Zhang et al 2014	-
marker	cQTL2-3	2	66240000	119270000	leaf length lateral root number, relative above-ground biomass dry weight, tolerance low phosphorus stress, rate of [H+] under low	Zhang et al 2014	-



marker	cQTL3-1	3	18630000	31990000	and normal-P, leaf area relative above-ground biomass dry weight leaf length rate od acid phosphatase activity	Zhang et al 2014	-
marker	cQTL3-2	3	212290000	213680000	plant height relative fibrous root number, tolerance low phosphorus stress rate of [H+] under low and normal-P,	Zhang et al 2014	-
marker	cQTL4-1	4	4810000	4980000	grain yield chlorophyll level percentage of colonized roots	Zhang et al 2014	-
marker	cQTL4-2	4	171670000	178110000	kernel weigth, leaf area, rate od acid phosphatase activity	Zhang et al 2014	-
marker	cQTL4-3	4	233690000	237710000	phosphorus efficiency, leaf length	Zhang et al 2014	-
marker	cQTL5-1	5	120680000	147230000	ear height, grain yield, rate od acid phosphatase activity	Zhang et al 2014	-
marker	cQTL5-2	5	167270000	167540000	relative above-ground biomass dry weight, leaf length, tolerance low phosphorus stress	Zhang et al 2014	-
marker	cQTL5-3	5	175330000	180150000	plant height, EN,	Zhang et al 2014	-

marker	cQTL5-4	5	199740000	202540000	row number, tolerance low phosphorus stress, relative tio soil root dry weight, relative root dry weight, phosphorus efficiency taproot length fibrous root number, fibrous root number, top sopsoil root dry weight,	Zhang et al 2014	-
marker	cQTL5-5	5	207160000	208070000	tolerance low phosphorus stress, relative fibrous root number, relative root dry weight, relative tio soil root dry weight, phosphorus efficiency	Zhang et al 2014	-
marker	cQTL5-6	5	212970000	214370000	kernel number phosphorus efficiencyr row, ear length kernel weight, taproot length, phosphorus efficiency	Zhang et al 2014	-
marker	cQTL6-1	6	150760000	156160000	relative above-ground biomass dry weight, shoot phosphorus utilization efficiency	Zhang et al 2014	-
marker	cQTL6-2	6	162230000	163820000	chlorophyll levear length root weight, phosphorus absorption efficienty	Zhang et al 2014	-

marker	cQTL9-1	9	16050000	17540000	grain yield, relative root dry weight	Zhang et al 2014	-
					relative to soil root dry weight		
marker	cQTL9-2	9	24990000	27310000	leaf area	Zhang et al 2014	-
					leaf width		
					leaf width		
marker	cQTL10-1	10	113850000	115160000	plant height, shoot phosphorus utilization efficiency, phosphorus utilization efficiency	Zhang et al 2014	-
marker	cQTL10-2	10	147710000	148840000	ear diameter	Zhang et al 2014	-
					kernel number per row		
					row number		
					phosphorus utilization efficiency		
					tolerance low phosphorus stress		
					relative above-ground biomass dry weight		
GRMZM2G037650	myb167 (MYB-transcription factor 167)	1	207176666	207182370	Encode an MYB transcription factor; induced by low P	Dai et al. 2012	-
GRMZM2G045473	pht3 (phosphate transporter protein3)	2	98751587	98753456	Encode high-affinity Pi transporters; induced by low P	Shin et al. 2004	-
GRMZM2G170208	-	2	99175457	99177363	Encode high-affinity Pi transporters	Shin et al. 2004	-
					; induced by low P		
GRMZM2G009779	-	2	99255085	99256744	Encode high-affinity Pi transporters; induced by low P	Shin et al. 2004	-
GRMZM2G009800	-	2	99298095	99299848	Encode high-affinity Pi	Shin et al. 2004	-

GRMZM5G829946	-	2	66712173	66712872	transporters; induced by low P Encode glycerophosphodiester phosphodiesterase; induced by low P	Cheng et al. 2011	-
GRMZM2G467671	-	3	19965649	19969019	Encode a protein kinase; induced by low P	Gamuyao et al. 2012	-
GRMZM2G094028	-	3	20059223	20068196	Encode a protein kinase; induced by low P	Gamuyao et al. 2012	-
GRMZM2G412760	-	3	20173885	20177370	Encode a protein kinase; induced by low P	Gamuyao et al. 2012	-
GRMZM2G426917	-	3	27772592	27775818	Encode a protein kinase; induced by low P	Gamuyao et al. 2012	-
GRMZM2G162709	myb137 (MYB- transcription factor 137)	3	23683344	23685084	Encode an MYB transcription factor; induced by low P	Baek et al. 2013	-
GRMZM2G054050	pza02427	3	24352754	24356876	Encode a multicopper oxidase; minor response to low P	Svistoonoff et al. 2007	-
GRMZM2G086727	-	3	24473522	24475933	Encode a multicopper oxidase; minor response to low P	Svistoonoff et al. 2007	-
GRMZM2G466545	-	4	171917079	171922792	Encode a member of SPX family genes; induced by low P	Stefanovic et al. 2007	-
Non-coding RNA	<i>miR399h;miR399i</i>				Non-coding RNA; induced by low P	Fujii et al. 2005	-
GRMZM2G081589	sut* GRMZM2G081589 (sucrose transporter GRMZM2G081589)	5	176487861	176491184	Encode a sucrose transporter; uncertain response to low P	Zakhleniuk et al. 2001	-
GRMZM2G092780	-	5	179926810	179929966	Encode a low-affinity	Versaw and Harrison 2002	-

GRMZM2G178892	-	5	214327389	214334953	transporter; minor response to low P Encode an MGDG synthase;		Kobayashi et al. 2009	-	
GRMZM2G169966	wrky57 (WRKY-transcription factor 57)	6	153264614	153267678	two member of this family are induced by low P Encode a WRKY transcription factor; induced by low P		Devaiah et al. 2007	-	
GRMZM2G099481	-	6	152079566	152085327	Encode phosphatidate phosphohydrolase; induced by low P		Nakamura et al. 2009	-	
GRMZM5G831009	-	10	113923825	113925962	Encode purple acid phosphatases; induced by low P		Wang et al. 2011	-	
<b>Inversion Inv1n</b>									
	inversion	1	65000000	115000000	no specific phenotype	molecular evolution study	(Fang, Pyhäjärvi et al. 2012) (Pyhäjärvi, Hufford et al. 2013)	SB_3 SB_4	
<b>Water logging tolerance</b>									
GRMZM2G067919	glb1 - globulin1	1	258363413	258366718	Plant height root length water logging tolerance	QTL mapping	qiu et al 2007	-	
GRMZM2G102356	(prp2 - pathogenesis-related protein2) phi083	2	40588194	40589174	Plant height root length water logging tolerance		Qiu et al 2007 (Malosetti, Ribaut et al. 2008)	-	
GRMZM2G123633	(incw3 - invertase cell wall3)	10	114288173	114290565	Plant height root length water logging tolerance	QTL mapping	qiu et al 2007	-	

marker	bnlg1808-phi034	7	129831204	50173540	Plant height root length water logging tolerance	QTL mapping	qiu et al 2007	-
marker	umc1748	1	191860023	191860425	Plant height root length water logging tolerance	QTL mapping	qiu et al 2007	-
marker	bnlg1556	1	207514669	209868319	Plant height root length water logging tolerance	QTL mapping	qiu et al 2007	-
marker	umc1265	2	5514764	5515120	Plant height root length water logging tolerance	QTL mapping	qiu et al 2007	-
marker	umc1117	4	25232841	25233582	Plant height root length water logging tolerance	QTL mapping	qiu et al 2007	-
marker	bnlg292b	4	213657288	214143007	Plant height root length water logging tolerance	QTL mapping	qiu et al 2007	-
marker	umc1231	9	131066648	133267815	Plant height root length water logging tolerance	QTL mapping	qiu et al 2007	-
marker	phi063	10	5793876	6213193	Plant height root length water logging tolerance	QTL mapping	qiu et al 2007	-
marker	bnlg1643	1	235531663	235987846	Plant height water logging tolerance grain yield	QTL mapping	qiu et al 2007 (Coque and Gallais 2006)	-
marker	bnlg1647	3	8211240	8211908	Root length water logging tolerance	QTL mapping	qiu et al 2007	-
marker	umc1025	3	27733976	29838475	Root length water logging tolerance	QTL mapping	qiu et al 2007	-

marker	umc1399	3	198386813	198387603	Root length water logging tolerance	QTL mapping	qiu et al 2007	-
marker	mmc0501	10	6032853	6409779	Root length water logging tolerance	QTL mapping	qiu et al 2007	-
GRMZM2G096475	mlg3 - lea protein group3 umc1063	6	161966268	161967646	water logging tolerance	QTL mapping	qiu et al 2007	-
marker	umc1772	3	14790768	15281332	water logging tolerance	QTL mapping	qiu et al 2007	-
marker	umc1983	7	105728832	108560688	water logging tolerance	QTL mapping	qiu et al 2007	-
marker	umc1519	9	126906736	130624535	water logging tolerance	QTL mapping	qiu et al 2007	SB_72 Fst_86 Fst_87
marker	bnlg1305	7	131009736	131984320	water logging tolerance	QTL mapping	qiu et al 2007	-
marker	sdw1 -(semi-dwarf plant1)	8	135858398	136042717	water logging tolerance	QTL mapping	qiu et al 2007	-
GRMZM2G152908	sus1 sucrose synthase1	9	122220190	122226863	water flood	functionnal	(Subbaiah and Sachs 2001)	-
GRMZM2G318780	sus2 sucrose synthase2	1	56785051	56792250	water flood	molecular evolution, gene expression	(Carlson, Chourey et al. 2002) Chourey 2006 in MaizeBook	-
marker	hb1 hemoglobin1	9	145303921	146261506	water flood	functionnal	(Dordas, Hasinoff et al. 2003)	MS_90 MS_91 SB_79 SB_80

<sup>1</sup> Le chromosome et la position des locus/gènes sur la version v2 du génome du maïs sont indiqués.

<sup>2</sup> La méthode de détection décrite dans le texte a été indiquée.

<sup>3</sup> Identification des SNP candidats pour le génotypage KASPar.

## **ANNEXE III.2 : TABLE DE FREQUENCES ALLELIQUES**



Table de fréquences alléliques ("F") des 218 SNP candidats et du nombre d'individus ("I") génotypés avec succès par la méthode KASPar dans chacune des 37 populations.

SNPID	P3_F	P3_I	P7_F	P7_I	P15_F	P15_I	PH1_F	PH1_I	P17_F	P17_I	P25_F	P25_I	PH2_F	PH2_I	P6_F	P6_I	PL1_F	PL1_I	P4_F	P4_I	PL2_F	PL2_I
Fst_03	0,98	132	0,99	144	0,03	144	1	146	0,87	134	0,04	139	0,99	145	0,99	144	0,92	145	0,92	113	0,05	127
Fst_04	0,4	128	0,41	145	0,69	140	0,77	145	0,23	142	0,9	134	0,68	140	0,4	142	0,03	143	0,23	119	0,07	126
Fst_05	0,23	61	0,16	71	0,01	71	0,46	69	0,63	72	0,13	65	0,33	62	0,12	66	0,9	68	0,38	68	0,99	59
Fst_07	0,71	132	0,76	143	0,74	138	0,9	146	0,62	136	1	131	0,84	138	0,77	142	0,06	143	0,73	108	0,03	126
Fst_08	0	140	0	152	0,01	144	0,82	130	0,01	143	0,01	139	0,99	145	0,01	145	0,01	147	0,01	122	0,01	130
Fst_09	0,98	138	1	150	0,94	140	0,01	147	0,87	129	1	138	0,99	143	0,99	145	0,03	147	0,99	118	0,06	128
Fst_12	0,41	122	0,38	144	0,58	142	0,08	143	0,46	139	0,27	125	0,03	139	0,33	135	0,92	141	0,2	113	0,44	122
Fst_13	0,19	136	0,41	132	0,17	140	0,26	145	0,35	143	0,04	132	0,1	142	0,33	144	0,96	143	0,15	117	0,02	130
Fst_14	0,99	140	1	150	0	144	0	147	0,01	143	0	139	1	144	1	144	0,01	147	0,99	122	0,51	129
Fst_16	0,46	134	0,29	146	0,45	141	0,09	144	0,06	133	0,48	131	0,14	138	0,02	141	0,64	145	0,3	113	0,98	127
Fst_18	0,8	137	0,89	150	0,95	140	0,84	144	0,86	140	0,67	132	0,84	141	0,53	139	1	145	0,69	113	1	127
Fst_19	0,31	128	0,92	142	0,3	119	0,78	143	0,03	135	0,92	135	0,88	142	0,74	142	0,36	142	0,5	101	0,16	129
Fst_20	0,38	126	0,08	148	0,22	140	0,27	139	0,13	93	0,13	131	0,04	142	0,03	140	0,87	139	0,16	119	0,45	124
Fst_21	0,56	136	0,54	148	0,91	138	0,07	141	0,82	142	0,13	129	0,14	141	0,34	140	0,98	142	0,44	117	0,97	128
Fst_22	0,47	122	0,7	139	0,98	137	0,53	134	0,98	139	0,57	122	0,46	132	0,5	139	0,92	142	0,54	112	0,98	129
Fst_23	0,31	126	0,01	148	0,01	143	0,01	144	0	128	0,04	139	0	145	0,01	145	0,69	145	0,07	113	0	130
Fst_24	0,52	128	0,3	134	0,38	137	0,5	142	0,65	131	0,39	129	0,17	142	0,28	142	0,13	146	0,39	109	0,99	127
Fst_25	0,01	140	0	152	0,01	143	0,03	147	0	140	0,1	137	0,08	146	0,01	145	0	145	0,03	119	0,02	130
Fst_27	0,03	140	0,01	152	0,01	144	0,02	147	0,03	143	0,01	139	0,01	146	0,01	145	0,01	147	0,02	122	0,02	130
Fst_28	0,24	136	0,35	140	0,22	138	0,92	141	0,06	139	0,87	126	0,84	134	0,19	143	0	143	0,14	113	0,02	127
Fst_29	0,16	135	0,53	149	0,76	143	0,05	144	0,71	139	0,39	132	0,5	144	0,12	143	0,98	146	0,25	115	0,98	129
Fst_30	0,83	136	0,98	147	0,98	142	0,98	146	0,65	139	0,98	135	0,9	144	0,76	143	0,04	144	0,64	119	0,64	129
Fst_31	0,87	131	0,76	149	0,95	142	0,73	143	0,61	137	0,71	131	0,8	144	0,48	141	0,96	144	0,8	116	0,99	129
Fst_32	0,81	136	0,78	149	0,21	137	0,94	145	0,35	142	0,88	137	0,97	144	0,73	143	0,79	145	0,58	116	0,02	129
Fst_34	0,65	122	0,43	135	0,09	137	0,01	143	0,72	130	0,03	128	0,09	140	0,16	142	0,89	144	0,38	111	0,98	124
Fst_35	0,44	133	0,18	141	0,17	136	0,69	142	0,17	136	0,66	132	0,82	139	0,55	133	0,03	143	0,63	116	0,05	130
Fst_36	0,41	120	0,23	146	0,12	143	0,29	143	0,29	122	0,32	134	0,26	142	0,27	144	0,95	142	0,52	114	0,02	129
Fst_37	0,76	127	0,61	136	0,83	139	0,98	139	0,4	133	0,93	132	0,83	137	0,99	139	0,05	139	0,59	111	0,63	123
Fst_38	0,98	138	0,02	151	0,03	141	0,01	146	0,98	129	0,01	135	0,02	143	0,03	133	0	136	0,98	120	0,98	129
Fst_40	0,86	132	0,77	147	0,9	142	0,16	137	0,98	142	0,18	109	0,64	141	0,89	145	0,99	147	0,74	112	1	127
Fst_41	0,08	127	0,3	138	0,04	141	0,95	144	0,02	135	0,38	106	0,94	134	0,33	135	0,06	145	0,15	113	0,01	125
Fst_45	0,67	139	0,6	152	0,14	143	0,77	147	0,18	142	0,73	133	0,83	144	0,8	145	0,5	146	0,8	120	0,02	129
Fst_46	0,62	131	0,72	147	0,37	142	0,12	144	0,69	135	0,16	136	0,1	143	0,24	142	0,98	142	0,44	110	0,1	125
Fst_47	0,19	127	0,49	131	0,09	141	0,62	146	0,16	128	0,44	126	0,81	140	0,31	52	0,01	144	0,14	107	0,02	125
Fst_49	0,66	131	0,68	137	0,59	141	0,98	146	0,68	140	0,99	138	0,99	137	0,98	141	0,53	143	0,68	117	0,11	126
Fst_50	0,96	133	0,84	144	0,01	144	0,17	144	0,99	142	0,06	136	0,19	142	0,9	134	0,99	144	0,92	112	0,99	130
Fst_51	0,56	131	0,57	146	0,99	144	0,18	141	0,99	140	0,56	124	0,55	138	0,55	141	0,93	144	0,56	114	0,93	127
Fst_53	0,07	132	0,26	136	0,11	141	0,79	146	0,01	141	0,94	136	0,94	145	0,07	140	0,02	142	0,04	119	0,02	129

