Diversité de l’abricotier sauvage à l’échelle phénotypique de deux loci et du génome complet
Marie Lepert

To cite this version:

HAL Id: dumas-01093362
https://dumas.ccsd.cnrs.fr/dumas-01093362
Submitted on 10 Dec 2014

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.
Diversité de l’abricotier sauvage à l’échelle phénotypique, de deux loci et du génome complet

Par : Marie LEPERT

Soutenu à Rennes le 11/09/14

Devant le jury composé de :
Président : Maria Manzanares-Dauleux
Maître de stage : Véronique Decroocq
Enseignant référent : Maria Manzanares-Dauleux
Autres membres du jury (Nom, Qualité) : Laurence Hibrand Saint Oyand, Rapporteur
Anne Laperche, Examinateur

Les analyses et les conclusions de ce travail d’étudiant n’engagent que la responsabilité de son auteur et non celle d’AGROCAMPUS OUEST
REMERCIEMENTS

Je tiens ici à remercier les personnes sans qui ce travail n’aurait pu être réalisé : Véronique Decrooq, pour m’avoir accueilli au sein de son laboratoire et m’avoir donné l’occasion de travailler sur ce sujet ; Stéphanie Mariette pour les discussions préliminaires et le professeur Albert Abbott, coordinateur du projet ANR ABRIWG, pour avoir permis le séquençage des individus sauvages. Je remercie également Véronique Decrooq, et David Tricon pour m’avoir transmis leur savoir sur le sujet, Stéphane Decrooq et Aurélie Chague pour avoir généré les données de génotypage par microsatellites et m’avoir aidée dans l’analyse des données et le travail de laboratoire, et Fabienne Wong pour la mise en place du pipeline d’analyse des données génomiques et m’avoir aidée au nettoyage des données de séquençage. Et enfin, l’ensemble du personnel, doctorants et stagiaires du laboratoire pour m’avoir occasionnellement apporté leur aide et leurs conseils.
SOMMAIRE
INTRODUCTION .......................................................................................................................... 1
Les enjeux de la production fruitière ......................................................................................... 1
L’abricotier ............................................................................................................................... 1
Importance économique ........................................................................................................ 1
Histoire ................................................................................................................................ 1
Biologie .................................................................................................................................. 2
Génétique ................................................................................................................................ 2
  Auto-incompatibilité .............................................................................................................. 2
  Résistance à la sharka .......................................................................................................... 3
    La maladie de la sharka ....................................................................................................... 3
    Dissémination de la maladie .............................................................................................. 3
    Méthodes de lutte .............................................................................................................. 3
    Mécanismes de résistance ................................................................................................. 4
  Adaptation au réchauffement climatique ............................................................................ 4
Hypothèses et objectifs de cette étude ...................................................................................... 5
MATÉRIEL et MÉTHODES ...................................................................................................... 6
Matériel végétal ........................................................................................................................ 6
Phénotypage ............................................................................................................................. 6
Analyse du marqueur d’auto-compatibilité ............................................................................. 7
Analyse du marqueur de résistance à la sharka .................................................................. 7
Analyse de la structure des populations avec les marqueurs SSR ..................................... 7
Extraction de l’ADN et séquençage ....................................................................................... 8
Nettoyage des données de séquençage issues de 77 accès d’abricotiers sauvages ........... 8
Analyse des données de séquençage ..................................................................................... 8
RESULTATS .............................................................................................................................. 9
  Analyse de la diversité phénotypique du caractère « débourrement » chez les abricotiers
d’Asie centrale et du Caucase .............................................................................................. 9
    Analyse en composantes principales (ACP) ................................................................. 9
    Analyse de variance ....................................................................................................... 9
  Structuration des abricotiers d’Asie centrale et du Caucase selon la variabilité phénotypique ...
    Analyse de la structuration chez les individus sauvages et les variétés anciennes .......... 10
    Analyse de la structuration au sein des individus sauvages uniquement ..................... 10
    Analyse sans individus apparentés .................................................................................. 10
  Synthèse des résultats obtenus lors de l’analyse de la diversité phénotypique chez les
  abricotiers provenant d’Asie Centrale et du Caucase ....................................................... 11
Diversité génétique aux loci PPVres et SI ............................................................................. 11
  Locus majeur contrôlant la résistance à la sharka .............................................................. 11
  Auto-Incompatibilité .......................................................................................................... 12
Diversité génétique globale à partir des données de génotypage issues de 21 loci
  microsatellites ...................................................................................................................... 12
  Distance génétique et flux de gènes entre populations .................................................... 13
  Différenciation des populations ........................................................................................ 14
  Phylogénie .......................................................................................................................... 14
  Synthèse des résultats obtenus lors de l’analyse de la diversité génétique au niveau
  de 21 loci microsatellites ................................................................................................. 14
Diversité génétique globale à partir des données de SNPs (Single Nucleotide
  Polymorphism) répartis sur l’ensemble du génome ............................................................. 15
Estimation de la diversité génétique des abricotiers sauvages d’Asie centrale .................15
Structuration des populations naturelles d’abricotiers sur la base de 2 076 SNPs .............15
Distance génétique et flux de gènes entre populations ...............................................15
Différenciation des populations .................................................................................15
Phylogénie des accessions séquencées à partir de 2076 SNPs ....................................16
Synthèse des résultats obtenus lors de l’analyse de la diversité génétique au niveau des SNPs ..................................................................................................................16

DISCUSSION .....................................................................................................................16
Des variétés anciennes avec de faibles besoins en froid ..................................................16
Le débourrement, une variable échouant à révéler une structuration au sein des abricotiers sauvages ................................................................................................................................17
Un potentiel des populations sauvages pour lutter contre la sharka ..............................18
Une mutation « récente » à l’origine de l’auto-compatibilité ? .......................................18
Une proximité des abricotiers issus du Caucase avec le matériel cultivé et caucasien révélée par les marqueurs SSR ..........................................................19
Des flux de gènes entre compartiments sauvage et cultivé .......................................19
Des populations sauvages isolées par distance ............................................................20
Des populations sauvages peu structurées ..................................................................20
Une faible structuration des populations observée avec les marqueurs SNP ............21
Une diversité importante révélée par les SNP ..........................................................21
Une séparation précoce du compartiment cultivé .........................................................22
Des interrogations qui demeurent.................................................................................22

CONCLUSION ET PERSPECTIVES .................................................................................24
BIBLIOGRAPHIE ............................................................................................................24
LISTE DES ABRÉVIATIONS/GLOSSAIRE

**ABRIWG** = ABRicot Whole Genome : Projet visant à étudier la résistance à la sharka au niveau pan-génomique sur les abricotiers cultivés

**ACP** = Analyse en Composantes Principales

**ADN** = Acide Desoxyribonucléique

**AFLP** = Amplification Fragment Length Polymorphism

**ANOVA** = ANalysis Of VAriance

**ANR** = Agence Nationale de la Recherche

**Apr** = richesse allélique privée

**ARN/RNA** = Acide Ribonucléique

**ARS-USDA** = Agricultural Research Service-United States Department of Agriculture

**ELISA** = Enzyme-Linked ImmunoSorbent Assay

**EPPO** = European and Mediterranean Plant Protection Organization

**F1** = première génération d’autofécondation

**FIS, FST, FIT** = Indicateurs statistiques de Wright

**GPS** = Global Positioning System

**He** = hétérozygotie attendue

**HF** = somme d’heures de froid (température inférieure à 7°C) nécessaires pour atteindre le stade 3

**HJ** = nombre d’heures de jour depuis l’équinoxe d’hiver pour atteindre le stade 3

**Ho** = hétérozygotie observée

**IBD** = Isolation By distance

**INRA** = Institut National de la Recherche Agronomique

**locus S** : gène participant à l’auto-incompatibilité

**MAF** = minor Allele Frequency

**MANOVA** = Multiple ANalysis Of VAriance

**MARS** = Molecular Assisted Resistance to Sharka : Projet similaire à ABRIWG

**MATH** = meprin and TRAF homology

**Mb** = Megabases

**Na** = nombre d’allèles

**Nm** = nombre de migrants : estimateur du flux de gènes

**pb** = paires de bases

**PCR** = Polymerase Chain Reaction

**PGS1.21** = marqueur lié à la résistance à la sharka

**PPV** = Plum Pox virus : virus provoquant la maladie de la sharka

**QTL** = Quantitative Trait Loci

**RG** = quantité de rayonnement solaire pour atteindre le stade 3 depuis le premier septembre

**SAM** = Sélection Assistée par Marqueur

**SEO** = Stark Early Orange : nom d’une variété d’abricot

**SFB** = S-haplotype specific F-Box : gène participant à l’auto-incompatibilité

**SI** = Self Incompatibility

**SNP** = Single Nucleotid Polymorphism

**S-RNase** = Ribonucléase codée par le locus S

**SSLP** = Simple Sequence Length Polymorphism

**SSR** = single Sequence Repeat

**STONE** = projet européen visant à étudier la diversité des arbres fruitiers à noyaux.

**TRAF** = Tumor Necrosis Factor-receptor associated factor

**UPGMA** = Unweighted Pair Group Method with Arithmetic Mean

**ZP002** = marqueur lié à la résistance à la sharka
LISTE DES ANNEXES
Annexe I : Distance entre les lieux de prélèvement (km)
Annexe II : Liste des accessions utilisées dans chaque partie
Annexe III : Amorces ciblant l’insertion de 358 pb dans le gène SFB et conditions PCR correspondantes.
Annexe IV : Protocole pour l’extraction d’ADN
Annexe V : Sorties graphiques de l’ACP sur les deux premiers axes à partir des données HF et RG
Annexe VI : Coordonnées du barycentre de chaque population et ses caractéristiques
Annexe VII : Résultats des ANOVAs sur le nombre d’heures de froid (HF), la quantité de rayonnement solaire (RG) et le nombre d’heures de jour (HJ) nécessaires pour atteindre le stade 3 du débourrement en fonction des critères ‘type d'accession’, ‘population d'origine’ et ‘pays d'origine’
Annexe VIII: Résultats de la MANOVA sur les notes attribuées tout au long du débourrement en fonction des effets du type d'accession, de la population et du pays d'origine
Annexe IX : Résultats de la MANOVA sur les notes attribuées tout au long du débourrement en fonction des effets du nombre d’heures de froid (HF), de la quantité de rayonnement solaire (RG) et du nombre d’heures de jour (HJ) nécessaires pour atteindre le stade 3 du débourrement
Annexe X : Résultats de la MANOVA sur le nombre d’heures de froid la quantité de rayonnement solaire et le nombre d’heures de jour nécessaires pour atteindre le stade 3 du débourrement en fonction des effets type d'accession, population d'origine et pays d'origine.
Annexe XI Structuration détectée par GENELAND sur le phénotype (HF=nombre d’heures de froid pour atteindre le stade 3) des individus sauvages
Annexe XII : Comparaison des Fis obtenus avec les marqueurs SSR et SNP
Annexe XIII : Valeur de Fis pour chaque SNP le long du génome.
LISTE DES ILLUSTRATIONS
Figure 1: Organisation génétique du locus S
Figure 2 : Fonctionnement du mécanisme de l'auto-incompatibilité gamétophytique
Figure 3 : Symptômes sur fruits
Figure 4 : Symptômes sur feuilles
Figure 5 : Position des différents marqueurs liés à PPVres
Figure 6: Lieux de prélèvement des échantillons.
Figure 7 : Parties des arbres récoltées et leur utilisation.
Figure 8 : Échelle de notation utilisée pour caractériser les stades de débourrement.
Figure 9: Sorties graphiques de l'ACP sur les deux premières dimensions. L’ensemble des notes de la période d’observation d’un notateur sont utilisées et l’individu 21 a été supprimé.
Figure 10 : Structuration détectée par GENELAND (fréquences corrélées) sur le phénotype (HF=nombre d’heures de froid pour atteindre le stade 3). Les individus apparentés ont été supprimés. Structuration en trois groupes des individus phénotypés.
Figure 11: Structuration détectée par GENELAND (fréquences non corrélées) sur le phénotype (HF=nombre d’heures de froid pour atteindre le stade 3). Les individus apparentés ont été conservés.
Figure 12: Exemple de gel de polyacrylamide obtenu avec le marqueur ZP002.
Figure 13: Structuration des populations d’abricotier détectée par STRUCTURE à partir des données des SSR.
Figure 14: Arbre phylogénétique construit selon la méthode UPGMA avec le logiciel Poptree à partir des données de SSR.
Figure 15: Traitement des données de séquençage avant leur utilisation.
Figure 16: Répartition des SNPs le long du génome et correspondance avec la taille de chaque scaffold (pseudomolécule).
Figure 17: Structuration en 3 populations détectée par STRUCTURE à partir des données de génotypage de 2076 SNPs (pas d’admixture, fréquences alléliques corrélées).
Figure 18: Arbre phylogénétique construit selon la méthode UPGMA avec le logiciel DARwin à partir des données de SNP.
LISTE DES TABLEAUX

Tableau 1 : Classement selon leur degré de résistance des principales variétés
Tableau 2 : Coordonnées du barycentre de chaque type d’accessions.
Tableau 3 : Fréquence de l’allèle résistant au marqueur ZP002 par population naturelle d’abricotiers sauvages
Tableau 4 : Répartition des types d’accession génotypées au locus PPVres selon le pays
Tableau 5 : Estimation à partir des données de génotypage par microsatellites du flux de gènes (Nm) entre matériel provenant des différents pays (SSR).
Tableau 6 : Estimation du flux de gènes (Nm) entre les différents sites d’échantillonnage (populations naturelles d’abricotiers sauvages uniquement) du Kazakhstan (données issues des SSR).
Tableau 7 : Flux de gènes (Nm) entre les 6 populations détectées par STRUCTURE (données issues des SSR).
Tableau 8 : Diversité génétique entre les pays (données issues des SSR).
Tableau 9 : Estimatores de diversité globaux résumés par marqueur SSR.
Tableau 10 : Estimatores de diversité de Weir et Cockerman calculés par GENEPOP sur les populations détectées par STRUCTURE (données SSR).
Tableau 11 : Estimatores de diversité de Weir et Cockerman calculés par GENEPOP sur les populations détectées au sein des kazakhs sauvages (données SSR).
Tableau 12 : Fis moyen par scaffold (pseudo-molécule) calculé sur l’ensemble des SNPs.
Tableau 13 : Flux de gènes (Nm) entre les 3 populations détectées par STRUCTURE (données SNP).
Tableau 14 : Diversité génétique entre les 3 populations détectées par STRUCTURE (données issues des SNPs).
Tableau 15 : Estimatores de diversité de Weir et Cockerman calculés par GENEPOP sur les populations détectées par STRUCTURE (données SNP).
INTRODUCTION

Les enjeux de la production fruitière

La production fruitière joue un rôle important dans l’alimentation humaine car elle constitue une source essentielle de vitamines et de fibres. Au niveau socio-économique, elle permet également de valoriser des terres peu adaptées aux grandes cultures, dans des régions rurales déjà défavorisées. Cependant, le suivi des vergers sur de longues périodes allant jusqu’à 20 ans et la contrainte de production d’un fruit sain et appétissant s’accompagnent généralement de nombreux traitements chimiques (pesticides, herbicides, insecticides) afin de lutter contre de nombreuses maladies affectant les arbres fruitiers (bactériennes, fongiques, virales, attaque de ravageurs). De plus, les arbres, de par leur nature pérenne, subissent de plein fouet les effets des changements climatiques. Cela peut se manifester par des pertes non négligeables, que ce soit par manque de froid en hiver ou à cause de gels tardifs (Olukolu et al., 2009). Parmi les espèces fruitières, l’abricot est le troisième fruit produit en France après la pomme, les pêches-brognons et avant la poire, d’où l’importance de l’amélioration variétale de cette espèce pour répondre aux enjeux liés à sa culture (FAOstats).

L’abricotier

Importance économique

La production mondiale d’abricots en 2012 s’élevait à 4 millions de tonnes dans le monde, ce qui représente plus de 1,5 millions d’euros. À l’échelle européenne, 876 mille tonnes ont été produites (22 % de la production mondiale), sachant que le bassin méditerranéen concoure à hauteur de 50 % à la production mondiale (FAOstat ; Bourguiba et al., 2012). Parmi les maladies touchant cette culture fruitière, la sharka est la plus dévastatrice. Elle est due à un virus appelé plus communément le Plum pox virus ou PPV. Il a entraîné des pertes économiques mondiales de plus de 10 milliards d’euros entre 1975 et 2005 sur les arbres du genre Prunus, pertes dues au coût des méthodes de lutte auxquelles s’ajoute la diminution des revenus par pertes de rendement (Cambra et al., 2006).