SNPID	P3_F	P3_I	P7_F	P7_I	P15_F	P15_I	PH1_F	PH1_I	P17_F	P17_I	P25_F	P25_I	PH2_F	PH2_I	P6_F	P6_I	PL1_F	PL1_I	P4_F	P4_I	PL2_F	PL2_I
Fst_54	0,75	133	0,68	149	0,35	142	0,91	142	0,5	137	0,84	131	0,96	144	0,93	144	0,16	141	0,95	113	0,03	128
Fst_55	0,8	132	0,86	145	0,8	139	0,13	139	0,86	139	0,5	126	0,4	137	0,28	116	0,99	141	0,52	115	1	122
Fst_56	0,12	138	0,13	148	0	142	0,63	142	0,01	139	0,77	132	0,85	141	0,01	144	0,01	145	0,17	114	0,03	129
Fst_57	0,78	138	0,97	152	0,92	119	0,97	143	0,85	141	0,99	137	0,99	146	0,99	143	0,43	144	0,87	121	0,03	129
Fst_58	1	137	0,88	149	0,97	143	0,1	145	0,99	143	0,49	136	0,29	143	0,99	144	0,99	145	1	117	0,99	129
Fst_59	0,72	130	0,12	143	0,03	142	0,71	142	0,06	137	0,73	131	0,66	142	0,84	143	0,02	143	0,75	114	0,01	129
Fst_60	0,99	139	0,84	149	0,01	144	0,14	143	0,01	143	0,5	135	0,35	140	0,99	143	0,99	144	0,02	122	0,99	130
Fst_61	0,74	128	0,64	142	0,81	143	0,98	144	0,31	133	0,77	133	0,96	144	0,94	143	0,09	143	0,67	113	0,02	128
Fst_62	0,08	132	0,16	145	0,57	135	0,01	142	0,26	137	0,14	127	0,03	142	0,01	144	0,56	144	0,19	113	0,97	127
Fst_66	0,92	136	0,64	132	0,82	142	0,28	143	0,96	140	0,16	132	0,45	142	0,71	130	0,94	145	0,73	118	0,89	128
Fst_68	0,25	140	0,95	152	0,33	144	0,64	147	0,43	143	0,43	139	0,58	146	0,23	145	0,08	147	0,15	122	0,78	130
Fst_69	0,86	137	0,5	148	0,64	141	0,25	144	0,99	143	0,27	132	0,44	143	0,84	143	0,97	147	0,82	118	1	129
Fst_70	0,65	127	0,62	135	0,11	133	0,84	141	0,08	124	0,46	124	0,71	140	0,05	133	0,02	143	0,6	112	0,01	125
Fst_71	0,01	140	0,01	152	0,03	144	0,01	147	0,01	143	0,01	139	0,01	146	0,01	145	0,01	147	0,02	122	0,01	130
Fst_73	0	140	0,01	152	0,02	144	0,01	147	0,03	143	0,02	139	0,01	146	0,03	145	0,01	147	0,02	122	0,02	130
Fst_74	0,65	124	0,77	148	0,58	141	0,21	138	0,97	141	0,36	120	0,39	138	0,04	136	0,98	146	0,67	102	0,98	128
Fst_75	0,44	126	0,71	151	0,96	140	0,24	142	0,94	141	0,51	138	0,38	146	0,37	135	0,99	146	0,66	106	0,99	130
Fst_76	0,55	131	0,65	139	0,72	143	0,19	143	0,49	137	0,4	124	0,12	143	0,65	138	0,95	146	0,56	115	0,83	130
Fst_77	0,99	133	0,01	152	0	144	0	147	0,01	143	0,06	139	1	143	0,01	145	0	147	0,99	116	0,4	111
Fst_78	0,59	136	0,68	150	0,02	143	0,93	147	0,72	138	0,6	134	0,77	144	0,61	144	0,08	143	0,71	116	0,02	126
Fst_80	0,14	121	0,18	150	0	142	0,81	143	0,01	140	0,44	132	0,85	143	0,01	144	0,01	146	0,27	112	0,01	128
Fst_82	0,58	70	0,32	79	0,15	73	0,88	76	0,54	67	0,8	70	0,72	78	0,76	78	0,83	76	0,64	47	0,03	69
Fst_83	0,03	140	0,02	152	0,03	144	0,83	146	0,01	143	0,01	137	0,01	146	0,01	143	0	140	0,02	122	0,01	130
Fst_84	0,02	140	0,05	152	0,01	144	0,87	143	0,03	143	0,01	133	0,01	146	0,01	123	0	144	0,02	122	0,01	130
Fst_85	0,89	72	0,86	80	0	73	0,17	75	0	71	0,01	72	0	82	0,99	78	0	76	0,06	50	0	70
Fst_87	0,11	119	0,15	152	0,1	144	0,88	139	0,01	137	0,06	133	0,03	144	0,43	143	0,01	138	0,09	112	0,01	129
Fst_88	0,82	137	0,6	146	0,87	141	0,09	143	0,85	141	0,41	133	0,13	136	0,08	143	0,99	145	0,71	119	0,98	129
Fst_89	0,01	140	0,01	152	0,99	144	0,18	140	0,01	143	1	137	0	146	0,99	143	1	146	0,03	122	0,01	130
Fst_90	0,01	105	0	100	0	108	0	105	0	103	0,01	96	0,01	107	0,01	102	0,02	105	0,01	93	0,75	92
Fst_91	0	128	0,07	152	0,06	144	0,84	137	0,06	143	0,08	139	0,03	146	0,01	140	0	145	0,03	112	0,03	130
Fst_92	0,36	128	0,45	147	0,02	143	0,89	147	0,03	133	0,39	135	0,21	144	0,29	142	0,05	146	0,44	112	0,49	126
Fst_94	0,01	140	0,03	152	0,03	144	0,83	140	0,02	143	0,02	139	0,01	146	0,02	145	0	145	0,02	122	0,01	130
Fst_95	0,33	104	0,15	100	0,5	105	0,01	105	0,3	100	0,04	92	0,02	108	0,09	102	0,92	103	0,33	92	0,69	94
Fst_96	0,07	104	0,57	97	0,38	104	0,92	104	0,01	101	0,76	96	0,83	105	0,24	100	0,03	106	0,1	94	0,01	94
Fst_97	0,88	137	0,78	151	0,58	140	0,96	147	0,4	141	0,95	137	0,76	143	0,89	143	0,23	142	0,7	118	0,42	128
MS_01	0,55	122	0,67	107	0,97	131	0,09	141	0,99	138	0,08	116	0,07	132	0,83	129	0,99	143	0,83	103	0,99	128
MS_02	0	140	0,01	152	0,01	144	0	147	0,01	143	0,01	139	0	146	0,01	145	0,01	147	0	122	0	130
MS_03	0,01	140	1	150	0	144	0,02	147	0,01	143	0,01	139	0,99	139	0,02	145	0	147	0,03	122	0,01	130
MS_09	0,76	119	0,97	146	0,99	140	0,24	139	0,99	136	0,11	131	0,05	139	0,46	128	0,98	146	0,79	113	0,99	124
MS_11	0,08	140	0,13	152	0,01	144	0,01	147	0,05	143	0,42	133	0,32	141	0,01	145	0,01	147	0	116	0,01	130
MS_13	0,01	140	0,01	152	0,01	144	0,01	147	0,01	143	0,01	139	0,02	146	0,01	145	0,02	147	0,01	122	0,02	130

SNPID	P3_F	P3_I	P7_F	P7_I	P15_F	P15_I	PH1_F	PH1_I	P17_F	P17_I	P25_F	P25_I	PH2_F	PH2_I	P6_F	P6_I	PL1_F	PL1_I	P4_F	P4_I	PL2_F	PL2_I
MS_14	0	140	0,01	148	0	144	0,01	146	0	142	0,36	128	0,38	139	0,04	145	0,01	147	0,03	122	0,01	130
MS_15	0,93	140	0,99	151	0,98	143	0,9	144	0,98	142	0,39	134	0,26	142	0,91	144	1	147	0,97	120	0,99	129
MS_16	0,91	56	0,21	60	0,86	46	0,03	68	0,46	62	0,03	65	0,15	61	0,5	52	0,81	68	0,84	63	0,99	60
MS_17	0,83	128	0,67	141	0,41	131	0,05	139	0,99	135	0,19	121	0,06	133	0,83	140	0,96	143	0,82	117	0,99	123
MS_18	0,99	137	0,8	147	0,03	144	0,05	142	0,99	141	0,2	130	0,07	138	0,99	142	0,98	147	0,04	122	0,99	130
MS_19	0,98	140	0,65	152	0,99	144	0,08	147	0,99	143	0,19	139	0,05	146	0,99	145	0,99	147	0,98	122	0,99	130
MS_20	0,01	120	0,34	98	0,03	144	0,95	138	0,01	129	0,79	124	0,94	139	0,01	142	0,02	142	0	112	0,01	125
MS_21	0,99	137	0,74	145	0,01	144	0,05	144	0,99	141	0,2	135	0,06	140	0,99	145	0,99	146	0,02	122	0,99	130
MS_22	1	131	0,75	139	0,06	144	0,06	136	0,99	135	0,21	124	0,07	132	0,99	136	0,98	143	1	112	0,99	125
MS_23	0,01	139	0,15	151	0,01	144	0,94	142	0,01	142	0,69	130	0,94	141	0,01	139	0,01	145	0,01	121	0,01	130
MS_24	0,99	139	0,75	147	0,01	144	0,05	144	0,99	141	0,2	133	0,06	143	0,94	142	0,98	146	0,06	122	0,99	129
MS_25	0,01	140	0	152	0	144	0,01	147	0	143	0,01	139	0	146	0	145	0	147	0,02	122	0,01	130
MS_26	0	140	0	152	0	144	0,01	147	0	143	0,01	139	0	146	0	145	0	147	0	122	0	130
MS_27	0,55	129	0,84	143	0,93	141	0,04	144	0,98	140	0,14	137	0,05	142	0,42	144	0,97	146	0,55	116	0,98	129
MS_28	0,02	140	0,01	152	0	144	0,01	147	0,01	143	0,01	139	0,01	146	0,01	145	0,01	147	0	122	0,01	130
MS_29	0	140	0	152	0	144	0,01	147	0,01	143	0,01	139	0	146	0	145	0,01	147	0,01	122	0,02	130
MS_30	0,29	140	0,17	152	0,07	144	0,93	147	0,13	143	0,78	139	0,84	146	0,47	145	0,05	147	0,25	122	0,02	130
MS_31	0,59	116	0,82	144	0,93	142	0,05	128	0,98	134	0,17	121	0,12	136	0,38	138	0,98	144	0,6	87	0,98	127
MS_32	0,48	64	0,15	71	0,06	70	0,97	69	0,01	72	0,89	62	0,87	63	0,72	65	0,04	66	0,43	69	0,02	59
MS_34	0,47	62	0,27	68	0,06	70	0,97	68	0,01	67	0,89	61	0,89	57	0,6	64	0,04	70	0,43	69	0,01	58
MS_35	0,11	134	0,07	148	0,01	144	0,05	145	0,06	143	0,19	134	0,34	139	0,01	135	0,01	145	0,15	117	0,02	130
MS_36	0	122	0	142	0,01	144	0,08	139	0,01	143	0,15	135	0,33	139	0,02	124	0,01	140	0,01	105	0,02	130
MS_37	0,88	138	0,84	151	1	142	0,37	144	0,99	143	0,24	136	0,18	145	0,99	138	0,91	124	0,8	117	0,97	130
MS_39	0,16	137	0,06	150	0,03	142	0,01	145	0,01	143	0,02	131	0,05	144	0,02	145	0,01	147	0,02	122	0	130
MS_40	0,04	140	0,01	152	0,01	144	0	147	0	139	0,09	133	0,4	131	0,01	145	0,01	146	0,05	122	0,01	130
MS_44	0,01	140	0	152	0	144	0	147	0,01	143	0,01	139	0	146	0,01	145	0,01	147	0,02	122	0,01	130
MS_45	0	66	0,01	72	0,01	71	0,92	68	0,01	72	0,84	63	0,82	63	0,01	67	0,02	70	0,01	72	0,02	60
MS_47	0,1	140	0,99	145	0,02	144	0,54	141	1	134	0,21	129	0,16	137	0,99	131	0,99	147	0,08	122	0,99	127
MS_48	1	139	0,99	151	0,01	144	0,16	143	1	140	0,57	132	0,22	145	0,99	141	0,99	142	1	118	0,02	130
MS_49	0,02	140	0	149	0,01	144	0,89	146	0	141	0,84	137	0,85	142	0,01	143	0,01	144	0	118	0,01	129
MS_50	0	66	0,01	70	0,01	71	0,93	67	0,01	57	0,88	64	0,83	62	0,01	67	0,01	69	0,01	71	0,02	60
MS_51	0,97	66	0,99	69	0,01	71	0,14	67	0,04	72	0,38	43	0,49	57	0,99	67	0,98	66	0,93	72	0	60
MS_55	0,02	131	0,01	150	0,03	128	0,95	147	0	141	0,91	135	0,86	140	0,02	138	0,01	146	0,01	106	0,01	129
MS_56	0,04	140	0,99	149	1	144	0,08	145	1	139	0,17	129	0,16	138	1	141	0,99	144	1	119	0,99	129
MS_57	0,61	130	0,83	146	0,67	142	0,29	139	0,91	137	0,14	135	0,09	140	0,36	140	0,82	146	0,75	114	0,98	128
MS_58	0,87	134	0,99	151	0,99	143	0,1	145	0,85	138	0,16	135	0,14	141	0,76	143	0,72	142	0,87	115	0,98	130
MS_59	0,05	125	0,07	145	0,03	144	0,9	142	0	137	0,49	124	0,71	140	0,01	142	0	145	0,03	111	0,01	128
MS_60	0,32	127	0,43	146	0,01	144	0,96	147	0,08	137	0,84	129	0,89	138	0,35	143	0,03	146	0,21	113	0,01	128
MS_61	0,01	127	0,01	134	0,03	144	0,95	142	0,01	133	0,92	126	0,89	141	0,08	126	0,06	142	0,02	107	0,01	127
MS_62	0,6	127	0,61	143	0,71	141	0,98	147	0,69	136	0,88	134	0,9	144	0,58	143	0,51	144	0,35	113	0,02	124
MS_63	0,99	139	0,99	121	1	144	0,08	145	1	128	0,17	129	0,13	145	0,99	138	0,99	147	0,98	118	0,99	128