Histoire

L’abricotier a probablement été domestiqué en Chine il y a plus de 5000 ans et s’est propagé vers l’ouest il y a environ 2000 ans. Il fut apporté sur le continent américain par les espagnols (Faust et al., 1998). Toutefois, trois centres d’origine ont été recensés (Bourguiba et al., 2012) : la Chine, l’Asie centrale et la zone Irano-caucasienne. La dernière zone est considérée comme un centre secondaire de diversification de par sa position géographique intermédiaire entre toutes les zones où l’abricotier est cultivé et l’aire de répartition des formes sauvages. Deux hypothèses sont évoquées pour expliquer l’arrivée de l’abricotier dans le bassin méditerranéen : directement depuis la zone Irano-caucasienne ou bien via deux routes différentes (par le Moyen-Orient grâce aux arabes et par la Hongrie-Europe centrale). La deuxième paraît plus vraisemblable selon Bourguiba et al. (2012).

L’abricotier étant un arbre fruitier, sa période juvénile est longue (jusqu’à 5 ans). Cette espèce aurait donc réalisé moins de cycles sexuels depuis sa domestication par rapport à une plante annuelle dans le même laps de temps. De ce fait, sa domestication peut être considérée comme récente par rapport à une espèce annuelle qui aurait été domestiquée en même temps et on pourrait s’attendre à ce que le goulot d’étranglement lié à la domestication soit plus marqué chez cette espèce. En effet, le temps pour que la diversité réapparaisse par mutation est plus court. La propagation végétative par greffe accentue encore ce phénomène puisqu’elle diminue encore le nombre de cycles sexuels réalisés et permettrait la fixation de génotypes intéressants (Zohary et al., 2012). Cela limite ainsi la diversité génétique de la population en diminuant le nombre de génotypes maintenus mais favorise le maintien de l’hétérozygotie.
(excès d'hétérozygotes attendu quand la reproduction asexuée est privilégiée) (Miller and Gross, 2011). Néanmoins, les premiers résultats obtenus chez l’abricotier (Bourguiba et al., 2012), le pommier (Cornille et al., 2012) et le cerisier (Mariette et al., 2010) montrent que le goulot d’étranglement est étonnamment faible. Cela pourrait s’expliquer par une domestication simultanée en plusieurs endroits et à des flux de gènes avec les populations sauvages (Miller and Gross, 2011; Bourguiba et al., 2012), flux de gènes d’autant plus favorisés par l’allogamie de l’espèce.

**Biologie**

L’abricotier est une espèce pérenne propagée par multiplication végétative (essentiellement greffage) dans le cas des génotypes cultivés. L’abricotier sauvage présente un système d’auto-incompatibilité gamétophytique tandis que parmi les abricotiers cultivés, certains sont auto-compatibles. En effet, dans son aire d’origine, l’abricotier est entouré d’autres arbres pouvant le polliniser. Or, dans son aire de distribution occidentale, ses pollinisateurs sauvages sont absents, ce qui aurait conduit à la sélection d'arbres auto-compatibles (Miller and Gross, 2011). La question est donc de savoir si ces mutations conduisant à l’auto-compatibilité sont aussi présentes dans les populations sauvages, puisque dans le cas contraire, ce caractère serait apparu récemment et sélectionné par l’homme.


**Génétique**

L’abricotier est une espèce diploïde à 2 paires de 8 chromosomes (2n=2x=16) (Doré C., Varoquaux F., 2006) dont la carte physique n'a pas encore été établie, la carte de référence utilisée est donc celle du pêcher (Verde et al., 2013). De nombreuses cartes génétiques sont néanmoins disponibles (Lambert et al., 2004 ; Marandel et al., 2009 ; Dondini et al., 2011 ; Pilarova et al, 2010). Plusieurs caractères agronomiques sont actuellement étudiés à l’INRA dans un but d’amélioration de cette espèce cultivée, dont, notamment, le contrôle de l’auto-incompatibilité, la résistance à la sharka et les effets du réchauffement climatique sur la production fruitière.

**Auto-incompatibilité**

L’auto-incompatibilité gamétophytique est apparentem controtée par un seul locus multi-allélique appelé locus S. Ce locus S est situé sur le chromosome 6, en position 26 422 200 pb. Il code pour une S-RNAs (ribonucléase) qui a une activité dans le pistil et contient un gène de reconnaissance du pollen permettant le rejet du gamète mâle ayant un allèle commun au pistil. Deux introns de tailles variables sont présents dans la séquence de la S-RNAs dont la séquence est elle-même très variable, excepté dans cinq régions conservées (structure observée chez les solanacées).

En plus de cette S-RNAs, un autre gène a été identifié dans ce locus S : le gène SFB (S-haplotype specific F-box) qui contient un motif F-box, deux régions variables et deux régions hypervariables. Ce gène est exprimé dans les tissus polliniques (Figure 1). Ces deux gènes sont en déséquilibre de liaison et sont éloignés de moins de 49 kilobases.
Figure 1: Organisation génétique du locus S (Vilanova et al., 2006)

Figure 2 : Fonctionnement du mécanisme de l’auto-incompatibilité gamétophytique : l’ensemble des S-RNAses est produit dans le pistil, SFB est exprimé dans le pollen et inhibe les S-RNAses correspondant aux pollens à accepter ou bien active celles correspondant au pollen à rejeter.

Figure 3 : Symptômes sur fruits (EPPO)

Figure 4 : Symptômes sur feuilles (EPPO)
Les abricotiers auto-compatibles présentent un gène SFB non fonctionnel dû à une insertion de 358 paires de bases en position 904, ce qui donne un codon stop prématuré à cette position. La protéine produite n'est donc pas fonctionnelle. Les protéines F-box produites pourraient avoir pour rôle d'inhiber ou de dégrader les S-RNAses ne correspondant pas aux allèles du pistil. Cependant, dans le cas où la SFB n'est pas fonctionnelle, il y aurait alors impossibilité pour tout pollen de germer. Une autre possibilité est donc que la SFB protège les S-RNAses ayant l'allèle commun au pistil d'un mécanisme général de dégradation des S-RNAses (Figure 2) (Romero et al., 2004, Vilanova et al., 2006). Cette mutation serait apparue sur l'allèle S6 d'auto-incompatibilité du pollen d'un individu S8S8, i étant inconnu (Halasz 2007). Un autre locus, appelé « modifier », intervient dans le déterminisme de l’auto-compatibilité ; il ne co-localise pas avec le locus d’auto-incompatibilité mais le gène contrôlant ce déterminisme n’a pas encore été cloné (Villanova et al., 2006). Il ne sera pas étudié dans le cadre de ce stage.

**Résistance à la sharka**

**La maladie de la sharka**

Le *Plum pox virus*, ou variole du prunier, est un potyvirus à ARN (10 000 pb) qui atteint bon nombre des espèces du genre *Prunus*, dont des espèces sauvages pouvant servir de réservoir. Les arbres atteints présentent des pertes de rendement liées à la chute des fruits, mais aussi une baisse de la qualité de ces derniers avec des déformations, des nécroses (Figure 3) et une baisse de la teneur en sucre. La quantité et la sévérité des symptômes sur les feuilles (Figure 4) n’a pas de lien avec la charge virale dans la plante. Il a été classé agent pathogène de quarantaine par l’Union européenne et la plupart des pays producteurs. Le virus est apparu pour la première fois en 1916 en Bulgarie, la maladie a été décrite en 1932 et s’est répandue dans la quasi-totalité des pays producteurs de fruits à noyaux. Il est arrivé en France dans les années 60 (Dallot S. et al, 2012 ; James, 2013).

**Dissémination de la maladie**

Le virus est transmis à faible distance par l’intermédiaire des pucerons (environ 20 espèces vectrices, qui ne sont pas toutes inféodées aux arbres fruitiers). Ceux-ci ne sont vecteurs que pendant quelques minutes car le mode de transmission est non persistant, c’est-à-dire que le virus reste temporairement sur les stylets du puceron sans pouvoir effectuer de cycle infectieux ni se multiplier. Ainsi, lorsque quelques arbres sont infectés, le virus se propage au reste du verger en 2 à 3 ans par les piqûres d’essai répétées des pucerons. Cependant, des variétés sensibles peuvent rester saines plus de 15 ans, même en présence du virus dans l’environnement. La transmission à longue distance se fait quant à elle par les échanges de greffons contaminés (Dallot S. et al, 2012 ; Marandel 2008 ; Doré C., Varoquaux F., 2006, EPPO).