SNPID	P3_F	P3_I	P7_F	P7_I	P15_F	P15_I	PH1_F	PH1_I	P17_F	P17_I	P25_F	P25_I	PH2_F	PH2_I	P6_F	P6_I	PL1_F	PL1_I	P4_F	P4_I	PL2_F	PL2_I
MS_64	0,1	128	0,01	148	0,01	142	0,94	141	0,27	139	0,85	134	0,69	144	0,02	131	0,07	145	0,06	116	0,01	130
MS_66	0,08	135	0,01	148	0	143	0,78	146	0,01	142	0,84	135	0,84	142	0,03	145	0,02	145	0,22	119	0,01	129
MS_67	0,08	128	0,01	146	0	139	0,9	140	0	133	0,86	132	0,84	143	0,21	141	0,02	140	0,17	115	0,01	128
MS_68	0,96	57	0,85	55	0,01	71	0,18	67	0,01	72	0,12	64	0,2	61	0,92	52	0,98	67	0,88	64	0	60
MS_69	0,63	120	0,64	125	0,89	126	0,52	124	0,64	116	0,13	105	0,1	130	0,76	104	0,71	131	0,53	102	0,93	118
MS_70	0,23	139	0,04	137	0,27	112	0,92	142	0,05	139	0,87	135	0,86	139	0,69	128	0,03	145	0,26	109	0,02	127
MS_74	0,17	130	0,01	149	0,13	127	0,91	137	0,16	136	0,83	130	0,86	141	0,02	139	0,01	146	0,01	110	0,22	125
MS_75	0	140	0,99	152	0,01	144	0,11	142	1	140	0,18	132	0,15	139	0,99	143	0,99	146	1	121	0,99	129
MS_76	0,01	140	0,01	150	0,02	144	0,89	145	0	143	0,83	135	0,84	144	0,01	145	0,01	146	0,02	122	0,01	129
MS_77	0,01	140	0,99	151	0,01	144	0,11	143	1	143	0,18	134	0,17	140	0,99	143	0,99	146	1	119	0,99	129
MS_78	0,02	140	0,01	147	0,01	142	0,9	142	0,03	143	0,88	121	0,87	138	0,01	144	0,01	147	0	119	0,01	130
MS_81	0,61	133	0,67	149	0,97	141	0,07	142	0,99	141	0,13	134	0,07	140	0,55	131	0,97	144	0,4	112	0,99	129
MS_82	0,99	73	0,99	80	0,99	73	0,99	77	0,99	71	0,43	71	0,46	82	0,99	70	0	76	0,96	47	0,99	70
MS_83	0,02	140	0,01	150	0	144	0,01	147	0	143	0,65	134	0,53	139	0,01	118	0,01	147	0	118	0,01	129
MS_87	0,01	140	0,99	151	1	142	1	147	0,01	143	0,61	135	0,52	141	0,01	145	0,01	147	1	120	0,02	130
MS_88	0,01	140	0,99	152	0,05	144	0	147	0,99	142	0,96	137	0,47	140	0,9	140	1	147	0,02	122	0,01	130
MS_89	0,06	140	0,99	152	0,01	144	0,99	141	0,99	138	0,65	126	0,17	134	1	145	0,99	146	0,05	122	0,01	130
SB_01	0,03	131	0,03	146	0,02	144	0	144	0,12	135	0,01	134	0,01	144	0,01	141	0,08	145	0,08	114	0,96	123
SB_02	0,99	139	1	152	1	144	0	147	0	143	0,01	139	1	145	0,99	145	0	147	0,99	120	0,08	128
SB_03	0,03	137	0,26	147	0,85	141	0,81	146	0,1	141	0,71	136	0,86	145	0,03	143	0,15	146	0,02	122	0,02	130
SB_04	0,02	131	0,33	146	0,58	141	0,96	146	0,03	138	0,93	132	0,97	141	0,02	144	0,02	145	0,02	114	0,17	128
SB_07	0,54	125	0,11	145	0,52	101	0,08	146	0,98	136	0,09	128	0,07	141	0,54	119	0,98	143	0,72	112	0,98	129
SB_08	0,78	134	0,67	147	0,33	140	0,98	144	0,66	139	0,88	132	0,86	142	0,95	135	0,28	143	0,75	114	0,09	124
SB_09	0,17	136	0,09	149	0,05	143	0,02	145	0,11	140	0,21	137	0,01	144	0,01	143	0,93	145	0,09	117	0,61	127
SB_10	0,39	62	0,25	71	0,85	70	0,38	64	0,85	72	0,38	65	0,33	62	0,58	63	0,96	70	0,58	71	0,98	59
SB_12	0,03	139	0,01	151	0,02	143	0,01	145	0,19	141	0,01	139	0	146	0,01	145	0,23	147	0,02	121	0,89	130
SB_13	0	140	0,01	152	0,01	144	0,01	147	0,02	143	0,02	139	0	146	0,01	145	0,01	147	0	122	0	130
SB_14	0,11	65	0,11	71	0,14	70	0,13	67	0,09	72	0,01	67	0,02	64	0,04	63	0,05	70	0,18	71	0,92	59
SB_15	0,06	123	0,01	149	0,21	140	0,8	142	0,01	141	0,5	129	0,96	142	0,03	140	0,01	139	0,1	114	0	130
SB_16	0,18	139	0,01	146	0,47	137	0,11	144	0,17	112	0,01	138	0,01	144	0,06	144	0,02	147	0,12	117	0,97	128
SB_20	0,01	140	0	152	0,01	144	0	147	0,01	143	0,01	139	0,01	146	0,01	145	0,01	147	0,02	122	0,02	130
SB_22	0,93	135	0,98	150	0,97	142	0,9	144	0,98	136	0,4	126	0,2	142	0,87	143	1	146	0,97	116	0,99	128
SB_25	0,77	133	0,83	151	0,56	138	0,74	144	0,73	141	0,8	133	0,89	143	0,83	141	0,84	146	0,77	119	0,19	129
SB_26	0,79	135	0,99	151	0,45	139	0,97	146	0,98	140	0,99	134	0,99	143	0,87	143	0,91	145	0,71	111	0,03	128
SB_27	0,11	127	0,01	148	0	143	0,77	145	0,01	139	0,6	129	0,87	143	0,07	139	0,08	146	0,15	115	0,01	129
SB_28	0,37	62	0,01	69	0,5	65	0,1	73	0,04	60	0,26	58	0,14	78	0,13	76	0,06	76	0,25	40	0,01	69
SB_29	0,58	133	0,7	145	0,99	143	0,14	142	1	138	0,09	119	0,53	139	0,58	137	0,99	146	0,61	110	0,99	129
SB_30	0,96	124	0,99	132	0,92	132	0,08	133	0,99	134	0,14	116	0,65	127	0,98	132	0,1	147	0,96	115	1	116
SB_32	0,97	132	0,96	146	0,53	139	0,82	146	0,96	135	0,06	139	0,98	146	0,91	140	0,96	143	0,97	116	0,03	122
SB_33	0	140	0	152	0,01	144	0,01	147	0,01	143	0,01	139	0	146	0,01	145	0	147	0,02	122	0	130
SB_34	0,3	130	0,38	148	0,27	142	0,97	145	1	141	0,98	139	0,97	146	0,98	144	0,03	144	0,52	112	0,03	126

SNPID	P3_F	P3_I	P7_F	P7_I	P15_F	P15_I	PH1_F	PH1_I	P17_F	P17_I	P25_F	P25_I	PH2_F	PH2_I	P6_F	P6_I	PL1_F	PL1_I	P4_F	P4_I	PL2_F	PL2_I
SB_35	0,65	119	0,93	140	0,79	136	0,99	143	0,73	128	1	133	0,99	142	0,99	140	0,72	142	0,88	112	0,03	118
SB_38	0,46	137	0,28	149	0,57	143	0,16	147	0,47	140	0,34	136	0,32	142	0,03	143	0,93	144	0,24	117	0,75	130
SB_39	0,98	132	0,96	148	1	142	0,99	145	0,89	135	0,97	134	0,99	144	0,94	138	0,96	147	1	117	0,17	123
SB_41	0,15	131	0,07	150	0,01	144	0	147	0,27	137	0,05	139	0,01	141	0,01	144	0,02	145	0,02	115	0,96	126
SB_43	0,18	134	0,84	148	0,74	140	0,2	145	0,69	137	0,72	131	0,87	142	0,4	141	0,91	144	0,32	112	0,95	128
SB_44	0,2	130	0,45	144	0,17	141	0,37	143	0,26	137	0,06	128	0,1	143	0,09	139	0,02	145	0,05	114	0	127
SB_45	0,92	140	0,56	152	0,85	144	0,07	147	0,66	143	0,29	139	0,11	146	0,5	145	0,98	147	0,95	122	0,25	130
SB_46	0,63	128	0,84	149	0,67	142	0,06	146	0,58	136	0,05	133	0,35	144	0,53	144	0,92	144	0,66	113	0,98	129
SB_47	0,89	121	0,89	147	0,99	143	0,38	137	0,99	140	0,23	124	0,21	119	0,99	143	0,99	144	0,84	113	0,97	128
SB_48	0,14	133	0,46	149	0	142	0,33	144	0,13	138	0,09	135	0,01	145	0,01	145	0,04	146	0,05	117	0,83	126
SB_49	0,72	130	0,44	130	0,27	59	0,96	139	0,02	141	0,75	127	0,97	136	0,93	119	0,03	144	0,66	114	0,02	128
SB_50	0,93	132	1	150	0,99	141	0,99	146	0,87	130	0,03	139	0,99	145	0,06	145	0,17	139	0,79	107	0,99	128
SB_51	0,01	140	0,05	152	0	144	0	147	0,02	143	0	135	0	144	0	145	0,02	147	0	120	0,49	113
SB_52	0,9	130	0,99	150	0,99	141	0,93	142	1	142	0,97	138	1	143	0,99	143	0,53	136	0,83	118	0,04	130
SB_54	0,3	133	0,7	149	0,38	142	0,86	144	0,03	139	0,89	131	0,67	141	0,75	138	0,11	146	0,48	114	0,02	129
SB_55	0,27	138	0,05	149	0,68	141	0,96	143	0,18	138	0,95	138	0,97	142	0,03	141	0,03	143	0,34	115	0,02	128
SB_56	0,61	136	0,46	149	0,43	140	0,1	139	0,94	137	0,13	131	0,14	142	0,83	145	0,62	145	0,5	117	0,98	127
SB_57	0,33	127	0,17	138	0,02	144	0,9	140	0,01	134	0,69	127	0,77	137	0,74	137	0,01	145	0,2	110	0,02	126
SB_58	0,11	66	0,08	72	0,35	71	0	70	0,01	72	0,39	65	0,49	63	0,48	67	0,04	71	0,95	66	0,02	60
SB_59	0,01	138	0	150	0,01	144	0,84	144	0	141	0,06	135	0,22	140	0,01	145	0	147	0,11	117	0,01	130
SB_60	0,02	106	0	100	0	108	0	105	0	103	0,63	87	0,52	99	0	101	0	106	0,01	93	0	95
SB_61	0	74	0	80	0	73	0,01	77	0	71	0	72	0	82	0	78	0	76	0	50	0	70
SB_63	0,89	135	0,68	149	0,85	142	1	145	1	143	0,82	134	0,99	144	0,97	80	0,91	146	0,83	119	0,03	130
SB_64	0,92	131	0,56	123	0,99	139	0,09	139	0,84	128	0,86	124	0,43	135	0,64	28	1	142	0,77	114	0,98	128
SB_65	0,94	106	0,97	100	0,64	105	0,92	103	0,7	102	0,97	94	0,83	106	0,46	96	0,99	104	0,96	93	0,42	90
SB_66	0,88	138	0,99	150	0,99	144	0,99	147	1	140	0,99	138	0,99	146	0,9	144	0,1	143	0,84	118	0,9	130
SB_69	0	40	0,24	25	0	37	0,84	31	0,02	31	0	29	0	44	0	35	0,01	35	0	22	0	35
SB_70	0,01	140	0,02	152	0,01	144	0,27	139	0,01	143	1	139	0	146	0,99	145	1	146	0	122	0,01	130
SB_71	0,01	140	0,01	152	0,02	144	0,88	141	0	143	0,01	137	0,02	146	0,01	143	0,01	144	0,02	122	0	130
SB_72	0,12	138	0,31	148	0,77	142	0,9	126	0,13	142	0,01	135	0	143	0,01	144	0,01	145	0,22	115	0,02	127
SB_73	0	140	0,01	152	0,01	144	0,19	107	0,02	143	0,03	139	0,01	146	0,99	142	1	146	0,02	122	0,01	130
SB_74	0,99	137	0,01	152	0,9	142	0,17	146	0,02	143	1	138	0	146	0,94	144	1	146	0,97	120	0,93	130
SB_75	0,06	140	0,01	152	0,01	144	0,83	144	0,03	143	0,01	136	0	145	0,01	144	0	146	0,05	122	0,02	130
SB_76	0,01	106	0,02	100	0,02	108	0,57	23	0,01	103	0,02	96	0,01	108	0,02	102	0,02	106	0	94	0	95
SB_77	0,05	40	0	28	0,03	37	0,84	32	0,03	31	0,02	29	0	44	0,03	35	0,03	35	0	22	0,03	35
SB_79	0,72	111	0,94	116	0,99	142	0,96	146	0,99	107	0,12	133	0,2	144	0,91	144	0,92	145	0,59	92	0,99	128
SB_80	0,05	132	0,01	150	0	144	0,02	146	0,01	142	0,88	128	0,79	144	0,01	143	0,06	146	0,16	116	0,02	130
Fst_03	0,88	17	0,1	20	0,05	20	0,63	19	0,11	19	0,96	13	0,96	12	0,05	20	0,72	18	0,8	20	0,94	18
Fst_04	0,19	16	0	20	0,84	19	0,16	19	0,22	16	0,25	16	0,9	20	0,72	18	0,16	19	0,29	7	0,88	8
Fst_05	0,18	17	0,2	20	0,76	19	0,1	20	0,05	19	0,38	21	0,76	17	0,68	19	0,05	20	0,65	20	0,83	18
Fst_07	0,69	13	0,03	18	0,95	19	0,5	17	0,47	17	0,6	15	0,15	20	0,85	20	0,53	20	0,83	3	0,94	18

SNPID	P22_F	P22_I	P2_F	P2_I	P28_F	P28_I	P16_F	P16_I	P21_F	P21_I	P19_F	P19_I	P29_F	P29_I	P27_F	P27_I	P18_F	P18_I	P10_F	P10_I	P24_F	P24_I
Fst_08	0	17	0	20	0,05	20	0	20	0	19	0	21	0,97	19	0	20	0	20	0,05	20	0	18
Fst_09	0,06	17	0,15	20	0,05	20	0,97	18	0,89	19	0,87	19	0	20	0	20	0,89	18	0,95	19	0,22	18
Fst_12	0,71	12	0,35	20	0,25	20	0,63	20	0,35	17	0,56	17	0,08	19	0,4	20	0,69	18	0,85	13	0,08	12
Fst_13	0,06	17	0,74	19	0,14	18	0,43	20	0,11	19	0,36	18	0,2	20	0,21	17	0,21	19	0,35	13	0,2	15
Fst_14	0	17	0,05	20	0,35	20	0,1	20	0,11	19	0	21	0,18	20	0,53	19	0	20	0,9	20	0,78	16
Fst_16	0,33	15	0,72	16	0,06	18	0,26	17	0,32	17	0,07	14	0,05	20	0,05	20	0,13	16	0,85	20	0,25	2
Fst_18	0,53	16	0	20	0,48	20	0,84	19	0,87	19	0,93	21	0,82	19	0,58	19	0,84	19	0,1	20	0,94	16
Fst_19	0,06	17	0,12	17	0,9	20	0	20	0,16	19	0,05	21	0	20	0,98	20	0,1	20	0,71	14	0,22	18
Fst_20	0,06	17	0,63	15	0,15	20	0,2	20	0,32	19	0,33	21	0,08	19	0,1	20	0,25	20	0,13	8	0,56	18
Fst_21	0,06	17	0,78	18	0,17	18	0,73	20	0,72	18	0,52	21	0,03	20	0,18	20	0,92	19	0,83	20	0,25	16
Fst_22	0,24	17	0,3	20	0,47	19	0,05	20	0	19	0,29	21	0,53	18	0,35	20	0,94	18	0,5	20	0,72	18
Fst_23	0	17	0,68	17	0,05	20	0	20	0	19	0,05	21	0	20	0,05	20	0	20	0,15	20	0,11	18
Fst_24	0,08	13	0,77	15	0,17	18	0,22	18	0,56	16	0,46	14	0,06	17	0,32	14	0,5	15	0,95	20	0,5	1
Fst_25	0,03	17	0	20	0,08	18	0	20	0,11	19	0	21	0,21	19	0,1	20	0	20	0,15	20	0,06	18
Fst_27	0,06	17	0	20	0,05	20	0	20	0,05	19	0,1	21	0	20	0	20	0	20	0,1	20	0	18
Fst_28	0,06	17	0	20	0,87	19	0	20	0,11	19	0,08	19	0,73	11	0,88	16	0,05	20	0,4	20	0,29	7
Fst_29	0,97	17	0,68	17	0,39	19	0,93	20	0,81	18	0,72	18	0,53	20	0,5	20	0,76	19	0,4	20	0,5	9
Fst_30	0,97	17	0,3	20	0	20	0,83	20	0,66	19	0,58	18	0,92	19	0	20	0,55	19	0,5	12	0	18
Fst_31	0,12	17	0,83	15	0,7	20	0,8	20	0,53	19	0,83	18	0,82	19	0,58	20	0,64	18	0,93	7	0,7	5
Fst_32	0,16	16	0,98	20	0,93	20	0,18	20	0,31	16	0,45	19	0	20	0,93	20	0,44	18	0,47	19	0,11	18
Fst_34	0,15	10	0,3	20	0,05	19	0,74	17	0,67	18	0,77	13	0,05	19	0,21	19	0,64	18	0,5	4	0,94	18
Fst_35	0,73	15	0,24	19	0,79	17	0,06	16	0,13	15	0,33	20	0,76	17	0,75	18	0,25	18	0,22	18	0,64	18
Fst_36	0,18	14	0,97	17	0,33	18	0,75	16	0,71	17	0,69	13	0,13	15	0,38	20	0,5	18	0,1	10	0,44	18
Fst_37	0,1	15	0,11	18	0,91	16	0,11	19	0,29	14	0,16	16	0,68	17	0,89	19	0,29	19	0,35	20	0,92	6
Fst_38	0	17	0	20	0	20	0,05	20	0,21	19	0,95	20	0	20	0	20	0,95	20	0,11	19	0,06	18
Fst_40	0,06	17	0,15	20	0,47	19	0,75	18	0,83	18	0,97	18	0,47	19	0,68	20	0,92	18	0,6	20	0,79	7
Fst_41	0,12	17	0,25	18	0,63	19	0,08	20	0,47	19	0,39	18	0,68	19	0,65	20	0,39	19	0,19	13	0,67	6
Fst_45	0,03	17	0,03	20	0,84	19	0,08	20	0,29	19	0,14	21	0,91	17	0,9	20	0,08	20	0,34	16	0,58	12
Fst_46	0,38	16	0,92	19	0,18	20	0,65	20	0,58	18	0,61	18	0,28	18	0,25	20	0,53	19	0,91	11	0,39	9
Fst_47	0,06	17	0	20	0,92	19	0,13	19	0,33	12	0,14	14	0,88	20	0,95	19	0,05	11	0,07	15	0,63	8
Fst_49	0,76	17	0,47	18	0,98	20	0,75	20	0,58	18	0,6	20	0	20	0	20	0,78	20	0,31	13	0,97	15
Fst_50	0,06	17	0,1	20	0,13	19	0	20	0	19	0,05	21	0,05	19	0,29	19	0	20	0,3	20	0,85	10
Fst_51	0,06	17	0,1	20	0,63	19	0,05	20	0,05	19	0,05	21	0,32	19	0,65	20	0	20	0,35	20	0,63	4
Fst_53	0,06	17	0,15	17	0,89	18	0	20	0,11	19	0	21	0,91	17	0,9	20	0	20	0,18	11	0,33	6
Fst_54	0,73	15	0,5	13	0,8	20	0,68	20	0,68	17	0,67	18	0	20	0,95	20	0,53	20	0,27	11	0,67	18
Fst_55	0,4	15	0,92	18	0,43	20	0,53	20	0,69	18	0,88	21	0,28	16	0,39	18	0,69	18	0,45	20	0,8	5
Fst_56	0,06	17	0,24	17	0,59	17	0	20	0,16	19	0	21	0,68	19	0,48	20	0	20	0,04	12	0,17	6
Fst_57	0	17	0,5	17	0,95	20	0,95	20	0,74	19	0,88	20	0	20	0,93	20	0,84	19	0,89	19	0,11	18
Fst_58	0,06	17	0,05	20	0,45	19	0	20	0,16	19	0	21	0,37	19	0,65	20	0	20	0	20	0,11	18
Fst_59	0,12	17	0,1	20	0,63	20	0	20	0,09	17	0,21	14	0,75	20	0,5	20	0	20	0,55	20	0,5	3
Fst_60	0,06	17	0	20	0,41	17	0	20	0	19	0,05	21	0,28	20	0,3	20	0	20	0,1	20	0,64	11

SNPID	P22_F	P22_I	P2_F	P2_I	P28_F	P28_I	P16_F	P16_I	P21_F	P21_I	P19_F	P19_I	P29_F	P29_I	P27_F	P27_I	P18_F	P18_I	P10_F	P10_I	P24_F	P24_I
Fst_61	0,54	12	0,15	20	0,95	19	0,26	19	0,5	18	0,19	13	0,9	20	0,9	20	0,21	19	0,79	12	0,81	8
Fst_62	0,15	13	0,5	15	0,05	20	0,14	18	0,15	17	0,4	15	0,03	20	0,08	20	0,18	19	0,8	20	0,89	18
Fst_66	0,06	17	0	20	0,24	19	0	20	0,92	19	0,92	18	0,55	19	0,55	20	0,94	18	0,2	20	0,44	8
Fst_68	0,65	17	0,45	20	0,4	20	0,15	20	0,32	19	0,71	21	0,8	20	0,5	20	0,35	20	0,6	20	0,83	18
Fst_69	0,43	15	0	20	0,32	19	0,98	20	0,11	19	0,05	21	0,48	20	0,2	20	0	20	0,65	20	0,5	4
Fst_70	0,76	17	0,67	12	0,42	18	0,21	14	0,06	16	0,38	21	0,63	15	0,39	19	0,2	20	0,63	4	0,94	18
Fst_71	0,06	17	0	20	0	20	0	20	0	19	0	21	0	20	0,1	20	0	20	0,1	20	0	18
Fst_73	0	17	0	20	0	20	0	20	0,05	19	0	21	0	20	0	20	0,05	20	0,15	20	0,11	18
Fst_74	0,06	17	0,05	20	0,53	17	0,05	20	0	19	0,97	18	0,37	15	0,23	20	0,05	20	0,85	13	0,5	9
Fst_75	0,06	17	0,88	20	0,28	18	0,05	20	0,89	18	0,95	21	0,53	19	0,35	20	0,8	20	0,95	19	0,63	15
Fst_76	0,6	15	0,78	18	0,2	20	0,53	19	0,67	18	0,71	17	0,03	18	0,21	19	0,58	19	0,32	11	0,4	5
Fst_77	0,06	17	0,15	20	0,1	20	0,05	20	0	19	0,14	21	0,05	20	0	20	0	20	0,7	20	0,94	18
Fst_78	0,06	17	0,19	18	0,5	19	0,68	19	0,47	15	0,33	18	0,63	19	0,39	19	0,74	19	0,25	20	0,28	18
Fst_80	0,24	17	0,3	20	0,58	18	0,15	20	0	19	0,05	21	0,65	20	0,43	20	0	20	0,3	15	0,22	9
Fst_82	0,35	17	0,74	17	0,7	20	0,29	19	0,53	15	0,75	18	0,61	19	0,61	19	0,32	17	0,22	16	0,97	16
Fst_83	0,03	17	0,1	20	0,05	20	0,05	20	0	19	0,05	21	0,08	20	0	20	0	20	0,05	20	0	18
Fst_84	0	17	0,1	20	0	20	0	20	0	19	0	21	0,05	19	0	20	0	20	0,1	20	0,17	18
Fst_85	0	17	0	20	0	20	0,98	20	0	19	0	21	0,95	20	0	20	0	20	0	20	0	18
Fst_87	0,18	17	0,4	20	0,05	19	0,05	20	0	19	0,19	21	0,03	18	0,05	20	0	20	0,85	20	0,94	18
Fst_88	0,91	17	0	20	0,18	20	0,9	20	0,94	18	0,98	21	0,4	20	0,2	20	0,8	20	0,94	18	0,17	18
Fst_89	0,06	17	0	20	0	20	0	20	0	19	0	21	0,95	20	0	20	0	20	0	20	0,06	18
Fst_90	0,18	17	0,38	12	0,05	19	0,1	20	0,11	19	0,1	21	0,2	20	0,18	20	0,1	20	0,55	20	0,56	18
Fst_91	0,53	17	0,5	20	0,05	20	0,05	20	0,05	19	0,33	21	0,03	16	0,05	20	0,05	20	0,9	20	0,94	18
Fst_92	0,06	16	0,1	20	0,97	19	0,08	19	0,12	17	0,12	21	0,76	19	0,83	20	0,17	18	0,03	16	0,88	12
Fst_94	0	17	0	20	0	20	0,05	20	0,05	19	0,05	21	0,05	20	0	20	0	20	0,05	20	0,11	18
Fst_95	0,53	16	0	20	0	20	0,05	20	0,05	19	0	21	0	20	0	20	0	20	0,94	17	0	18
Fst_96	0,38	17	0,2	20	0,94	18	0,43	20	0,5	19	0,28	20	0,83	20	0	20	0,5	19	0,19	18	0,81	16
Fst_97	0,88	17	0,97	17	0,08	18	0,5	16	0,53	17	0,5	12	0,03	19	0,32	14	0,41	17	0,07	14	0,2	5
MS_01	0,12	17	0,35	20	0,03	18	0	20	0,05	19	0,29	21	0,06	17	0,16	19	0	20	0,75	4	0,94	18
MS_02	0	17	0	20	0	20	0	20	0	19	0	21	0	20	0	20	0	20	0	20	0	18
MS_03	0	17	0,05	20	0,97	18	0	20	0	19	0	21	0,15	20	0,97	17	0	20	0,1	20	0,28	18
MS_09	0,18	17	0,25	20	0	20	0,05	20	0,11	19	0,19	21	0,15	20	0	20	0	20	0,55	20	0,94	18
MS_11	0,06	17	0	20	0,13	20	0	20	0	19	0	21	0,28	20	0,2	20	0	20	0	20	0,39	18
MS_13	0	17	0	20	0	20	0	20	0	19	0	21	0	20	0	20	0	20	0,35	20	0,33	18
MS_14	0,06	17	0,05	20	0,4	20	0	20	0,05	19	0	21	0,5	20	0,25	20	0,05	20	0,05	20	0,11	18
MS_15	0,97	17	0	20	0,32	17	0	20	0,11	19	0	21	0,45	20	0,5	20	0	20	0,05	20	0,77	15
MS_16	0,12	17	0,83	20	0,06	18	0	20	0,11	19	0	21	0	20	0,1	20	0	20	0,05	20	0,11	18
MS_17	0,06	17	0,1	20	0	20	0	20	0,95	19	0,14	21	0,05	20	0,05	20	0,98	20	0,15	20	0,42	6
MS_18	0,97	17	0	20	0,1	20	0,05	20	0,11	19	0	21	0,05	20	0	20	0	20	0	20	0,31	18
MS_19	0	17	0	20	0,15	20	0	20	0	19	0	21	0,05	20	0	20	0	20	0	20	0,94	18
MS_20	0,12	17	0,2	20	0,05	20	0,3	20	0,16	19	0,33	21	0,3	20	0	20	0,05	19	0,85	20	0,5	1

SNPID	P22_F	P22_I	P2_F	P2_I	P28_F	P28_I	P16_F	P16_I	P21_F	P21_I	P19_F	P19_I	P29_F	P29_I	P27_F	P27_I	P18_F	P18_I	P10_F	P10_I	P24_F	P24_I
MS_21	0,94	17	0	20	0,1	20	0	20	0	19	0	21	0	20	0	20	0	20	0	20	0,29	17
MS_22	0,94	16	0,1	20	0,25	20	0,1	20	0,11	19	0,05	21	0	20	0,05	20	0,2	20	0,05	20	0,2	10
MS_23	0,06	17	0	20	0,97	18	0	20	0	19	0	21	0	20	0	20	0	20	0,1	20	0,56	8
MS_24	0,06	17	0	20	0,05	20	0,05	20	0,05	19	0	21	0,05	20	0,05	20	0	20	0	20	0,29	17
MS_25	0	17	0	20	0	20	0	20	0	19	0	21	0	20	0,05	20	0	20	0	20	0	18
MS_26	0	17	0,05	20	0	20	0	20	0	19	0,05	21	0	20	0	20	0	20	0	20	0	18
MS_27	0,06	17	0,1	20	0,24	19	0,05	20	0,05	19	0,05	21	0,08	20	0,38	20	0	20	0	20	0,93	15
MS_28	0,06	17	0	20	0	20	0	20	0	19	0	21	0	20	0	20	0	20	0	20	0	18
MS_29	0,06	17	0	20	0	20	0	20	0	19	0	21	0	20	0	20	0	20	0	20	0	18
MS_30	0,29	17	0,35	20	0,55	20	0,15	20	0	19	0,52	21	0,95	20	0,5	20	0,3	20	0,85	20	0,94	18
MS_31	0,12	17	0,15	20	0,23	15	0,05	20	0	19	0,24	21	0,07	14	0,38	17	0	20	0,85	20	0,89	18
MS_32	0,47	17	0,2	20	0,39	18	0,05	20	0,05	19	0,14	21	0,1	20	0,3	20	0,05	20	0,7	20	0,89	18
MS_34	0,18	17	0,25	20	0,31	18	0,35	20	0	19	0,14	21	0,11	19	0,14	18	0,2	20	0,75	20	0,94	18
MS_35	0,06	17	0,05	20	0,55	19	0,05	20	0,11	19	0,24	21	0,2	20	0,5	17	0	20	0,07	15	0,67	18
MS_36	0	17	0,1	20	0,64	18	0	20	0	19	0	21	0,18	19	0,58	20	0	20	0,35	20	0,39	18
MS_37	0,94	16	0,38	13	0,44	18	0	20	0,05	19	0	21	0,33	20	0,53	20	0	20	0,97	16	0,72	9
MS_39	0,06	17	0,05	20	0,46	14	0	20	0,05	19	0,03	20	0,1	20	0,16	16	0,1	20	0,35	20	0,72	18
MS_40	0,06	17	0,2	20	0,3	20	0	20	0	19	0,19	21	0,2	20	0,28	18	0	20	0,25	20	0,22	18
MS_44	0	17	0,05	20	0	20	0,1	20	0	19	0	21	0	20	0	20	0,05	20	0,15	20	0	18
MS_45	0,18	17	0,2	20	0,36	18	0	20	0	19	0,24	21	0,14	14	0,39	19	0,05	20	0,65	20	0	18
MS_47	0,06	17	0	20	0,09	17	0	20	0,05	19	0,05	21	0,18	20	0,28	20	0	20	0,1	20	0,72	18
MS_48	0,06	17	0,05	20	0,5	20	0	20	0	19	0	21	0,31	18	0,58	19	0,05	20	0,15	20	0	18
MS_49	0,06	17	0	20	0,63	19	0	20	0,11	19	0	21	0,9	20	0,7	20	0,05	20	0	20	0,11	18
MS_50	0,12	17	0,61	18	0,94	18	0,53	19	0,29	17	0,08	19	0,05	20	0,88	20	0,31	18	0,25	8	0,42	6
MS_51	0,06	16	0,1	20	0,71	19	0	20	0	19	0	21	0,9	20	0,7	20	0	20	0,1	20	0,39	18
MS_55	0,14	7	0,05	20	0,68	19	0,1	20	0,16	19	0,05	21	0,88	20	0,78	20	0,05	20	0,15	20	0	18
MS_56	0,06	17	0,1	20	0,42	18	0,05	20	0	19	0,05	21	0,06	17	0,33	18	0	20	0,5	20	0,5	18
MS_57	0,12	17	0	20	0,28	20	0,83	20	0,86	18	0,85	20	0,11	19	0,2	20	0,98	20	0,39	18	0,93	15
MS_58	0,94	17	0,78	18	0,21	19	0,92	19	0,94	16	0,86	18	0,1	20	0,16	19	0,89	19	0,96	13	0,83	9
MS_59	0,24	17	0,25	20	0,53	16	0,1	20	0	19	0,38	21	0,81	18	0,73	20	0,3	20	0,33	3	0,5	1
MS_60	0,18	17	0,21	17	0,72	18	0	20	0,03	17	0,14	21	0,9	20	0,79	19	0	20	0,19	8	0,1	5
MS_61	0,18	17	0,03	16	0,92	18	0	20	0,11	19	0,38	21	0,95	19	0,79	19	0	20	0,9	20	0,94	18
MS_62	0,61	14	0,5	13	0,83	18	0,48	20	0,79	19	0,91	17	0,97	19	0,89	18	0,95	20	0,83	3	0,94	18
MS_63	0,12	17	0	20	0,29	19	0,3	20	0,16	19	0,1	21	0,11	18	0,18	19	0,97	19	0,15	20	0,22	18
MS_64	0,06	17	0,15	20	0,72	18	0,13	19	0,24	17	0,53	20	0,85	20	0,73	20	0,1	20	0,4	20	0,61	18
MS_66	0,06	17	0	20	0,76	19	0	20	0,11	19	0	21	0,9	20	0,8	20	0	20	0	20	0,08	13
MS_67	0,24	17	0,2	20	0,75	18	0,05	20	0,05	19	0,29	21	0,92	18	0,8	20	0,15	20	0,85	20	0	18
MS_68	0,24	17	0	20	0,83	20	0	20	0,17	18	0,14	21	0,95	19	0,74	19	0,03	19	0,03	16	0,63	16
MS_69	0,18	17	0,27	11	0,17	15	0,75	14	0,73	11	0,92	12	0,06	17	0,11	18	0,91	11	0,25	2	0,4	5
MS_70	0,06	17	0,05	20	0,53	19	0	20	0,05	19	0,05	21	0,13	20	0,33	20	0,05	20	0,05	20	0	18
MS_74	0,47	17	0,2	20	0,53	20	0,31	18	0,13	19	0,34	16	0,9	20	0,71	19	0,39	18	0,3	5	0,83	18