**Méthodes de lutte**

Du fait de sa faible durée de vie dans le puceron, la lutte insecticide en éliminant les pucerons est inefficace car les produits mettent plus longtemps à agir qu’il n’en faut au virus pour être inactif. De plus, les traitements insecticides n’ont aucun effet sur les populations de pucerons qui ne sont pas inféodés aux arbres fruitiers car ceux-ci ne s’installent pas sur l’arbre. Ils ne font que passer, acquérir et transmettre le virus par piqûres d’essai. Enfin, les produits chimiques utilisés ont tendance à exciter le puceron, qui piquera plus souvent et propagera plus vite le virus, avant de quitter le verger. Comme il n’existe aucun moyen pour lutter contre le virus une fois qu’un arbre est infecté, des méthodes de détection des arbres infectés ont été développées pour éliminer les arbres et greffons atteints et diminuer ainsi la taille du réservoir. Cependant, le coût d’éradication de la maladie reste élevé et fragilise encore plus l’économie.
Tableau 1 : Classement selon leur degré de résistance des principales variétés (Decroocq et al., 2011 ; Martinez-Gomez et Dicenta, 2000)

<table>
<thead>
<tr>
<th>Cultivar</th>
<th>Origine</th>
<th>Pedigree</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Partiellement résistant</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Goldrich</td>
<td>USA</td>
<td>(Sunglo x Perfection)$^1$</td>
</tr>
<tr>
<td>Veecot</td>
<td>Canada</td>
<td>(pollinisation libre de Reliable)$^1$</td>
</tr>
<tr>
<td>Henderson</td>
<td>USA</td>
<td>Unknown$^1$</td>
</tr>
<tr>
<td>Sunglo</td>
<td>USA</td>
<td>Columbia &amp; Okanagan Nursery Co$^1$</td>
</tr>
<tr>
<td>Pandora</td>
<td>Grèce</td>
<td>(SEO x P. Tirynthos)$^2$</td>
</tr>
<tr>
<td>Harcot</td>
<td>Canada</td>
<td>[(Geneva x Naramata) x Morden 604] x (Phelps x Perfection)</td>
</tr>
<tr>
<td>Bora</td>
<td>Italie</td>
<td>‘Early Blush’ (NJA53) x ‘PA 7005-2’ (‘Rival’ x ‘PA 63-265’)</td>
</tr>
<tr>
<td>Robada</td>
<td>USA</td>
<td>(Orangered x K113-40)</td>
</tr>
<tr>
<td><strong>Résistant</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NJA 2</td>
<td>USA</td>
<td>Pollinisation libre de Morden 604</td>
</tr>
<tr>
<td>NJA 17</td>
<td>USA</td>
<td>(pollinisation libre de Erli Orange)x(Morden 604 x Perfection)</td>
</tr>
<tr>
<td>NJA 42</td>
<td>USA</td>
<td>(NJA13(=NJA2x[Perfection x [Perfection x (Royal x Blush)]) x NJA12(=Goldcot x Priusadelnyi)</td>
</tr>
<tr>
<td>Orange Red</td>
<td>USA</td>
<td>(Lasgerdi Mashhad x NJA 2)$^3$</td>
</tr>
<tr>
<td><strong>Très résistant</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SEO</td>
<td>USA</td>
<td>Stark Bros. Nurseries &amp; Orchards$^1$</td>
</tr>
<tr>
<td>Stella</td>
<td>USA</td>
<td>inconnu</td>
</tr>
<tr>
<td>Harlayne</td>
<td>Canada</td>
<td>[(Reliable x pollinisation libre.) x pollinisation libre.] x Sunglo$^3$</td>
</tr>
<tr>
<td>Lito</td>
<td>Grèce</td>
<td>(SEO x P.Tirynthos)$^2$</td>
</tr>
<tr>
<td><strong>Résistance indéterminée</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Traian</td>
<td>Romania</td>
<td>Viceroy x NJA2</td>
</tr>
</tbody>
</table>

Figure 5 : Position des différents marqueurs liés à PPVres (Decroocq S. et al., 2014)
agricole dans des zones rurales déjà marginalisées et/ou fragiles (bassin méditerranéen, zones de cultures en terrasse, pays de l’Est Européen).

La seule solution à long terme pour contrer le virus passe donc par la lutte variétale. Cependant, la difficulté d’avoir un phénotypage fiable en moins de trois ans ralentit le processus de sélection qui est déjà long. La SAM (Sélection Assistée par Marqueur) pourrait être un outil permettant de raccourcir le processus de sélection de variétés résistantes au virus et donc de développer des cultivars adaptés aux différents agrosystèmes européens (Dallot S. et al, 2012 ; EPPO ; Decrooocq et al., 2011).

Mécanismes de résistance

Il existe actuellement peu de variétés résistantes à la sharka et la plupart sont d’origine nord-américaine (Tableau 1). Ces variétés ne présentent aucun symptôme, mais peuvent présenter, de manière plus ou moins concentrée, des particules virales. Les plus résistantes ne sont pas immunes mais ne permettent pas l’établissement et la propagation de la maladie dans la zone où elles sont plantées. Il a été montré que ces variétés n’appartiennent ni au groupe européen ni au groupe irano-caucasien. Même si elles semblent être apparentées avec le germplasm européen, elles partagent probablement une même origine génétique, provenant d’Asie centrale et/ou de Chine comme l’ont démontré Zhebentayeva et al. (2008) au travers de marqueurs AFLP et SSR.

La résistance à cette maladie est quantitative et probablement contrôlée par plusieurs gènes. Parmi ceux-ci, Marandel et al. (2009) indique qu’un locus à effet majeur, nommé MetaPPV1a, a été trouvé dans toutes les études réalisées jusqu’à présent et est positionné sur le haut du groupe de liaison 1. La position de ce locus a été affinée et il a été renommé PPVres (Soriano et al., 2011). Trois marqueurs SSR en déséquilibre de liaison et jalonnant PPVres ont été développés : PGS1.21, PGS1.23 et PGS1.24. Les allèles résistants à chacun de ces marqueurs sont dominants. (Soriano et al., 2012 ; Decrooocq S. et al., 2014 ; Zuriaga E. et al., 2013). Un autre marqueur SSLP (Simple Sequence Length Polymorphism), nommé ZP002 (Figure 5) est situé dans l’intervalle de PPVres, à 8 157 718 pb sur le haut du chromosome 1. Il correspond à une délétion de 5 paires de bases chez les individus résistants (Decrooocq et al., 2014). Cependant, quel que soit le marqueur, PGS ou ZP, il existe des accessions ayant un génotype résistant à ces marqueurs et qui présentent un phénotype sensible. Ceci pourrait être dû au fait qu’un autre locus intervient dans l’expression de la résistance et module l’activité de PPVres (Decrooocq S. et al., 2014). Selon cette hypothèse, il faudrait alors cumuler un haplotype résistant pour PPVres et pour cet autre locus afin d’avoir un individu résistant à la sharka.