SNPID	P22_F	P22_I	P2_F	P2_I	P28_F	P28_I	P16_F	P16_I	P21_F	P21_I	P19_F	P19_I	P29_F	P29_I	P27_F	P27_I	P18_F	P18_I	P10_F	P10_I	P24_F	P24_I
MS_75	0,06	17	0	20	0,47	18	0	20	0	19	0,1	21	0,08	19	0,3	20	0,05	20	0,05	20	0,06	18
MS_76	0,06	17	0	20	0,55	20	0	20	0	19	0	21	0,9	20	0,7	20	0	20	0	20	0	18
MS_77	0,06	17	0	20	0,5	18	0	20	0,05	19	0,05	21	0,11	18	0,31	18	0	20	0	20	0,11	18
MS_78	0,06	17	0,1	20	0,55	19	0	20	0	19	0,14	21	0,9	20	0,74	17	0	20	0,1	20	0,22	18
MS_81	0,94	16	0,05	20	0,2	20	0,05	20	0	19	0	21	0,25	20	0,1	20	0	20	0,83	18	0,06	17
MS_82	0,06	17	0,97	19	0,05	20	0	20	0	19	0	21	0,5	20	0,5	20	0	20	0,82	17	0,27	11
MS_83	0,06	17	0,05	20	0,05	20	0	20	0	19	0	21	0,55	20	0,65	20	0	20	0,17	18	0,86	14
MS_87	0,06	17	0,2	20	0,21	19	0	20	0,11	19	0	21	0,37	19	0,45	19	0	20	0,05	20	0,57	15
MS_88	0,06	17	0	20	0,83	20	0,05	20	0	19	0,05	21	0,78	18	0,68	20	0,05	20	0,25	20	0,78	9
MS_89	0,24	17	0,25	20	0,33	18	0,1	20	0	19	0,24	21	0,53	16	0,38	17	0,15	20	0,3	20	0,14	7
SB_01	0,18	17	0,15	20	0,05	20	0,36	18	0	19	0,03	16	0	20	0	20	0,28	20	0,7	20	0,89	18
SB_02	0	17	0	20	0	20	0	20	0	19	0	21	0	20	0	20	0	20	0,05	20	0	18
SB_03	0,4	15	0,13	19	0,79	19	0,1	20	0,08	18	0,03	20	0,93	20	0,79	19	0,23	20	0,07	15	0,21	7
SB_04	0,38	13	0,09	17	0,89	19	0	20	0,05	19	0,14	21	0	20	0,9	20	0	20	0,85	20	0,5	1
SB_07	0,83	3	0,35	20	0,03	17	0,05	20	0,11	19	0,43	21	0,03	19	0,1	20	0,3	20	0,67	6	0,94	18
SB_08	0,13	15	0,2	15	0,89	19	0,47	18	0,56	18	0,61	19	0,97	19	0,83	18	0,61	18	0,69	8	0,63	12
SB_09	0	17	0,7	20	0,03	20	0,33	18	0,15	17	0,19	18	0,03	20	0	20	0,05	19	0,35	17	0,19	16
SB_10	0,55	10	0,19	16	0,98	20	0,63	20	0,72	16	0,33	15	0,98	20	0,95	20	0,44	18	0,5	7	0,56	18
SB_12	0	17	0	20	0,05	20	0,34	19	0,38	17	0,21	19	0	20	0	20	0,16	19	0	20	0	18
SB_13	0	17	0	20	0	20	0,05	20	0	19	0	21	0	20	0	20	0,05	20	0	20	0	18
SB_14	0,18	17	0,05	20	0	20	0	20	0,11	19	0,1	21	0,98	20	0	20	0	20	0	20	0,11	18
SB_15	0,06	17	0,05	20	0,71	19	0	20	0	19	0	21	0,97	18	0,73	20	0	20	0,2	15	0,44	9
SB_16	0,88	16	0	20	0,03	20	0,15	20	0,19	18	0,21	21	0,03	20	0,03	19	0,2	20	0,25	16	0,11	18
SB_20	0	17	0	20	0	20	0	20	0	19	0	21	0	20	0	20	0	20	0,1	20	0	18
SB_22	0,12	17	0,15	20	0,21	19	0,05	20	0	19	0,33	21	0,38	16	0,29	19	0	20	0,93	7	0,88	4
SB_25	0,12	17	0,13	16	0,17	18	0,13	20	0,21	19	0,12	17	0,14	18	0,15	20	0,03	18	0,58	6	0,89	18
SB_26	0,06	17	0,75	20	0	20	0,84	19	0,87	19	0,79	19	0	20	0,05	20	0,9	20	0	20	0,17	18
SB_27	0,06	17	0,15	20	0,86	18	0	20	0,11	19	0,1	21	0,74	19	0,45	20	0	20	0,08	6	0,5	2
SB_28	0,11	9	0,15	20	0,22	18	0,03	19	0	19	0,38	21	0,15	13	0,08	20	0,03	18	0,67	3	0,94	18
SB_29	0,12	17	0,46	14	0,47	18	0,05	20	0	19	0,1	21	0,43	15	0,42	18	0	20	0,5	20	0,5	2
SB_30	0,24	17	0,15	20	0,5	18	0,1	20	0	19	0,14	21	0,53	17	0,72	18	0,25	20	0,25	20	0,91	11
SB_32	0,73	15	0,15	20	0	20	0,82	19	0,83	18	0,24	21	0	20	0	20	0,81	18	0,33	3	0,89	18
SB_33	0	17	0,1	20	0	20	0	20	0	19	0	21	0	20	0	20	0	20	0,05	20	0,06	18
SB_34	0,06	17	0,15	20	0,98	20	0,98	20	0,97	17	0,98	21	0,9	20	0,98	20	0,97	19	0,65	20	0,56	18
SB_35	0,5	4	0,45	20	0,98	20	0,67	15	0,73	15	0,71	12	0,3	20	0	20	0,77	13	0,85	20	0,94	18
SB_38	0,06	17	0,89	18	0,1	20	0,28	20	0,42	18	0,6	20	0,25	20	0,05	20	0,28	20	0,37	19	0,15	17
SB_39	0,12	17	0,97	17	0,98	20	0,7	15	0,76	17	0,87	15	0,05	20	0,95	20	0,81	18	0,65	20	0,83	18
SB_41	0,06	17	0,31	13	0	20	0,63	16	0,33	18	0,25	18	0	20	0	20	0,18	20	0,9	20	0,94	18
SB_43	0,97	15	0,81	16	0,33	18	0,83	20	0,76	19	0,74	17	0,53	18	0,42	19	0,88	20	0,83	9	0,67	18
SB_44	0,41	17	0,45	20	0,15	20	0,53	19	0,41	17	0,36	14	0,32	19	0,2	20	0,23	20	0,05	10	0,67	18
SB_45	0,71	17	0	20	0,15	20	0,8	20	0,68	19	0	21	0,05	20	0,35	20	0,45	20	0	20	0	18

SNPID	P22_F	P22_I	P2_F	P2_I	P28_F	P28_I	P16_F	P16_I	P21_F	P21_I	P19_F	P19_I	P29_F	P29_I	P27_F	P27_I	P18_F	P18_I	P10_F	P10_I	P24_F	P24_I
SB_46	0,68	14	0,7	15	0,61	18	0,45	19	0,64	18	0,65	17	0,1	15	0,45	20	0,59	17	0,33	3	0,5	1
SB_47	0,24	17	0,8	15	0,42	18	0	20	0	19	0	21	0,34	16	0,56	17	0,05	20	0,97	18	0,72	9
SB_48	0	17	0,08	18	0,03	20	0,08	20	0,16	19	0,22	18	0,03	20	0,15	20	0,03	20	0,03	19	0,09	17
SB_49	0,47	17	0,2	20	0,91	16	0	20	0,11	19	0,03	20	0,88	20	0,89	18	0	20	0,65	20	0,44	18
SB_50	0	17	0,38	17	0	20	0,95	20	0,94	17	0,97	16	0	20	0,05	20	0,91	17	0,65	20	0,61	18
SB_51	0,06	17	0	20	0	20	0	20	0	19	0	21	0	20	0,05	20	0	20	0,1	20	0,28	18
SB_52	0	17	0,91	16	0	20	0,05	20	0	19	0	21	0	20	0,05	20	0	20	0,25	20	0,95	10
SB_54	0,25	14	0,05	20	0,75	20	0	20	0,03	17	0,1	21	0,71	17	0,55	20	0	20	0,45	20	0,72	18
SB_55	0,38	17	0	20	0,97	18	0	20	0,03	17	0,02	21	0,95	20	0,95	20	0,25	20	0,11	18	0,97	16
SB_56	0,13	16	0,58	18	0,16	19	0,93	20	0,97	15	0,94	17	0,05	19	0,25	20	0,69	18	0,3	15	0,68	11
SB_57	0,53	17	0,35	20	0,94	18	0,1	20	0	19	0,33	21	0,88	17	0,87	19	0,05	20	0,08	6	0,5	1
SB_58	0,41	17	0,25	20	0,05	20	0,05	20	0	19	0,29	21	0,76	19	0,92	19	0	20	0,1	5	0,89	18
SB_59	0,06	17	0	20	0	20	0	20	0	19	0	21	0,23	20	0,21	19	0	20	0,1	20	0,06	18
SB_60	0	17	0,71	19	0	20	0,93	20	0,74	19	0,8	20	0,9	20	0,9	20	0,78	20	0,78	18	0,97	18
SB_61	0	17	0	20	0,05	20	0	20	0	19	0	21	0	20	0	20	0,05	20	0,05	20	0,06	18
SB_63	0,39	14	0,63	15	0,98	20	0	20	0,05	19	0	21	0	20	0,95	20	0	20	0,15	20	0,06	18
SB_64	0,06	17	0,15	20	0,58	20	0,65	20	0,91	16	0,81	18	0,37	19	0,36	18	0,66	19	0,64	7	0,5	18
SB_65	0,86	14	0,2	20	0	20	0,19	18	0,21	17	0,03	18	0	20	0,03	20	0	20	0	20	0	18
SB_66	0	17	0,24	17	0,98	20	0	20	0	19	0	21	0,05	20	0	20	0	20	0,94	16	0	18
SB_69	0,65	17	0,76	19	0,93	20	0,8	20	0,81	18	0,76	21	0,9	20	0,95	20	0,66	19	0,77	15	0,06	18
SB_70	0	17	0	20	0	20	0	20	0	19	0	21	0,95	19	0	20	0	20	0	20	0	18
SB_71	0,06	17	0	20	0,05	20	0	20	0	19	0	21	0,03	18	0,03	20	0,1	20	0,15	20	0,11	18
SB_72	0,38	16	0	20	0,08	20	0,4	20	0,11	19	0,17	21	0,1	20	0,18	20	0,15	20	0,21	19	0,06	18
SB_73	0	17	0	20	0	20	0	20	0	19	0,05	21	0,95	19	0	20	0	20	0	20	0,06	18
SB_74	0,88	16	0,1	20	0	20	0,05	20	0	19	0	21	0,95	20	0	20	0	20	0,1	20	0,06	18
SB_75	0,18	17	0,25	20	0	20	0,05	20	0	19	0,24	21	0,03	19	0	20	0	20	0,45	20	0,61	18
SB_76	0,22	16	0,03	17	0,1	20	0,08	20	0,22	16	0,29	19	0	20	0,88	20	0,15	20	0,44	8	0,22	18
SB_77	0	17	0	20	0	20	0	20	0	19	0	21	0	20	0	20	0	20	0	20	0	18
SB_79	0,06	17	0	20	0,05	20	0,1	20	0,75	12	0,96	13	0,35	20	0,08	20	0,94	17	0,98	20	0,17	18
SB_80	0	17	0	20	0,92	19	0	20	0,05	19	0	21	0,66	19	0,9	20	0	20	0	20	0,74	17

SNPID	P20_F	P20_I	P31_F	P31_I	P26_F	P26_I	P23_F	P23_I	P14_F	P14_I	PIP1_F	PIP1_I	P12_F	P12_I	PIM1_F	PIM1_I	P5_F	P5_I	P37_F	P37_I	P36_F	P36_I
Fst_03	0,86	18	0,35	20	0,15	20	0,9	20	0,95	20	0,95	20	0	20	0	20	0,75	20	0	10	0	19
Fst_04	0,17	18	0,69	16	0,92	18	0,2	20	0,75	20	0,83	3	0,65	20	0,35	20	0,13	15	0,89	9	0,55	19
Fst_05	0,03	18	0,93	15	0,94	16	0,06	18	0	20	0,95	20	0	20	0,2	20	0,1	10	0,1	10	0,97	16
Fst_07	0,66	16	0,93	14	0,94	18	0,91	17	0	20	0,95	20	0	20	0	20	0,6	20	0,8	10	0	19
Fst_08	0,05	20	0,97	19	0	20	0	20	0	20	0,1	20	0	20	0	20	0	20	0	10	0,75	16
Fst_09	0,94	17	0,2	20	0	20	0,05	20	0,05	20	0,05	20	0	20	0	20	0	20	0	10	0	19
Fst_12	0,53	19	0,03	17	0,1	20	0,74	17	0,45	11	0,56	9	0,09	11	0,25	20	0,36	18	0,05	10	0,08	18
Fst_13	0,29	19	0,13	19	0,13	20	0,05	20	0,26	17	0,28	18	0,05	20	0,05	20	0,03	19	0,3	10	0,08	19
Fst_14	0,05	20	0,15	20	0,33	20	0	20	0,88	20	0,92	19	0	20	0,98	20	0	20	0,7	10	0,5	18
Fst_16	0,21	17	0,03	15	0,11	18	0,39	14	0	20	0,5	1	0	20	0,5	16	0,06	16	0,15	10	0,14	18
Fst_18	0,92	19	0,57	15	0,79	19	0,5	20	0,15	20	0,35	20	0,05	20	0,82	19	0,81	16	0,45	10	0,53	19
Fst_19	0,05	20	0,91	17	0,95	20	0,1	20	0,45	11	0,5	8	0,6	10	0,45	20	0,61	14	0,95	10	0,84	19
Fst_20	0,2	20	0,13	15	0	20	0,03	20	0,95	20	0,85	20	0,9	20	0,03	16	0,05	10	0,1	10	0,37	19
Fst_21	0,79	19	0,08	20	0,5	20	0,95	20	0,8	20	0,53	18	0	20	0,33	20	0,44	17	0,05	10	0,03	17
Fst_22	0,1	20	0,25	14	0,46	14	0,05	20	0,45	20	0,75	20	0,75	20	0,45	19	0,79	7	0,55	10	0,33	18
Fst_23	0,1	20	0,05	20	0	20	0	20	0,06	17	0,03	18	0,05	20	0	20	0,11	19	0	10	0	19
Fst_24	0,38	16	0,08	13	0,24	19	0,03	16	0,95	20	0,95	20	0	20	0	20	0,27	11	0,75	10	0,55	19
Fst_25	0,05	20	0,24	17	0,03	20	0,03	20	0,03	20	0,3	20	0,1	20	0,03	20	0,05	20	0,1	10	0,03	19
Fst_27	0,05	20	0,05	20	0	20	0	20	0	20	0,05	20	0	20	0	20	0	20	0	10	0	19
Fst_28	0,03	18	0,96	13	0,93	14	0,08	19	0,35	20	0,5	2	0,65	20	0,5	16	0,59	16	0	10	0,87	19
Fst_29	0,75	18	0,34	16	0,33	20	0,9	20	0,25	20	0,5	2	0,2	20	0,38	20	0,05	20	0,05	10	0,11	19
Fst_30	0,64	18	0,94	18	0	20	0	20	0,2	20	0,3	20	0,1	20	0,95	19	0	20	0	10	0	19
Fst_31	0,56	18	0,54	14	0,76	19	0,98	20	0,9	20	0,5	1	0	20	0,58	20	0,83	18	0,83	9	0,86	18
Fst_32	0,22	18	0,87	19	0,9	20	0,05	20	0,92	18	0,85	17	0	20	0	20	0,61	18	0,95	10	0,68	19
Fst_34	0,37	15	0,04	13	0,15	20	0,03	17	0	20	0,8	20	0	20	0,11	19	0,33	3	0	10	0	19
Fst_35	0,18	19	0,87	15	0,94	17	0,82	17	0,23	11	0,63	15	0,35	20	0,89	19	0,56	17	0,72	9	0,86	18
Fst_36	0,57	15	0,37	15	0,2	20	0,25	18	0,2	20	0,23	11	0,05	20	0,43	20	0,5	4	0,5	10	0,13	19
Fst_37	0,31	16	0,8	15	0,97	19	0,18	19	0,56	18	0,4	15	0,93	20	0,59	17	0,82	19	0,1	10	0,97	18
Fst_38	0,1	20	0,3	20	0	20	0	20	0,1	20	0	20	0	20	0	20	0,95	20	0	10	0	19
Fst_40	0,86	18	0,36	14	0,83	20	0	20	0,2	20	0,89	14	0,15	20	0,89	19	0,5	20	0	10	0,26	19
Fst_41	0,09	17	0,26	17	0,5	20	0	20	0,5	6	0,67	3	0,05	20	0,79	17	0,31	16	0,89	9	0,87	19
Fst_45	0,16	19	0,88	20	0,75	20	0	20	0	20	0,93	14	0	20	0,8	20	0,79	19	0	10	0,97	19
Fst_46	0,53	18	0,27	15	0,71	17	0,38	17	0,33	12	0,56	8	0,05	20	0,45	20	0,33	18	0,4	10	0,16	19
Fst_47	0,27	15	0,82	17	0	20	0,08	19	0,59	11	0,35	20	0	20	0,89	19	0,63	8	0,35	10	0,79	19
Fst_49	0,74	19	0	20	0	20	0,39	19	0,64	14	0	20	0	20	0,97	19	0,85	17	0	10	0	19
Fst_50	0,1	20	0,41	16	0,08	20	0	20	0,1	20	0,2	20	0	20	0,65	20	0,15	20	0	10	0,18	19
Fst_51	0,1	20	0,54	14	0,89	19	0,95	20	0,2	20	0,25	2	0,2	20	0,23	20	0,67	18	0	10	0,03	19
Fst_53	0,05	20	0,88	16	0	20	0,05	19	0,2	20	0,04	13	0,2	20	0,44	17	0,1	15	0	10	0,82	19
Fst_54	0,5	19	0,97	15	0,98	20	0,78	20	0,5	20	0,77	11	0,1	20	0	20	0,88	16	0	10	0,97	19
Fst_55	0,67	18	0,35	13	0,19	18	0,37	19	0,6	20	0,93	7	0,6	20	0,71	17	0,9	15	0,2	10	0,18	19
Fst_56	0,1	20	0,77	15	0	20	0,03	20	0,2	20	0,2	20	0,1	20	0,13	20	0	20	0,85	10	0,47	19