Adaptation au réchauffement climatique

Du fait du réchauffement climatique, les arbres voient leurs besoins en froid de plus en plus difficilement satisfaits. Il est donc important de sélectionner des variétés ayant de faibles besoins en froid. Cependant, les sources de résistance à la sharka (variétés nord-américaines) utilisées jusqu’à présent ont des caractéristiques agronomiques (faible rendement en fruits, auto-incompatibilité, besoins en froid hivernal élevés) qui les rendent impropres à une utilisation directe en vergers. Elles sont donc utilisées en croisement avec les variétés locales mais cette introgression se heurte à un problème d’adaptation aux conditions locales du fait de leurs besoins en froid très importants difficilement satisfaits. Le contrôle génétique de la levée de dormance et de la réponse au froid chez l’abricotier a été décrit par approche QTL dans une population F1 issue d’un croisement entre le cultivar Perfection et l’accession A1740 (Olukolu et al., 2009). Il fait intervenir au moins 5 loci, répartis sur l’ensemble du gênome de l’abricotier.
Hypothèses et objectifs de cette étude

L’étude « genome-wide », dite également pangénomique, de la résistance à la sharka est en cours sur les abricotiers cultivés (Projets ANR ABRIWG et européen FP7 MARS). Une étude similaire est menée à partir de données de floraison, en lien avec l’adaptation aux changements climatiques, mais ceci concerne l’induction florale qui n’est que partiellement corrélée au débourrement végétatif (Fan et al., 2010). En comparaison, l’étude du compartiment sauvage fut jusqu’à présent délaissée. Or, Lasnier (2013) a montré précédemment que les variétés actuelles d’abricotiers sélectionnés n’intègrent pas la diversité des variétés locales asiatiques, ce qui laisse supposer que les individus sauvages présentent un intérêt non négligeable pour la sélection. Dans le cas de la résistance à la sharka ou des besoins en froid, il est également possible que du matériel génétique intéressant se trouve dans les populations naturelles originaires d’Asie Centrale, ce qui confirmerait les données de Zhebentiayeva et al. (2008). Ainsi, l’hypothèse sous-jacente à l’ensemble de ce travail est que les individus faisant partie des populations sauvages présenteraient une diversité significativement plus grande que ceux du compartiment cultivé. Si cette hypothèse s’avère vraie, ceci devrait permettre au sélectionneur de trouver plus de variabilité phénotypique dans ce matériel sauvage, utilisable dans de futurs programmes de sélection alliant à la fois la résistance à la sharka, une meilleure adaptation au réchauffement climatique et bien d’autres caractères d’intérêt agronomique.

Ce travail s’inscrit dans la continuité du projet européen STONE (FP7-PEOPLE-2009-IRSES, ‘Genetic Diversity of Stone Fruit trees (Apricot, Peach and Cherry) in Europe, Caucasus and Central Asia’). Il a pour objectif de caractériser la diversité et la structuration de l’abricotier sauvage afin d’estimer son intérêt et son potentiel dans les programmes de sélection de l’abricotier en Europe. Cette étude sera menée à trois niveaux : phénotypique (débourrement des bourgeons végétatifs à la levée de dormance), à deux loci d’intérêt (système d’auto-incompatibilité et résistance à la sharka) et à l’échelle du génome complet avec des marqueurs SSR (microsatellites) et SNP (Single Nucleotide Polymorphism). L’intérêt de cette étude est multiple :

- La diversité phénotypique au débourrement permet d’appréhender la diversité de la réponse au froid (température inférieure à 7°C), qui est un critère important dans le processus de sélection des arbres fruitiers, surtout dans le contexte actuel du réchauffement climatique. En effet, il serait logique que les arbres issus de climats froids aient des besoins plus forts que ceux de climats plus méditerranéens, l’idée est donc de caractériser les besoins en froid des populations sauvages.
- La diversité au locus contrôlant, pour partie, la résistance à la sharka permet d’étudier le potentiel des populations sauvages comme réservoir de sources de résistance à cette maladie. De plus, l’étude de ces populations permet également de savoir si la présence de l’allèle résistant au QTL PPVires, repéré dans les populations cultivées, est également associée à un phénotype résistant dans les populations naturelles.
- La diversité au locus d’auto-incompatibilité permet d’étudier l’évolution du système de reproduction lors de la domestication et d’étudier l’origine de l’auto-compatibilité (apparition uniquement dans le compartiment cultivé ou bien provenant du compartiment sauvage).
- La diversité à l’échelle du génome entier permet, quant à elle, de reconstituer l’histoire évolutive de l’espèce sur plusieurs aspects : taille des populations, domestication, sélection, adaptation, structuration, flux de gènes entre populations et entre compartiments sauvages et cultivés.
Figure 6 : Lieux de prélèvement des échantillons. Les lieux de prélèvement en jaune concernent les individus sauvages, en vert, ce sont les variétés anciennes.
MATÉRIEL et MÉTHODES
Matériel végétal
Des feuilles et des noyaux d’abricotiers sauvages ont été récoltés de 2011 à 2013, dans le cadre du projet STONE, dans la zone d’origine de l’abricotier (genre Prunus, sous-genre Prunophora, section Armeniaca, espèce Prunus armeniaca selon la classification de Rehder, 1940 ; renommé Prunus armeniaca vulgaris par Lingdi & Bartholomew, 2003) soit en Chine de l’ouest, au Kazakhstan, au Kirghizstan et en Ouzbékistan ; dans les bassins de domestication secondaire : en Turquie, Azerbaïdjan, et dans des collections d’abricotier cultivé (variétés anciennes). Les sites de prélèvement sont répertoriés en Annexe et localisés sur la Figure 6. Cependant, les individus utilisés dans chaque partie ne sont pas toujours les mêmes du fait des contraintes d’échantillonnage (feuilles des pieds-mères en mauvais état, absence de noyaux pour les semis suite à des périodes de gel printanier) (Figure 7). En ce qui concerne les notations du caractère « débourrement », il s’agit d’individus issus exclusivement de semis. Dans le cas du matériel génotypé avec des marqueurs de type SSR (Single Sequence Repeat ou microsatellites), les feuilles lyophilisées issues de plants mère ont été analysées, tandis que pour les SNP, le choix a été fait de ne conserver que des populations équilibrées (au moins 8 individus par population) et des individus dont suffisamment d’ADN a pu être extrait pour le séquençage par la technique ILLUMINA. Les données sont donc issues de semis pour cette partie. Enfin, pour les deux marqueurs ZP002 et SFB, un maximum d’accessions, plants mère et semis, ont été étudiées. Les accessions utilisées pour chaque partie sont résumées en Annexe II. Il faut préciser ici que dans les noms de chaque accession, les deux premières lettres correspondent au pays d’origine ou au lieu de conservation de la collection. Lorsque deux chiffres sont présents comme dans kz_55_1 et kz_55_2, il s’agit de demi-frères car les noyaux sont issus des fruits d’un seul arbre. Différentes variétés cultivées («SE0», «Bakour», «Bergeron», «Canino», «Currot», «Goldrich», «Luizet», « Moniqui » et « Stella »), européennes ou nord-américaines ont été utilisées comme contrôles pour les marqueurs ZP002, SFB et SSR.
Phénotypage
Les accessions (exception faite des arbres kirghiz qui n’avaient pas encore subi de vernalisation) sont conservées sous tunnel et notées après une période de froid hivernal (températures inférieures à 7°C) sur leur date de débourrement, avec un suivi tous les 2 jours. Une note de 0 à 6 (Figure 8) (avec des intermédiaires) a été attribuée :

0= pas de débourrement
1= les écailles des bourgeons s’écartent et la couleur verte des feuilles est visible
2= les feuilles percent le bourgeon
3= les feuilles s’écartent du bourgeon sur moins de 50 % des bourgeons
4= les feuilles s’écartent du bourgeon sur plus de 50 % des bourgeons
5= au moins trois feuilles sont étalées sur chaque bourgeon
6= les branches axillaires s’allongent et s’écartent de l’axe principal, les feuilles sont étalées le long de la branche.

Pour l’analyse statistique, trois variables de cumul supplémentaires ont été utilisées : la somme d’heures de froid (température inférieure à 7°C) nécessaires pour atteindre de stade 3 (HF), la quantité de rayonnement solaire pour atteindre le stade 3 (RG) depuis le premier septembre, et le nombre d’heures de jour depuis l’équinoxe d’hiver pour atteindre le stade 3 (HJ).
Figure 7 : Parties des arbres récoltées et leur utilisation.

Figure 8 : Échelle de notation utilisée pour caractériser les stades de débourrement.
Les données météorologiques nécessaires au calcul de ces cumuls sont issues des relevés horaires de la station météorologique du centre INRA-La Ferrade (Villenave d’Ornon) disponibles sur le site Climatik (https://intranet.inra.fr/climatik). Ces données ont été traitées par des ACPs (Analyse en Composantes Principales) et des ANOVAs (Analyse de variance) grâce au logiciel R. Ensuite, une analyse de la structure restreinte à la seule variable HF a été réalisée avec GENELAND avec les conditions suivantes : burn-in : 200, nombre d’itérations : 100000, thinning : 100, modèle spatial, coordonnées incertaines : delta=0,005 (un seul relevé GPS pour plusieurs arbres proches), fréquences corrélées ou non, 10 répétitions.

Analyse du marqueur d’auto-compatibilité


Analyse du marqueur de résistance à la sharka

Vera-Ruiz et al. (2011) ont montré précédemment que la résistance à la sharka est contrôlée par un locus majeur, nommé PPVres. Un certain nombre de marqueurs liés à ce locus ont été développés et testés en sélection assistée par marqueurs sur des populations cultivées (Soriano et al., 2012 ; Decroocq et al., 2014). Le marqueur choisi dans notre cas est le ZP002 dont les amorces sont, en sens, AAC A TT TTC TGA TTC AAT GCC A et en anti-sens TGT ATC CTC CAG CTT CAA AGTC. La PCR est réalisée dans les conditions suivantes : 94°C pendant 2 min, 30 cycles avec 94°C pendant 1 min, 57°C pendant 40s, 72°C pendant 40s, et pour finir un cycle à 72°C pendant 20 min. La vérification de l’amplification est réalisée sur gel d’agarose à 2%. Ce marqueur génère des fragments de 112 pb chez les individus résistants et 117 pb chez les individus sensibles. La lecture des résultats a lieu sur gel d’acrylamide à 6 %. Les données ont ensuite été analysées par une comparaison de moyennes selon l’origine. Pour ce faire, une note issue de la moyenne de la taille de ses deux allèles a été affectée à chaque individu, ce qui permet de conserver l’information sur l’hétérozygotie

Analyse de la structure des populations avec les marqueurs SSR

Avant le début de ce stage, vingt-trois marqueurs de type SSR (microsatellites) ont été testés sur 301 individus issus de collecte de variétés anciennes et de populations naturelles du Kazakhstan, Chine et Ouzbékistan. Les individus génotypés avec ces marqueurs sont les pieds mère sur lesquels 3 à 4 feuilles ont été prélevées et desséchées sur place avant leur retour au laboratoire. Dans certains cas, la qualité des feuilles n’étant pas suffisante ou les feuilles n’ayant pas pu être prélevées (branches inatteignables), du matériel végétal issu de semis a été analysé. «SE0», «Bakour», «Bergeron», «Canino», «Currot», «Goldrich», «Luizet», «Moniqui » et « Stella » ont servi de contrôles. Les données brutes (taille des allèles) obtenues par PCR puis lecture sur séquenceur automatique ont été reprises et utilisées dans les logiciels STRUCTURE (version 2.3.4, burn-in : 10000, running : 100000 ou 250000, admixture, modèle spatial ou non) et GENELAND (version 4.0.0, burn-in : 200, nombre
d’itérations : 100000, thinning : 100, modèle spatial à coordonnées incertaines : delta=0,005) afin d’étudier la structuration. Les analyses ont été réalisées sur 10 répétitions avec ou sans les fréquences alléliques corrélées (c’est-à-dire que les fréquences alléliques de chaque population sont dérivées des fréquences alléliques d’une population ancestrale). Le logiciel PopTREE2 a été utilisé pour construire un arbre phylogénétique selon la méthode UPGMA (Unweighted Pair Group Method with Arithmetic Mean). GENEPOP (version 4.2), ADZE (version 1.0) (G=taille de la population la plus petite-1) et GENETIX (version 4.05.2) ont été utilisés conjointement pour leur complémentarité pour étudier la différenciation des populations.

**Extraction de l’ADN et séquençage**

Les extractions d’ADN, que ce soit pour le séquençage haut-débit ou les PCRs, ont été réalisées selon le protocole disponible en Annexe III. Le séquençage de 77 individus a été réalisé en paired-end avec une couverture minimum de 15X (soit 15 fois le génome abricotier de 230Mb) selon la méthode Illumina® par DHMRI (David H. Murdock Research Institute, 150 Research Campus Drive, Kannapolis, NC 28081, USA).

**Nettoyage des données de séquençage issues de 77 accessions d’abricotiers sauvages**

Les données brutes ont été soumises à un test de qualité qui élimine les « reads » (lectures de séquence) n’ayant pas un score minimal de 20 sur au moins 70 paires de bases. Les reads restants ont été alignés sur le génome du pêcher grâce au logiciel BWA. A partir de cet alignement, seuls les reads correspondants à des séquences géniques (introns et exons) ont été conservés. Une première détection des SNPs par rapport au pêcher a été réalisée grâce à SAMtool. Un filtrage pour repérer les individus hétérozygotes aux loci polymorphes a été effectué grâce à l’outil testassoc (Kim et al., 2012) avec les critères suivants : une profondeur de lecture minimum de 8x et au minimum deux reads portant l’allèle mineur. Si un SNP est observé avec le pêcher, il est conservé uniquement s’il est également polymorphe au sein des abricotiers. Pour ce faire, les alignements des accessions avec le pêcher ont donc été comparés deux à deux (pair-wise) entre abricotiers dans une même région. Finalement, ne sont conservés que les SNPs intra abricotier aux loci ne présentant pas de données manquantes parmi tous les individus séquencés. L’ensemble de la procédure est résumé sur la Figure 15.

**Analyse des données de séquençage**

Une première analyse est réalisée grâce au logiciel mlRho pour estimer Θ (estimateur de la diversité) et ε (taux d’erreur de séquençage) pour chaque accession. Ces valeurs servent également dans le logiciel testassoc pour l’identification des vrais hétérozygotes en comparaison des erreurs de séquençage. Les données filtrées sont ensuite analysées sur 10 répétitions avec TESS (version 2.3.1), STRUCTURE (version 2.3.4) avec et sans admixture, avec et sans les fréquences alléliques corrélées, burn-in=100000, thinning=100000. Ensuite, le logiciel DARwin (version 5) a été utilisé pour construire un arbre phylogénétique selon la méthode UPGMA (Unweighted Pair Group Method with Arithmetic Mean). Puis, les estimateurs de diversité de Wright et Fischer ont été calculés grâce à la version web de GENEPOP (version 4.2), aux logiciels ADZE (version 1.0) et GENETIX (version 4.05.2). Les fichiers d’entrée ont été générés à l’aide de PGDspider version 2.0.5.2.
Figure 9 : Sorties graphiques de l’ACP sur les deux premières dimensions. L’ensemble des notes de la période d’observation d’un notateur sont utilisées et l’individu 21 a été supprimé. Un effet taille est observable sur le cercle des corrélations et un effet Guttman sur le nuage des individus.

Tableau 2 : Coordonnées du barycentre de chaque type d’accessions. Dim 1 correspond aux coordonnées sur le premier axe de l’ACP et Dim 2 correspond aux coordonnées sur le deuxième axe de l’ACP.

<table>
<thead>
<tr>
<th>Type</th>
<th>Dim1</th>
<th>Dim2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variété ancienne</td>
<td>1.75815453</td>
<td>-0.110562786</td>
</tr>
<tr>
<td>Cultivar local</td>
<td>-0.85264460</td>
<td>0.982401980</td>
</tr>
<tr>
<td>Accession sauvage</td>
<td>-1.50823931</td>
<td>0.060073222</td>
</tr>
</tbody>
</table>

Figure 10 : Structuration détectée par GENELAND (fréquences corrélées) sur le phénotype (HF=nombre d’heures de froid pour atteindre le stade 3). Les individus apparentés ont été supprimés. Structuration en trois groupes des individus phénotypés. Les chiffres correspondent aux groupes mentionnés dans le texte.
RESULTATS
Durant ce stage, des études de diversité phénotypique et génétique ont été réalisées sur des abricotiers d’Europe orientale (Caucase, Turquie) et d’Asie centrale, pour sa partie nord (Kazakhstan, Ouzbékistan, Chine de l’ouest et Kirghizstan). Nous disposions de trois types de matériel : 1) des variétés anciennes, traditionnelles (arbres maintenus dans les jardins ou en bord de chemin ou bien encore en collection de germplasm issu d’arbres prélevés localement), 2) des cultivars locaux (vergers de production (ex : marché d’Almaty)) et 3) des accessions sauvages (populations naturelles dans les forêts du massif du Tian Shan) (voir Figure 6).

Analyse de la diversité phénotypique du caractère « débourrement » chez les abricotiers d’Asie centrale et du Caucase

Analyse en composantes principales (ACP)
L’ACP réalisée (Figure 9) explique 87 % de la variabilité observée uniquement grâce aux deux premiers axes qui représentent la précocité. L’étude des variables supplémentaires qualitatives (type d’accession, pays d’origine, population) montre que les variétés anciennes sont les plus précoces et les arbres sauvages les plus tardifs (Tableau 2). Refaire les analyses par type d’accession ne modifie pas beaucoup les résultats et les interprétations (résultats non présentés).