SNPID	P20_F	P20_I	P31_F	P31_I	P26_F	P26_I	P23_F	P23_I	P14_F	P14_I	PIP1_F	PIP1_I	P12_F	P12_I	PIM1_F	PIM1_I	P5_F	P5_I	P37_F	P37_I	P36_F	P36_I
Fst_57	0,67	18	0,95	20	0	20	0	20	0,74	17	0,91	17	0,1	20	0	20	0,97	19	0	10	0,95	19
Fst_58	0,05	20	0,21	19	0,33	20	0,05	20	0	20	0,1	20	0	20	0,93	20	0,98	20	0	10	0	19
Fst_59	0,03	17	0,62	13	0,76	19	0,05	19	0	20	0,95	20	0	20	0,86	18	0,19	13	0,9	10	0,87	19
Fst_60	0,95	20	0,18	19	0,23	20	0,98	20	0,05	20	0	20	0	20	0,74	19	0	20	0	10	0,13	19
Fst_61	0,28	18	0,94	16	0,95	19	0,75	16	0,95	20	0,86	7	0,9	20	0,89	19	0,6	20	0	10	0	19
Fst_62	0,18	17	0,03	15	0	20	0,21	19	0,95	20	0	20	0	20	0,08	20	0,08	18	0	10	0	19
Fst_66	0,97	18	0,57	15	0,23	20	0,98	20	0,5	20	0,8	20	0,25	20	0,73	13	0,93	14	0,75	10	0,47	19
Fst_68	0,45	20	0,5	20	0,5	20	0,35	20	0,9	20	0,95	20	0	20	0,5	20	0,35	20	0,5	10	0,89	19
Fst_69	0,95	19	0,27	15	0,42	19	0,39	19	0,85	20	0,85	20	0,7	20	0,5	20	0,83	15	0,2	10	0,11	19
Fst_70	0,11	14	0,64	11	0,78	20	0,6	20	0	20	0,95	20	0	20	0,69	16	0,9	20	0,64	7	0,94	17
Fst_71	0,05	20	0	20	0	20	0	20	0	20	0,05	20	0	20	0	20	0	20	0	10	0,11	19
Fst_73	0,05	20	0	20	0	20	0	20	0	20	0,2	20	0,15	20	0	20	0,05	20	0	10	0	19
Fst_74	0,15	20	0,47	16	0,35	17	0,95	20	0,2	20	0,36	7	0,05	20	0,11	19	0,73	13	0,22	9	0,08	19
Fst_75	0,94	17	0,26	19	0,18	20	0	20	0,83	18	0,47	19	0	20	0,23	20	0,5	15	0,72	9	0,24	19
Fst_76	0,58	18	0,08	19	0,34	19	0,82	17	0,85	20	0,67	3	0,85	20	0,1	20	0,11	9	0,1	10	0,14	18
Fst_77	0,1	20	0,35	20	0,1	20	0	20	0	20	0,95	20	0	20	0	20	0,15	20	0	10	0	19
Fst_78	0,41	17	0,81	16	0,83	20	0,05	20	0,17	12	0,79	12	0,3	20	0,93	20	0,66	16	0	10	0	19
Fst_80	0,1	20	0,71	14	0,85	20	0	20	0,04	14	0,23	11	0,05	20	0,5	19	0,22	16	0,1	10	0,69	18
Fst_82	0,26	19	0,71	19	0,48	20	0,33	20	0,22	18	0,75	18	0,05	20	0,76	17	0,86	18	0,95	10	0,74	19
Fst_83	0,05	20	0,1	20	0	20	0,05	20	0,15	20	0,1	20	0,05	20	0	20	0,05	20	0	10	0,97	18
Fst_84	0,05	20	0,05	20	0	20	0	20	0	20	0,15	20	0,15	20	0,05	20	0	20	0	10	0,68	19
Fst_85	0,05	20	0,05	20	0	20	0	20	0,3	20	0,15	20	0,05	20	0	20	0,15	20	0	10	0,18	19
Fst_87	0,1	20	0,04	14	0,05	20	0,1	20	0	20	0,95	20	0,95	20	0	20	0,25	20	0	10	0,83	18
Fst_88	0,95	19	0,29	19	0,38	20	0,93	20	0,87	19	0,4	20	0,05	20	0,29	19	0,81	18	0	10	0,06	18
Fst_89	0,05	20	0,05	20	0,05	20	0	20	0	20	0,05	20	0,05	20	0	20	0	20	0,1	10	0,19	18
Fst_90	0,15	20	0,07	14	0,05	20	0,03	18	0,08	13	0,14	11	0,94	17	0,47	19	0,2	15	0	10	0,82	19
Fst_91	0,25	20	0,35	20	0,1	20	0,25	20	0	20	0,95	20	0	20	0,5	20	0,65	20	0	10	0,82	19
Fst_92	0,28	18	0,97	19	0,8	20	0,05	20	0,05	20	0,17	9	0,06	18	0,97	19	0,33	18	0,85	10	0,86	18
Fst_94	0,05	20	0,1	20	0	20	0	20	0,1	20	0,05	20	0,05	20	0	20	0	20	0	10	0,82	19
Fst_95	0,05	20	0	20	0	20	0,53	20	0,05	20	0,05	20	0,5	20	0	20	0	20	0	10	0	19
Fst_96	0,28	16	0,88	20	0,93	20	0,18	19	0,85	17	0,78	18	0,5	19	0,8	20	0,82	14	0	10	0	19
Fst_97	0,62	17	0,04	13	0,06	17	0,75	6	0,93	14	0,3	20	0,97	15	0,14	18	0,33	6	0,36	7	0,11	18
MS_01	0,1	20	0,23	13	0,29	12	0,95	20	0	20	0	20	0	20	0,05	20	0,93	15	0	10	0,09	17
MS_02	0,05	20	0,05	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	10	0	19
MS_03	0,05	20	0,45	20	0,92	13	0	20	0,3	20	0,2	20	0,1	20	0	20	0,05	20	0	10	0	19
MS_09	0,1	20	0,5	20	0,35	20	0,94	18	0,75	20	0,8	20	0,95	20	0,05	20	0,2	20	0,4	10	0,14	18
MS_11	0,05	20	0,5	17	0,53	19	0	20	0,35	20	0	20	0,05	20	0	20	0,03	19	0	10	0	19
MS_13	0,05	20	0,05	20	0	20	0	20	0,15	20	0,4	20	0,65	20	0	20	0	20	0	10	0	19
MS_14	0,05	20	0,63	16	0,64	18	0	20	0,09	17	0	20	0	20	0	20	0	20	0	10	0	19
MS_15	0,05	20	0,25	18	0,4	20	0,98	20	0	20	0	20	0	20	0	20	0	20	0,9	10	0,84	19
MS_16	0,05	20	0,03	18	0,05	19	0,95	20	0,3	20	0,83	6	0	20	0	20	0,38	16	0	10	0,08	19

SNPID	P20_F	P20_I	P31_F	P31_I	P26_F	P26_I	P23_F	P23_I	P14_F	P14_I	PIP1_F	PIP1_I	P12_F	P12_I	PIM1_F	PIM1_I	P5_F	P5_I	P37_F	P37_I	P36_F	P36_I
MS_17	0,1	20	0,1	15	0	20	0,98	20	0,3	20	0,55	20	0,1	20	0,05	20	0,85	13	0	10	0,05	19
MS_18	0,05	20	0,1	20	0	20	0,98	20	0	20	0,95	20	0	20	0,29	19	0,95	20	0	10	0	19
MS_19	0	20	0,4	20	0,05	20	0,95	20	0	20	0,95	20	0	20	0,1	20	0	20	0	10	0	19
MS_20	0,2	20	0,89	14	0,05	20	0,03	17	0	20	0,5	1	0,95	20	0,89	19	0,7	20	0	10	0,11	19
MS_21	0,05	20	0,08	19	0	20	0,98	20	0	20	0,97	19	0	20	0,13	20	0,95	20	0	10	0	19
MS_22	0	20	0,08	19	0,05	20	0,97	19	0,1	20	0,97	16	0,25	20	0,13	15	0,97	15	0,1	10	0,11	19
MS_23	0,1	20	0,91	17	0	20	0,03	20	0	20	0,03	18	0	20	0,53	19	0,03	19	0	10	0,95	19
MS_24	0,05	20	0,08	19	0	20	0,98	20	0	20	0,97	19	0,05	20	0,13	16	0,97	17	0	10	0	19
MS_25	0,05	20	0,05	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	10	0	19
MS_26	0,05	20	0	20	0	20	0	20	0	20	0,05	20	0	20	0,05	20	0	20	0	10	0	19
MS_27	0,1	20	0,03	16	0,08	19	0,1	20	0	20	0,1	20	0,05	20	0,47	19	0,7	15	0	10	0,05	19
MS_28	0,05	20	0	20	0	20	0,05	20	0	20	0,05	20	0	20	0	20	0	20	0	10	0,05	19
MS_29	0,05	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	10	0	19
MS_30	0,15	20	0,95	20	0,65	20	0,15	20	0	20	0,95	20	0	20	0,6	20	0	20	0	10	0	19
MS_31	0,1	20	0,08	13	0,26	17	0,1	20	0,95	20	0,9	20	0,95	20	0,31	18	0,5	20	0	10	0,05	19
MS_32	0,1	20	0,07	15	0,37	19	0,15	20	0,9	20	0,85	20	0,9	20	0,13	19	0,35	20	0,05	10	0,11	19
MS_34	0,1	20	0,03	15	0,44	18	0,15	20	0,85	20	0,95	20	0,9	20	0,13	20	0,1	20	0,05	10	0,11	18
MS_35	0,05	20	0,81	16	0,86	18	0,05	20	0,03	20	0,2	20	0,1	20	0	20	0,3	15	0	10	0,03	19
MS_36	0,1	20	0,84	16	0,92	19	0	20	0,2	20	0,15	20	0,05	20	0,15	20	0,3	20	0	10	0,03	19
MS_37	0,1	20	0,26	17	0	20	0,98	20	0,97	17	0,15	20	0,05	20	0,83	20	0,05	20	0,15	10	0,53	19
MS_39	0,05	20	0,17	15	0,05	20	0	20	0,25	4	0,35	20	0,25	20	0	20	0	20	0	10	0,05	19
MS_40	0,1	20	0,46	14	0,16	19	0	20	0,1	20	0	20	0,05	20	0	20	0,25	20	0	10	0	19
MS_44	0,05	20	0,05	20	0	20	0	20	0,15	20	0,1	20	0,1	20	0,05	20	0,05	20	0	10	0	19
MS_45	0,1	20	0,07	15	0,44	16	0,1	20	0	20	0,95	20	0,95	20	0,2	20	0,4	20	0,25	10	0,58	18
MS_47	0,1	20	0,03	15	0,18	17	0,95	20	0,55	20	0,92	13	0,05	20	0,48	20	0,95	19	0,2	10	0,29	19
MS_48	0,05	20	0,5	16	0,45	20	0,05	20	0,05	20	0,15	20	0,05	20	0,5	20	0,05	20	0,05	10	0,11	19
MS_49	0,05	20	0,95	20	0,66	19	0	20	0	20	0	20	0	20	0,89	19	0,05	20	0,95	10	0,89	19
MS_50	0,11	18	0,25	20	0,7	20	0,03	20	0,45	20	0,5	10	0,15	20	0	20	0,33	15	0,95	10	0,95	19
MS_51	0,1	20	0,95	19	0,63	19	0,25	20	0,1	20	0,05	20	0,05	20	0,88	20	0,05	20	0	10	0,92	19
MS_55	0,05	20	0,97	19	0,85	17	0,23	13	0,2	20	0,05	20	0	20	0,85	20	0	20	0	10	0,92	19
MS_56	0,05	20	0,08	13	0,41	16	0	20	0,4	20	0,65	20	0,35	20	0,28	16	0,15	20	0	10	0,06	17
MS_57	0,95	19	0,03	18	0,34	19	0	20	0,83	18	0,94	16	0	20	0,2	15	0,89	18	0,15	10	0,08	19
MS_58	0,94	18	0,03	18	0,32	19	0	20	0,93	15	0,45	20	0,15	20	0,15	20	0,92	19	0,05	10	0,11	19
MS_59	0,1	20	0,71	14	0,64	18	0,2	20	0	20	0,95	20	0	20	0,9	20	0,4	20	0,95	10	0,97	19
MS_60	0,1	20	0,93	15	0,73	20	0	20	0,4	20	0,08	6	0,45	20	0,92	19	0,69	8	0	10	0,95	19
MS_61	0,1	20	0,45	20	0,88	16	0,06	16	0	20	0,95	20	0,95	20	0,88	20	0,95	20	0,94	9	0,94	18
MS_62	0,79	17	0,3	20	0,84	19	0,83	15	0	20	0,95	20	0,95	20	0,9	20	0,33	6	0,95	10	0,95	19
MS_63	0,15	20	0,07	14	0,32	19	0	20	0	20	0,05	20	0,15	20	0,06	18	0,1	20	0,05	10	0,11	19
MS_64	0,41	17	0,85	13	0,63	19	0,05	20	0,4	20	0,45	20	0,3	20	0,92	19	0,24	17	0,95	10	0,89	19
MS_66	0,05	20	0,95	19	0,88	20	0,03	20	0,05	20	0,03	18	0	20	0,88	20	0,05	20	0,6	10	0,86	18