Concernant l'origine géographique, il ressort un gradient lié à la latitude puisque les individus ouzbeks sont plus précoces, puis viennent les azéris, les chinois et les kazakhs. Au niveau des populations, cette hiérarchisation latitudinale se confirme (Annexe V).

Afin d’appréhender les données autrement, les stades de débourrement ont été remplacés par les variables HF et RG, plus quantitatives (Annexe IV). Il ressort des analyses que les deux variables HF et RG sont fortement corrélées positivement (r=0,942) sur la première dimension et expliquent, à elles seules, 100% de la variabilité observée. La répartition par type d’accession est conservée (les individus en retard sont ici à droite), ainsi que celle par pays (à l’exception de la Chine et du Kazakhstan qui sont inversés).

Analyse de variance
Une MANOVA a été réalisée sur les notes obtenues sur la totalité de la période de notation. Il en ressort qu’à la fois HF, RG et HJ ont un effet significatif sur les notes obtenues, de même que leurs interactions (effets combinés) (Annexe VIII). Lorsque l’analyse est décomposée en variables qualitatives (Annexe VII), seuls les effets du type d’accession et de la population d’origine sont significatifs, ce qui rejoint les résultats de l’ACP. La recherche des variables qualitatives ayant des effets significatifs sur les données HF, RG et HJ séparément (ANOVA), indique que seule la population d’origine a un effet significatif (Annexe VI). Cependant, si ces trois variables HF, RG et HJ sont considérées comme dépendantes (Annexe IX), le type d’accession est alors également un facteur significatif. Les individus sauvages puis les variétés anciennes ont donc été étudiés séparément. Que ce soit pour les accessions sauvages ou les variétés anciennes, l’effet de la population reste significatif. Les effets des variables quantitatives HF, RG et HJ sont également conservés, excepté l’interaction RG-HJ dans le cas des individus sauvages.
Figure 11: Structuration détectée par GENELAND (fréquences non corrélées) sur le phénotype (HF=nombre d'heures de froid pour atteindre le stade 3). Les individus apparentés ont été conservés. Structuration en quatre groupes des individus phénotypés.
Structuration des abricotiers d’Asie centrale et du Caucase selon la variabilité phénotypique

Analyse de la structuration chez les individus sauvages et les variétés anciennes

Lorsque les données phénotypiques (ici, données sur HF) sont utilisées dans GENELAND afin de tester une possible structuration dans ce matériel végétal, les variétés anciennes et les individus sauvages ne sont pas nettement séparés (Figure 10 et Figure 11). Cependant, dans le cas des fréquences alléliques corrélées (Figure 10), les individus provenant d’Aksu (population naturelle du parc national d’Aksu-Zhabagly), de Soujetskoe et des environs de Sayram (kz_139 à kz_154), c’est-à-dire tous à l’ouest du kazakhstan (environ à 600km d’Almaty) sont regroupés avec les ouzbeks (groupe 2) tandis que les azéris forment un groupe intermédiaire (groupe 3) et que les autres kazakhs forment un groupe homogène (groupe 1). En comparaison, dans le cas des fréquences alléliques non corrélées (Figure 11), la moitié des kazakhs du lac de Big Almaty forme un groupe à part (1), de même que les ouzbeks de Samarcande et les azéris qui intègrent également un individu chinois d’Urumqui (ch_137) (4) et que les individus kazakhs des environs de Sayram, de Chymkent et de Soujestkoe (ouest du Kazakhstan) sont associés avec les ouzbeks de Boukhara et de Chimgan Beldersay (3). Les autres forment un grand groupe (2). Ces résultats sont obtenus que les individus apparentés soient conservés ou non.

Analyse de la structuration au sein des individus sauvages uniquement

L’analyse a tout d’abord été réalisée sans intégrer les coordonnées spatiales. Lorsque les fréquences ne sont pas considérées comme corrélées, le nombre de populations le plus probable est de 8. Cependant, chaque « run » (répétition) donne un nombre de populations différent (entre 7 et 13) et les demi-frères sont presque toujours assignés à des populations différentes, ce qui pourrait indiquer que le caractère de débourrement est tout aussi variable au sein des populations qu’entre les populations. Au contraire, si les fréquences alléliques sont considérées comme corrélées (dérivées d’une population ancestrale), une seule population est obtenue. L’ajout des coordonnées permet d’identifier deux populations principales séparant les individus provenant du parc national d’Aksu Zhabagly des autres. Lorsque les fréquences ne sont pas corrélées, la séparation du groupe d’Aksu-Zhabagly des autres abricotiers est conservée avec deux sous-populations dans la grande population précédente : d’un côté la moitié des individus issus de Big Almaty Lake et la population de la vallée de Medeu, de l’autre les chinois et les autres kazakhs (Annexe X). Il est intéressant de noter ici que la population naturelle d’abricotiers sauvages de Big Almaty Lake se structure en deux groupes distincts, sur la base de leur besoin en froid pour débourrer (Annexe X, kz_100 à 104 d’une part et kz_122 à 127, d’autre part). Aucune structure plus fine n’a pu être détectée intra-population.

Analyse sans individus apparentés

Les résultats présentés jusqu’ici ont tous été obtenus en ayant des individus apparentés (demi-frères), ce qui constitue une entorse à l’équilibre de Hardy-Weinberg (hypothèse utilisée dans les analyses de structure des populations). Il nous paraissait néanmoins important de travailler ainsi du fait du faible nombre d’individus non apparentés disponibles. D’autant plus que Cornille et al (2012) ont obtenu des résultats identiques en intégrant ou non les apparentés dans leur étude. Finalement, si les apparentés sont retirés, nous retrouvons toujours la séparation des individus d’Aksu-Zhabagly du reste des sauvages, à condition de prendre en compte les coordonnées spatiales.
Figure 12 : Exemple de gel de polyacrylamide obtenu avec le marqueur ZP002. « Stella » et « SE0 » sont des variétés résistantes utilisées comme contrôles, la première étant homozygote au locus PPVres et la seconde hétérozygote. Le premier puits correspond à l’échelle de taille (100pb).

Tableau 3 : Fréquence de l’allèle résistant au marqueur ZP002 par population naturelle d’abricotiers sauvages

<table>
<thead>
<tr>
<th>Population</th>
<th>Pays</th>
<th>Fréquence des individus portant l’allèle associé à la résistance-homozygotes compris (%)</th>
<th>Fréquence des individus présentant l’allèle associé à la résistance à l’état homozygotes (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aksu reserve</td>
<td>Kazakhstan</td>
<td>37.5</td>
<td>0</td>
</tr>
<tr>
<td>Sary Chelek</td>
<td>Kirghizstan</td>
<td>35</td>
<td>3</td>
</tr>
<tr>
<td>Chimgan Beldersay</td>
<td>Ouzbékistan</td>
<td>30</td>
<td>10</td>
</tr>
<tr>
<td>Chyckan river</td>
<td>Kirghizstan</td>
<td>16</td>
<td>0</td>
</tr>
<tr>
<td>Arslanbob</td>
<td>Kirghizstan</td>
<td>30</td>
<td>2</td>
</tr>
<tr>
<td>Kaskelen canyon</td>
<td>Kazakhstan</td>
<td>25</td>
<td>10</td>
</tr>
<tr>
<td>Belbulak canyon</td>
<td>Kazakhstan</td>
<td>50</td>
<td>0</td>
</tr>
<tr>
<td>Turgen valley</td>
<td>Kazakhstan</td>
<td>32</td>
<td>0</td>
</tr>
<tr>
<td>Esik lake</td>
<td>Kazakhstan</td>
<td>75</td>
<td>42</td>
</tr>
<tr>
<td>Mehe</td>
<td>Chine</td>
<td>44</td>
<td>0</td>
</tr>
<tr>
<td>Medeu valley</td>
<td>Kazakhstan</td>
<td>46</td>
<td>8</td>
</tr>
<tr>
<td>Big Almaty lake</td>
<td>Kazakhstan</td>
<td>55</td>
<td>7</td>
</tr>
</tbody>
</table>

Tableau 4 : Répartition des types d’accession génotypées au locus PPVres selon le pays