SNPID	P20_F	P20_I	P31_F	P31_I	P26_F	P26_I	P23_F	P23_I	P14_F	P14_I	PIP1_F	PIP1_I	P12_F	P12_I	PIM1_F	PIM1_I	P5_F	P5_I	P37_F	P37_I	P36_F	P36_I
MS_67	0,1	20	0,92	13	0,88	17	0,1	20	0	20	0,95	20	0	20	0,97	19	0,05	10	0,7	10	0,88	16
MS_68	0,03	17	0,95	19	0,75	20	0,18	19	0,05	20	0,15	13	0	20	0,98	20	0,18	19	0	10	0,97	19
MS_69	0,59	16	0,08	12	0,25	16	0,97	18	0,8	20	0,67	6	0,55	20	0,41	17	0,67	9	0,5	9	0,5	15
MS_70	0,1	20	0,06	16	0,37	19	0	20	0	20	0,05	20	0	20	0,13	20	0,05	20	0,11	9	0,22	18
MS_74	0,15	17	0,96	14	0,63	19	0,2	20	0,55	20	0,65	20	0,6	20	0,97	15	0,35	20	0,9	10	0,86	18
MS_75	0,05	20	0,07	15	0,37	19	0	20	0	20	0	20	0,05	20	0,13	19	0	20	0,1	10	0,13	19
MS_76	0,1	20	0,95	19	0,63	20	0	20	0	20	0	20	0	20	0,88	20	0	20	0,9	10	0,87	19
MS_77	0,05	20	0,06	17	0,37	19	0	20	0	20	0,05	20	0	20	0,13	19	0	20	0,1	10	0,13	19
MS_78	0,05	20	0,96	14	0,63	19	0,05	20	0,1	20	0,25	20	0,1	20	0,88	20	0,05	20	0,94	9	0,87	19
MS_81	0,05	20	0,13	20	0,23	20	0,95	20	0,05	20	0,93	20	0,05	20	0	20	0,72	18	0,2	10	0,11	19
MS_82	0,1	20	0,13	20	0,28	20	0,95	19	0	20	0	20	0	20	0,05	20	0,95	20	0	10	0	19
MS_83	0,15	20	0,88	16	0,73	20	0,05	20	0,05	20	0,15	20	0,05	20	0	20	0	20	0	10	0	19
MS_87	0,05	20	0,3	20	0,5	20	0,95	20	0,05	20	0,05	20	0	20	0	20	0	20	0,95	10	0,97	19
MS_88	0,1	20	0,22	18	0,39	19	0,95	20	0,1	20	0	20	0,1	20	0	20	0	20	0,2	10	0	19
MS_89	0,15	20	0,54	13	0,47	18	0,95	20	0,15	20	0,3	20	0,05	20	0,05	20	0,15	20	0	10	0	19
SB_01	0,12	17	0,35	20	0	20	0,05	20	0,85	20	0,9	20	0,95	20	0	20	0,5	20	0	10	0	19
SB_02	0,05	20	0	20	0	20	0	20	0	20	0	20	0,5	19	0	20	0	20	0	10	0,05	19
SB_03	0,13	20	0,89	19	0,9	20	0,38	20	0,85	13	0,17	12	0	20	0,88	20	0	20	0,65	10	0,74	19
SB_04	0,1	20	0,25	20	0,97	17	0,39	19	0,95	20	0,5	1	0,95	20	0	20	0,2	20	0,9	10	0	19
SB_07	0,1	20	0,21	14	0,11	19	0,92	12	0,95	20	0,95	20	0	20	0	20	0,5	3	0	10	0,03	19
SB_08	0,58	18	0,91	17	0,83	20	0,17	18	0,22	9	0,9	20	0,8	20	0,93	20	0,9	10	0	10	0,67	18
SB_09	0,21	17	0,03	18	0	20	0,05	20	0,03	20	0,11	19	0,5	16	0	20	0,03	20	0	10	0	19
SB_10	0,58	18	0,25	20	0	20	0,79	17	0,55	20	0,55	20	0,9	20	0,9	20	0,77	11	0,5	10	0,58	18
SB_12	0,18	19	0	20	0	20	0	20	0,03	18	0,03	20	0	20	0	20	0,05	19	0	10	0	19
SB_13	0,15	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	10	0	19
SB_14	0,1	20	0,1	20	0	20	0,1	20	0,05	20	0,25	20	0,45	20	0,2	20	0,05	20	0	10	0,97	19
SB_15	0,05	20	0,94	17	0,86	14	0,05	20	0,15	20	0,2	20	0,05	20	0,17	18	0,15	20	0,9	10	0,82	19
SB_16	0,03	18	0,1	20	0	20	0,8	20	0,08	18	0,09	17	0	20	0,05	20	0,13	19	0	10	0,03	19
SB_20	0,05	20	0,05	20	0	20	0	20	0,1	20	0	20	0,1	20	0	20	0	20	0,1	10	0	19
SB_22	0,1	20	0,25	14	0,35	17	0,1	20	0	20	0,95	20	0	20	0,9	20	0,4	20	0,9	10	0,84	19
SB_25	0,15	20	0,37	15	0,61	19	0,1	20	0,9	20	0	20	0,95	20	0,2	20	0,33	6	0,6	10	0,53	19
SB_26	0,89	19	0,97	19	0,98	20	0,03	19	0,89	9	0,2	20	0	20	0,98	20	0,81	18	0	10	0,97	19
SB_27	0,1	20	0,86	14	0,55	19	0,05	20	0,85	20	0,85	20	0,75	20	0,25	20	0,2	20	0,8	10	0,61	18
SB_28	0,03	18	0,29	14	0,08	19	0,08	20	0	20	0,95	20	0	20	0,13	20	0	20	0	10	0,11	19
SB_29	0,05	20	0,5	14	0,44	18	0,95	20	0	20	0,5	1	0	20	0,09	17	0,56	9	0	10	0,05	19
SB_30	0,05	20	0,45	11	0,9	15	0,15	20	0,1	20	0,2	20	0,3	20	0,89	18	0,3	20	0,06	9	0,13	12
SB_32	0,86	18	0,25	20	0,05	20	0,64	18	0	20	0,95	20	0,85	20	0	20	0,55	20	0	10	0,97	19
SB_33	0,05	20	0,05	20	0	20	0	20	0,15	20	0	20	0,1	20	0	20	0	20	0	10	0	19
SB_34	0,94	18	0,15	20	0,1	20	0,08	20	0,03	17	0,5	20	0,1	20	0,33	20	0,05	20	0	10	0	19
SB_35	0,71	17	0,3	20	0,05	20	0,73	11	0	20	0,95	20	0	20	0,92	19	0,9	20	0	10	0,89	18
SB_38	0,58	19	0,32	19	0	20	0,2	20	0,29	19	0,16	19	0,08	19	0,05	20	0,1	20	0,2	10	0,13	19

SNPID	P20_F	P20_I	P31_F	P31_I	P26_F	P26_I	P23_F	P23_I	P14_F	P14_I	PIP1_F	PIP1_I	P12_F	P12_I	PIM1_F	PIM1_I	P5_F	P5_I	P37_F	P37_I	P36_F	P36_I
SB_39	0,92	18	0,3	20	0	20	0,05	20	0,7	20	0,85	20	0	20	0	20	0,25	20	0	10	0	19
SB_41	0,17	18	0,3	20	0,05	20	0	20	0	20	0,95	20	0	20	0	20	0,2	20	0	10	0	19
SB_43	0,72	18	0,61	14	0,78	20	0,92	19	0,7	20	0,92	6	0,5	20	0,1	20	0,33	9	0,4	10	0,21	19
SB_44	0,28	18	0,1	15	0,18	20	0	20	0,8	20	0,95	20	0,85	20	0	20	0,06	8	0,15	10	0,44	17
SB_45	0,7	20	0,5	20	0,15	20	0,5	20	0	20	0	20	0,95	20	0,6	20	0	20	0	10	0	19
SB_46	0,56	18	0,75	14	0,87	19	0,88	16	0	20	0,95	20	0	20	0,53	19	0,4	20	0	10	0,03	18
SB_47	0,1	20	0,2	15	0,15	20	0,98	20	0,35	20	0,15	20	0	20	0,88	20	0	20	0,21	7	0,56	16
SB_48	0,08	19	0,03	19	0,05	20	0	20	0,29	19	0,25	18	0,75	20	0,2	20	0,1	20	0,05	10	0,33	18
SB_49	0,15	20	0,94	16	0,92	18	0,09	11	0,45	20	0,71	7	0,25	20	0,89	19	0,76	17	0	10	0	19
SB_50	0,94	16	0,15	20	0,05	20	0	20	0,7	20	0,9	5	0,8	20	0	20	0	20	0	10	0	19
SB_51	0,05	20	0	20	0	20	0	20	0	20	0	20	0,15	20	0	20	0,05	20	0	10	0	19
SB_52	0,05	20	0,15	20	0,1	20	0	20	0,2	20	0,2	20	0,05	20	0,05	20	0,91	17	0	10	0,94	17
SB_54	0,05	20	0,46	14	0,65	17	0,44	17	0,75	20	0,65	20	0,35	20	0,58	19	0,08	13	0,65	10	0,86	18
SB_55	0,14	18	0	20	0	20	0,71	17	0,05	20	0,03	17	0	20	0,95	20	0,13	16	0	10	0	19
SB_56	0,2	20	0,06	17	0,18	20	0,03	19	0,88	17	0,75	18	0,05	20	0,03	20	0,55	19	0,05	10	0,11	18
SB_57	0,1	20	0,79	14	0,97	18	0,07	14	0,8	20	0,9	20	0,9	20	0	20	0,9	20	0,85	10	0,94	17
SB_58	0,1	20	0,82	14	0,76	19	0,1	10	0,65	20	0,25	2	0,9	20	0	20	0,45	20	0,8	10	0,89	19
SB_59	0,1	20	0,05	20	0,05	20	0	20	0	20	0	20	0	20	0,21	19	0,03	20	0,85	10	0,94	17
SB_60	0,76	19	0	20	0	20	0	20	0,15	20	0,95	20	0,86	18	0,63	20	0,9	20	0	10	0,87	19
SB_61	0,1	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	20	0	10	0	19
SB_63	0,05	20	0,97	19	0	20	0,88	20	0,86	18	0,81	18	0,1	20	0	20	0,93	20	0	10	0	19
SB_64	0,83	18	0,82	14	0,45	19	0,97	19	0,88	17	0,8	5	0,73	11	0,73	20	0,96	14	0,25	10	0,16	19
SB_65	0,08	18	0,03	20	0	20	0,9	20	0,08	18	0,08	19	0,05	20	0,15	20	0,2	20	0,05	10	0,08	18
SB_66	0,05	20	0	20	0	20	0	20	0	20	0,93	15	0,05	20	0	20	0,9	20	0	10	0	19
SB_69	0,89	19	0,93	20	0,95	20	0,8	20	0,86	18	0,64	14	0,39	14	0,71	19	0,95	20	0,7	10	0,89	18
SB_70	0,05	20	0,05	20	0	20	0	20	0	20	0	20	0,05	20	0,05	20	0	20	0,7	10	0,42	18
SB_71	0,05	20	0	20	0	20	0,05	20	0,4	20	0,05	20	0	20	0	20	0,05	20	0	10	0,88	17
SB_72	0,11	19	0	20	0	20	0,58	20	0,08	18	0,06	18	0	20	0,03	20	0,13	19	0	10	0,89	19
SB_73	0,05	20	0,05	20	0,05	20	0	20	0,05	20	0	20	0	20	0,05	20	0	20	0,1	10	0,19	18
SB_74	0,05	20	0,05	20	0	20	0,83	20	0,98	20	0	20	0	20	0	20	0	20	0	10	0,18	19
SB_75	0,1	20	0,25	20	0	20	0	20	0,2	20	0,45	20	0,3	20	0,05	20	0,1	20	0	10	0,83	18
SB_76	0,32	19	0,1	20	0	20	0,25	20	0,85	13	0,93	15	0	20	0	20	0,15	20	0	10	0	19
SB_77	0,05	20	0,05	20	0	20	0	20	0	20	0,05	20	0	20	0	20	0	20	0	10	0	19
SB_79	0,75	12	0,05	20	0,4	20	0,95	20	0,86	18	0,32	14	0	20	0,1	20	0,1	20	0,25	6	0,97	18
SB_80	0,05	20	0,9	20	0,6	20	0,05	20	0,24	19	0,05	19	0	20	0	20	0	20	0,65	10	0,03	19

SNPID	P35_F	P35_I	P32_F	P32_I	P33_F	P33_I	P34_F	P34_I
Fst_03	0	16	0,1	10	0	16	0	12
Fst_04	0,57	14	0,65	10	0,81	16	0,82	11
Fst_05	0,97	16	0,2	10	0,06	16	0,96	12
Fst_07	0	16	0,95	10	0	16	0,96	12
Fst_08	0	16	0	10	0,7	15	0,96	12
Fst_09	0	16	0	10	0,06	16	0	12
Fst_12	0,09	11	0,28	9	0,06	16	0,08	12
Fst_13	0,03	16	0,2	10	0,06	16	0,05	10
Fst_14	0,19	16	0,9	10	0,31	16	0,32	11
Fst_16	0,06	16	0	10	0,06	16	0,25	12
Fst_18	0,97	16	0,2	10	0,72	16	0,45	11
Fst_19	0,94	16	0	10	0,93	15	0,96	12
Fst_20	0,06	16	0,15	10	0,18	14	0,08	12
Fst_21	0,07	15	0,1	10	0	16	0,17	12
Fst_22	0,71	7	0,05	10	0,32	11	0,28	9
Fst_23	0	16	0	10	0	16	0,08	12
Fst_24	0,67	15	0	10	0,35	13	0,42	12
Fst_25	0,03	16	0,35	10	0,03	16	0,17	12
Fst_27	0	16	0	10	0	16	0	12
Fst_28	0,8	15	0,6	10	0,9	15	0,96	12
Fst_29	0,25	16	0,35	10	0,13	16	0,13	12
Fst_30	0	16	0	10	0	16	0	12
Fst_31	0,75	16	0,89	9	0,87	15	0,82	11
Fst_32	0	16	0	10	0,72	16	0,92	12
Fst_34	0	16	0,15	10	0	16	0,08	12
Fst_35	0,82	14	0,75	8	0,82	14	0,77	11
Fst_36	0,2	15	0	10	0,23	15	0,29	12
Fst_37	0,97	15	0,2	10	0,97	16	0,08	12
Fst_38	0	16	0	10	0	16	0	12
Fst_40	0,13	16	0,25	10	0,07	15	0,17	12
Fst_41	0,5	10	0	10	0,72	16	0,92	12
Fst_45	0,87	15	0	10	0	16	0	12
Fst_46	0,19	16	0,3	10	0,13	15	0,17	12
Fst_47	0,57	15	0,3	10	0,91	16	0,71	12
Fst_49	0,06	16	0	10	0	16	0	12
Fst_50	0,03	16	0	10	0,03	15	0,21	12
Fst_51	0,37	15	0	10	0	16	0	12
Fst_53	0	16	0	10	0,83	15	0,96	12
Fst_54	0,91	16	0	10	0	16	0,08	12
Fst_55	0,25	16	0	10	0	16	0,05	11
Fst_56	0,75	14	0,17	9	0,69	16	0,5	12



SNPID	P35_F	P35_I	P32_F	P32_I	P33_F	P33_I	P34_F	P34_I
Fst_57	0	16	0,85	10	0,93	15	0,96	12
Fst_58	0,44	16	0,2	10	0	16	0,04	12
Fst_59	0,4	15	0,35	10	0,88	16	0,08	12
Fst_60	0,59	16	0,65	10	0,03	16	0,08	12
Fst_61	0	16	0,63	8	0	16	0,08	12
Fst_62	0	16	0,4	10	0	16	0	12
Fst_66	0,63	15	0,95	10	0,27	15	0,25	12
Fst_68	0,44	16	0	10	0,81	16	0,92	12
Fst_69	0,37	15	0,25	10	0,22	16	0,05	10
Fst_70	0,8	10	0,6	10	0,25	16	0,96	12
Fst_71	0	16	0	10	0	16	0	12
Fst_73	0	16	0	10	0	16	0	12
Fst_74	0,03	16	0	10	0,13	16	0,25	12
Fst_75	0,63	16	0,3	10	0,22	16	0,35	10
Fst_76	0,69	16	0,4	10	0,19	16	0	12
Fst_77	0,06	16	0	10	0	16	0	12
Fst_78	0,94	16	0,75	10	0,97	16	0	12
Fst_80	0,43	15	0,65	10	0,65	13	0,54	12
Fst_82	0,5	16	0,4	10	0,88	16	0,96	12
Fst_83	0,91	16	0,2	10	0,93	15	0,82	11
Fst_84	0,8	15	0	10	0,83	15	0,63	12
Fst_85	0,13	16	0	10	0,16	16	0,21	12
Fst_87	0,88	16	0,2	10	0,84	16	0,79	12
Fst_88	0,06	16	0,15	10	0,03	16	0	12
Fst_89	0,13	16	0	10	0,17	12	0,18	11
Fst_90	0,88	16	0,4	10	0,97	15	0,92	12
Fst_91	0,93	15	0	10	0,92	13	0,79	12
Fst_92	0,97	16	0,7	10	0,94	16	0,96	12
Fst_94	0,88	16	0	10	0,82	14	0,79	12
Fst_95	0	16	0	10	0	16	0	12
Fst_96	0	16	0	10	0,7	15	0	12
Fst_97	0,08	13	0,22	9	0,1	15	0,2	10
MS_01	0,11	9	0	10	0,07	14	0,17	12
MS_02	0,06	16	0	10	0,06	16	0	12
MS_03	0,06	16	0	10	0	16	0,08	12
MS_09	0,09	16	0,38	8	0,13	16	0,17	12
MS_11	0,22	16	0	10	0	16	0	12
MS_13	0,06	16	0	10	0	16	0	12
MS_14	0,09	16	0,3	10	0	16	0	12
MS_15	0,91	16	0,7	10	0,69	16	0,95	11
MS_16	0,13	16	0,05	10	0,07	15	0,04	12

SNPID	P35_F	P35_I	P32_F	P32_I	P33_F	P33_I	P34_F	P34_I
MS_17	0	16	0,05	10	0,13	16	0	12
MS_18	0	16	0,05	10	0	16	0	12
MS_19	0	16	0	10	0,19	16	0	12
MS_20	0,06	16	0,95	10	0,25	16	0	12
MS_21	0	16	0,05	10	0	16	0	12
MS_22	0,19	16	0,05	10	0	16	0,08	12
MS_23	0	16	0,95	10	0,06	16	0,96	12
MS_24	0	16	0,05	10	0	16	0	12
MS_25	0,06	16	0	10	0	16	0	12
MS_26	0	16	0	10	0	16	0	12
MS_27	0,06	16	0	10	0	16	0	12
MS_28	0	16	0	10	0	16	0	12
MS_29	0,06	16	0	10	0	16	0	12
MS_30	0	16	0	10	0	16	0	12
MS_31	0,25	16	0	10	0,06	16	0,08	12
MS_32	0,06	16	0,05	10	0,03	15	0,04	12
MS_34	0,07	15	0	10	0,03	15	0,04	12
MS_35	0,09	16	0	10	0	16	0	12
MS_36	0,09	16	0,4	10	0,06	16	0	12
MS_37	0,27	15	0,85	10	0,31	16	0,5	11
MS_39	0,06	16	0	10	0,06	16	0,08	12
MS_40	0,03	16	0,56	9	0	16	0	12
MS_44	0	16	0	10	0	16	0	12
MS_45	0,44	16	0,1	10	0,41	16	0,21	12
MS_47	0,28	16	0,05	10	0,2	15	0,21	12
MS_48	0,25	16	0,45	10	0,28	16	0,13	12
MS_49	0,91	16	0,95	10	0,97	15	0,96	12
MS_50	0,97	16	0,95	10	0	16	0,96	12
MS_51	0,91	16	0,95	10	0	16	0,96	12
MS_55	0,94	16	0,95	10	0,97	15	0	12
MS_56	0,07	15	0,05	10	0,07	14	0	12
MS_57	0,09	16	0	10	0	16	0,08	12
MS_58	0,1	15	0,05	10	0,03	15	0,08	12
MS_59	0,91	16	0,95	10	0,97	15	0,95	11
MS_60	0,88	13	0,95	10	0,96	14	0,96	12
MS_61	0,96	14	0,95	10	0,94	16	0,96	12
MS_62	0,94	16	0,95	10	0,93	14	0	12
MS_63	0,13	16	0,05	10	0,06	16	0,04	12
MS_64	0,91	16	0	10	0,94	16	0,96	12
MS_66	0,88	16	0,95	10	0,88	16	0,68	11
MS_67	0,93	15	0,95	10	0,19	16	0,96	12

SNPID	P35_F	P35_I	P32_F	P32_I	P33_F	P33_I	P34_F	P34_I
MS_68	0,91	16	0,95	10	0,94	16	0	12
MS_69	0,79	12	0,06	9	0,5	10	0,55	11
MS_70	0,25	16	0,05	10	0,06	16	0,04	12
MS_74	0,88	13	0,95	10	0,97	15	0,96	12
MS_75	0,22	16	0,05	10	0,06	16	0,04	12
MS_76	0,8	15	0,95	10	0,97	15	0,96	12
MS_77	0,22	16	0,06	9	0,03	15	0,04	12
MS_78	0,85	13	0,95	10	0,19	16	0,96	12
MS_81	0,44	16	0,5	10	0,06	16	0,08	12
MS_82	0,91	16	0,5	10	0	16	0,96	12
MS_83	0,07	15	0,5	9	0	16	0	12
MS_87	0,97	16	0	10	0	16	0	12
MS_88	0	16	0,5	10	0	16	0,08	12
MS_89	0,13	16	0	10	0	16	0	12
SB_01	0	16	0	10	0	16	0	12
SB_02	0	16	0	10	0	16	0	12
SB_03	0,88	16	0	10	0,5	16	0,59	11
SB_04	0,88	12	0	10	0	16	0	12
SB_07	0,06	16	0	10	0,06	16	0	12
SB_08	0,91	16	0,3	10	0,06	16	0,92	12
SB_09	0,13	16	0,15	10	0	16	0,17	12
SB_10	0,88	16	0	10	0,67	15	0,67	12
SB_12	0	16	0	10	0	16	0	12
SB_13	0	16	0	10	0,03	16	0	12
SB_14	0	16	0	10	0	16	0	12
SB_15	0,56	16	0	10	0,88	16	0,82	11
SB_16	0,07	15	0	10	0	16	0	12
SB_20	0	16	0	10	0	16	0	12
SB_22	0,81	16	0,7	10	0,69	16	0,92	12
SB_25	0,33	15	0	10	0,38	16	0,32	11
SB_26	0	16	0,8	10	0	16	0	12
SB_27	0,47	16	0,75	10	0,5	15	0,83	12
SB_28	0,13	16	0	10	0,69	16	0,11	9
SB_29	0,03	16	0,6	10	0	16	0,08	12
SB_30	0,1	15	0,25	8	0,03	15	0,04	12
SB_32	0,97	16	0	10	0	16	0	12
SB_33	0	16	0	10	0	16	0	12
SB_34	0	16	0	10	0	16	0	12
SB_35	0	16	0	10	0,96	13	0	12
SB_38	0,09	16	0	10	0,19	16	0	12
SB_39	0,06	16	0	10	0	16	0	12

SNPID	P35_F	P35_I	P32_F	P32_I	P33_F	P33_I	P34_F	P34_I
SB_41	0	16	0	10	0	16	0	12
SB_43	0,06	16	0	10	0,23	15	0,21	12
SB_44	0,06	16	0	10	0,23	13	0,25	12
SB_45	0,06	16	0,1	10	0	16	0	12
SB_46	0	16	0	10	0	16	0	12
SB_47	0,31	13	0,85	10	0,3	15	0,5	11
SB_48	0,03	16	0,3	10	0,06	16	0,38	12
SB_49	0,83	15	0,55	10	0,97	15	0,08	12
SB_50	0	16	0	10	0	16	0	12
SB_51	0	16	0	10	0	16	0	12
SB_52	0	16	0	10	0,97	15	0	12
SB_54	0,78	16	0	10	0	16	0,83	12
SB_55	0	16	0,8	10	0,06	16	0	12
SB_56	0,13	16	0,05	10	0,06	16	0,05	11
SB_57	0,38	16	0,5	10	0,95	11	0,08	12
SB_58	0,56	16	0,5	10	0	16	0,95	10
SB_59	0,5	16	0	10	0,94	16	0,92	12
SB_60	0	16	0	10	0,06	16	0,96	12
SB_61	0	16	0	10	0	16	0	12
SB_63	0	16	0,55	10	0	16	0	12
SB_64	0,41	16	0	10	0,06	16	0,33	12
SB_65	0,03	16	0	10	0,03	16	0,09	11
SB_66	0	16	0	10	0	16	0,08	12
SB_69	0,6	15	0	10	0,88	16	0,88	12
SB_70	0,31	16	0	10	0,28	16	0,38	12
SB_71	0,94	16	0	10	0,84	16	0,79	12
SB_72	0,94	16	0	10	0,91	16	0,17	12
SB_73	0,1	15	0	10	0,11	14	0,18	11
SB_74	0,13	16	0	10	0,16	16	0,21	12
SB_75	0,88	16	0	10	0,38	16	0,85	10
SB_76	0	16	0,1	10	0	16	0	12
SB_77	0,06	16	0	10	0,06	16	0	12
SB_79	0,93	15	0,65	10	0	16	0	12
SB_80	0,09	16	0,22	9	0	16	0	12

## ANNEXE IV.1 : PLANTES-MERES

Plantes-mères utilisées dans le dispositif de génétique d'association

Plante-mère	Année 2013	Année 2014	
Pop 1	Pop1_1	Pop1_1	
	Pop1_10	Pop1_10	
	Pop1_11	Pop1_11	
	Pop1_12	Pop1_12	
	Pop1_13	Pop1_13	
	Pop1_14	Pop1_14	
	Pop1_16	Pop1_16	
	Pop1_18	Pop1_18	
	Pop1_19	Pop1_19	
	Pop1_23	Pop1_23	
	Pop1_24	Pop1_24	
	Pop1_25	Pop1_25	
	Pop1_26	Pop1_26	
	Pop1_28	Pop1_28	
	Pop1_29	Pop1_29	
	Pop 4	Pop4_1	Pop4_1
			Pop4_2
		Pop4_3	
		Pop4_4	
Pop4_6		Pop4_6	
		Pop4_7	
Pop4_8		Pop4_8	
Pop4_10		Pop4_10	
Pop4_11		Pop4_11	
Pop4_12		Pop4_12	
		Pop4_18	
		Pop4_19	
Pop4_21		Pop4_21	
Pop4_23		Pop4_23	
Pop4_25		Pop4_25	
		Pop4_26	
Pop4_27		Pop4_27	
Pop4_28	Pop4_28		
Pop4_29	Pop4_29		
Pop4_31	Pop4_31		
Pop4_33	Pop4_33		
Pop4_34	Pop4_34		
Pop 5	Pop5_1	Pop5_1	
	Pop5_4	Pop5_4	
	Pop5_5	Pop5_5	
	Pop5_8	Pop5_8	
	Pop5_12	Pop5_12	
	Pop5_13	Pop5_13	

	Pop5_14	Pop5_14
	Pop5_15	Pop5_15
	Pop5_17	Pop5_17
	Pop5_19	Pop5_19
	Pop5_20	Pop5_20
	Pop5_22	Pop5_22
	Pop5_23	Pop5_23
	Pop5_26	Pop5_26
	Pop5_32	Pop5_32
Pop 6	Pop6_1	Pop6_1
	Pop6_3	Pop6_3
	Pop6_7	Pop6_7
	Pop6_8	Pop6_8
	Pop6_9	Pop6_9
	Pop6_11	Pop6_11
	Pop6_13	Pop6_13
	Pop6_14	Pop6_14
	Pop6_16	Pop6_16
	Pop6_17	Pop6_17
	Pop6_19	Pop6_19
	Pop6_21	Pop6_21
	Pop6_25	Pop6_25
	Pop6_26	Pop6_26
	Pop6_31	Pop6_31
Pop 8	Pop8_5	Pop8_5
	Pop8_9	Pop8_9
	Pop8_11	Pop8_11
	Pop8_13	Pop8_13
	Pop8_14	Pop8_14
	Pop8_15	Pop8_15
	Pop8_16	Pop8_16
	Pop8_17	Pop8_17
	Pop8_25	Pop8_25
	Pop8_26	Pop8_26
	Pop8_27	Pop8_27
	Pop8_28	Pop8_28
	Pop8_31	Pop8_31
	Pop8_36	Pop8_36
	Pop8_40	Pop8_40
Pop 10	Pop10_1	Pop10_1
	Pop10_2	Pop10_2
	Pop10_3	Pop10_3
	Pop10_4	Pop10_4
	Pop10_5	Pop10_5
	Pop10_6	Pop10_6
	Pop10_7	Pop10_7
	Pop10_8	Pop10_8

	Pop10_11	Pop10_11
	Pop10_12	Pop10_12
	Pop10_18	Pop10_18
	Pop10_19	Pop10_19
	Pop10_22	Pop10_22
	Pop10_26	Pop10_26
	Pop10_28	Pop10_28
Pop 13	Pop13_4	Pop13_4
	Pop13_5	Pop13_5
	Pop13_7	Pop13_7
	Pop13_9	Pop13_9
	Pop13_12	Pop13_12
	Pop13_13	Pop13_13
	Pop13_15	
	Pop13_16	Pop13_16
	Pop13_17	Pop13_17
	Pop13_18	Pop13_18
	Pop13_25	Pop13_25
	Pop13_27	
	Pop13_30	
	Pop13_31	Pop13_31
		Pop13_33
		Pop13_34
	Pop13_35	Pop13_35
Pop 17	Pop17_4	Pop17_4
	Pop17_5	Pop17_5
	Pop17_6	Pop17_6
	Pop17_7	Pop17_7
	Pop17_9	Pop17_9
	Pop17_10	Pop17_10
	Pop17_11	Pop17_11
	Pop17_16	Pop17_16
	Pop17_20	Pop17_20
	Pop17_29	Pop17_29
	Pop17_30	Pop17_30
	Pop17_38	Pop17_38
	Pop17_40	Pop17_40
	Pop17_42	Pop17_42
	Pop17_43	Pop17_43
Pop 22	Pop22_1	Pop22_1
	Pop22_2	Pop22_2
	Pop22_3	Pop22_3
	Pop22_4	Pop22_4
	Pop22_5	Pop22_5
	Pop22_6	Pop22_6
	Pop22_7	Pop22_7
	Pop22_8	Pop22_8

	Pop22_9	Pop22_9
	Pop22_10	Pop22_10
	Pop22_11	Pop22_11
	Pop22_12	Pop22_12
	Pop22_13	Pop22_13
	Pop22_14	Pop22_14
	Pop22_15	Pop22_15
Pop 28	Pop28_2	Pop28_2
	Pop28_6	Pop28_6
	Pop28_9	Pop28_9
	Pop28_10	Pop28_10
	Pop28_11	Pop28_11
	Pop28_12	Pop28_12
	Pop28_13	Pop28_13
	Pop28_14	Pop28_14
	Pop28_21	Pop28_21
	Pop28_22	Pop28_22
	Pop28_25	Pop28_25
	Pop28_28	Pop28_28
	Pop28_37	Pop28_37
	Pop28_39	Pop28_39
	Pop28_43	Pop28_43
Pop 29	Pop29_1	Pop29_1
		Pop29_8
	Pop29_9	Pop29_9
	Pop29_11	Pop29_11
	Pop29_13	Pop29_13
	Pop29_16	Pop29_16
	Pop29_18	Pop29_18
	Pop29_19	
	Pop29_21	
		Pop29_22
	Pop29_29	Pop29_29
		Pop29_30
	Pop29_31	Pop29_31
	Pop29_35	Pop29_35
	Pop29_36	Pop29_36
	Pop29_38	Pop29_38
	Pop29_42	
	Pop29_43	Pop29_43

---







**Titre :** Adaptation locale des téosintes *Zea mays* ssp. *parviglumis* et ssp. *mexicana* le long de gradients altitudinaux

**Mots clés :** Adaptation locale, Populations naturelles, Séquençage en mélange, Différentiation allélique, Scan génomique, Génétique d'association

**Résumé :** Les téosintes *Zea mays* ssp. *parviglumis* et ssp. *mexicana* sont les apparentés sauvages les plus proches du maïs cultivé. Elles occupent au Mexique des niches écologiques distinctes bien délimitées par l'altitude. *Zea mays* ssp. *parviglumis* pousse dans des environnements chauds et humides situés à moins de 1800m d'altitude, tandis que *Zea mays* ssp. *mexicana* pousse dans des environnements plus secs et froids (au dessus de 1800m). Nous nous sommes intéressés aux bases génétiques de l'adaptation locale des téosintes à l'altitude. Dans ce but, 37 populations ont été échantillonnées le long de deux gradients altitudinaux, utilisés comme réplicats biologiques. Deux populations extrêmes de chaque gradient ainsi que deux populations de moyenne altitude de chaque sous-espèce du gradient 1 ont été séquencées à faible profondeur afin de découvrir des polymorphismes dont les fréquences alléliques sont contrastées entre les extrêmes tout en contrôlant pour la différenciation entre les sous-espèces. A partir de méthodes de détection de la sélection intra- et inter-populationnelle, des locus candidats ont été détectés.

Une revue bibliographique a permis d'établir des correspondances entre nos candidats et des régions précédemment décrites pour leurs effets phénotypiques sur des caractères adaptatifs. Nous montrons ainsi un rôle important des interactions sol-plante et de la pilosité des feuilles dans l'adaptation à l'altitude. Pour valider certains de nos candidats, 270 polymorphismes ont été génotypés sur les 37 populations afin de réaliser des études de clines de fréquence. Parallèlement, nous avons mis en œuvre un dispositif de caractérisation phénotypique (2 lieux x 2 années) afin de tester l'association entre ces polymorphismes et 18 caractères mesurés en champ. Nous discutons des apports méthodologiques de notre étude aussi bien du point de vue des technologies haut débit que de la détection de la sélection. Notre dispositif devra être pleinement exploité pour valider nos candidats. Les perspectives incluent la poursuite de l'étude sur d'autres types de polymorphismes ainsi qu'un suivi temporel des populations.

**Title :** Local adaptation of teosintes subspecies *Zea mays* sp *parviglumis* and *Zea mays* sp *mexicana* along altitudinal gradients

**Keywords :** Local adaptation, Natural populations, Pool-sequencing, Allelic differentiation, Genome scan, Association mapping

**Abstract:** Teosintes *Zea mays* ssp *parviglumis* (*parviglumis*) and *Zea mays* ssp *mexicana* (*mexicana*) are the closest wild relatives of cultivated maize. They occupy distinct ecological niches in Mexico, well separated by altitude. *Zea mays* ssp. *parviglumis* encountered below 1800m grows in wet and warm conditions, while *mexicana* grows in drier and colder climates (above 1800m). In order to investigate the genetic bases of local adaptation to altitude, we sampled 37 populations along two altitudinal gradients. We sequenced the two most extreme populations of each gradient as well as two intermediate populations - one from each subspecies along gradient 1. We searched for polymorphisms with contrasted allele frequencies between the extremes while accounting for subspecies differentiation. Using both inter- and intra- population methods we identified several candidate loci.

with regions previously described as involved in phenotypic variation of adaptive traits. Our results highlight the role of plant-soil interactions and leaves hairiness in the adaptation to altitude. To validate further a subset of 270 polymorphisms chosen among our best candidates, we genotyped them on the 37 populations with the aim of performing clinal analyzes of allele frequencies. In parallel, we undertook phenotypic evaluation trials (2 locations x 2 years) to test the association of these polymorphisms with 18 traits measured in the field. We discuss the methodological contributions of our study both from the standpoint of high throughput technologies and detection of selective footprints. Our setting will be fully exploited to validate our candidates. Perspectives include the discovery and assessment of the contribution of other types of polymorphisms and the temporal follow-up of populations.

Based on a literature review, we confronted them