<table>
<thead>
<tr>
<th></th>
<th>variété ancienne</th>
<th>cultivar local</th>
<th>mélange</th>
<th>Accession sauvage</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chine</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>18</td>
</tr>
<tr>
<td>Kazakhstan</td>
<td>66</td>
<td>10</td>
<td>0</td>
<td>123</td>
<td>199</td>
</tr>
<tr>
<td>Kirghizstan</td>
<td>41</td>
<td>0</td>
<td>26</td>
<td>172</td>
<td>239</td>
</tr>
<tr>
<td>Ouzbékistan</td>
<td>13</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>15</td>
</tr>
</tbody>
</table>
**Synthèse des résultats obtenus lors de l’analyse de la diversité phénotypique chez les abricotiers provenant d’Asie centrale et du Caucase**

Les analyses réalisées sur les données de débourrement ont permis de montrer que : 1) les variétés anciennes sont plus précoces que les individus sauvages mais ne forment pas une population distincte, 2) la précocité est liée à la population d’origine et non au pays, 3) la structuration est plutôt faible, les kazakhs de l’ouest étant tantôt regroupés avec les autres kazakhs, tantôt avec les ouzbeks ou les azéris, et la population d’Aksu-Zhabagly semblant particulière, 4) la variabilité phénotypique inter-populations est équivalente à la variabilité intra-populations. L’apparente contradiction entre les résultats d’analyse de variance et de l’ACP avec ceux de structuration peut être liée au fait qu’une seule variable a été utilisée pour étudier la structuration. Cependant, l’étude de ce caractère aura permis d’identifier des individus avec un potentiel améliorateur.

**Diversité génétique aux loci PPVres et SI**

*Locus majeur contrôlant la résistance à la sharka*

Le marqueur étudié ici, ZP002, est lié au locus PPVres qui contribue à la résistance au virus. Nous avons, dans ce cas, inclus dans l’analyse les individus kirghiz pour lesquels l’ADN a pu être extrait à partir des semis durant ce stage. Le type de résultats obtenu est présenté dans la Figure 12. A première vue, l’allèle lié à la résistance à la sharka semble répandu de manière très hétérogène dans les populations naturelles mais est quasiment absent des variétés cultivées (Tableau 3). Afin d’étudier plus en détail la répartition de cet allèle de résistance au sein des populations, des comparaisons de moyennes ont été effectuées.

Il ressort d’une première analyse que les individus provenant du Kazakhstan et de l’Ouzbékistan présentent la fréquence la plus élevée d’individus portant l’allèle lié à la résistance et ceux du Kirghizstan la fréquence la moins élevée; et que les cultivars locaux ont significativement l’allèle lié à la résistance moins fréquent que les variétés anciennes traditionnelles et les individus sauvages. Cependant, la répartition par type d’accession est fonction du pays (Tableau 4). L’analyse a donc été refaite par étapes successives.

Si l’on ne travaille que sur les variétés anciennes et les accessions sauvages, la différence de moyenne est significative, bien que très faible, selon le type (variétés anciennes=111.64, sauvages=111.96). Si cette l’analyse est refaite par pays, il en ressort que la différence entre les types n’est significative que pour la Chine (variétés anciennes=110.5, sauvages=112.68). De plus, il apparait que chez les variétés anciennes chinoises l’allèle résistant est plus fréquent que chez les variétés anciennes des autres pays. Ceci pourrait signifier un flux de gène plus important entre sauvages et variétés locales en Chine que dans les autres pays.

Au niveau des populations kazakhes, trois populations se distinguent : vallée de Turgen (b), lac d’Esik (a) et canyon de Kaskelen (c). La population du lac d’Esik est celle qui présente la fréquence la plus élevée de l’allèle résistant, notamment à l’état homozygote. Les autres populations sont intermédiaires. Parmi les populations kirghizes, seules celles de la rivière Chychkan et du canyon de Boom (rivière de Chuy) se démarquent, le canyon de Boom étant le plus riche en allèle lié à la résistance.
Figure 13 : Structuration des populations d’abricotier détectée par STRUCTURE à partir des données des SSR. Les individus apparentés ont été éliminés.
**Auto-Incompatibilité**


**Diversité génétique globale à partir des données de génotypage issues de 21 loci microsatellites**

Les résultats présentés dans cette partie ont été obtenus grâce aux données issues de 20 microsatellites parmi les 24 utilisés par Bourguiba et al. (2012) sur l’abricotier cultivé (les 3 autres générant un taux élevé d’allèles nuls, de profil monomorphe ou reflétant une mauvaise qualité d’amplification ou de lecture) et auxquels a été rajouté PGS1.21 (marqueur lié au locus PPVres). Un total de 21 loci microsatellites (Tableau 9) a donc été retenu pour l’analyse suivante.

**Structuration**

L’analyse réalisée avec le logiciel STRUCTURE sépare les individus en 6 populations (Figure 13) comme suit : en jaune, les azéris, les cultivars de contrôle (excepté « SEO » et « Goldrich »), les turkmens et les turcs (groupe 3), en bleu les chinois de Turpan et Luntai, les ouzbeks et les kazakhs de l’ouest (2), en violet les chinois de Xiongyue et les cultivars « SEO » et « Goldrich » (6), les Pakistanais en orange (1), les cultivars locaux (marché d’Almaty) et les variétés anciennes récoltées dans le vignoble kazakh en vert (5), et enfin les kazakhs sauvages ainsi que les chinois de Yilli en rose (4).

Il apparaît ici que les abricotiers chinois se structurent en trois groupes distincts : le germplasm collecté à travers le pays et rassemblé dans la collection de Xiongyue (Apricot national repository) (en violet, Figure 13), les variétés locales de la province du Xinjiang (en bleu) et les abricotiers issus de populations naturelles (Vallée de Yilli et Mehe, en rose). Seul ch_43 (collection de germplasm du Xiongyue) se distingue des autres et groupe avec les variétés locales du Xinjiang, ce qui est compréhensible car il provient de cette province. De plus, il est intéressant de remarquer que les variétés locales du Xinjiang (en bleu, Figure 13), de l’Ouzbékistan et du Kazakhstan de l’ouest (Chymkent, Sayram) sont apparentées entre elles mais distinctes du compartiment sauvage chinois et kazakh (en rose, Figure 13). Ceci démontre une séparation nette entre les individus cultivés et les variétés anciennes d’un côté et les individus sauvages de l’autre.

Afin d’affiner les résultats, les kazakhs sauvages pour lesquels nous disposons de populations naturelles plus fournies ont été analysés séparément. Il ressort 5 populations ancestrales qui correspondent globalement aux différents sites d’échantillonnage (données non présentées). Ceux-ci constituent des populations homogènes, excepté pour celle du lac de Big Almaty et de la vallée de Turgen. Dans ce dernier cas, deux sous-populations se dessinent selon un gradient altitudinal, contrairement à la vallée du Lac de Big Almaty qui a été échantillonnée de manière plus intensive durant l’été 2014 afin de comprendre l’origine de cette structuration intra-population.
**Tableau 5 :** Estimation à partir des données de génotypage par microsatellites du flux de gènes (Nm) entre matériel provenant des différents pays (SSR). Calculs effectués avec GENETIX. USA correspond à des accessions de différentes origines conservées en collection, cultivé correspond aux cultivars utilisés en référence dans l’analyse.

<table>
<thead>
<tr>
<th></th>
<th>Azerbaïdjan cultivé</th>
<th>Chine</th>
<th>Kazakhstan</th>
<th>Ouzbékistan</th>
<th>Turquie</th>
<th>USA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canyon Belbulak</td>
<td>4.79</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Canyon Kaskelen</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Canyon Zhabagly</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Big Almaty Lake</td>
<td>2.15</td>
<td>2.32</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Big Almaty Lake</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lake Esik Lake</td>
<td>1.73</td>
<td>1.86</td>
<td>11.73</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lake Ak-Kain</td>
<td>1.47</td>
<td>1.69</td>
<td>16.36</td>
<td>19.98</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medeu Valley</td>
<td>25.51</td>
<td>3.94</td>
<td>2.00</td>
<td>1.65</td>
<td>1.51</td>
<td></td>
</tr>
<tr>
<td>Medeu Valley</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Turgen Valley</td>
<td>3.67</td>
<td>3.63</td>
<td>8.52</td>
<td>5.01</td>
<td>6.32</td>
<td>2.91</td>
</tr>
<tr>
<td>Turgen Valley</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>USA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Tableau 6 :** Estimation du flux de gènes (Nm) entre les différents sites d’échantillonnage (populations naturelles d’abricotiers sauvages uniquement) du Kazakhstan (données issues des SSR). X signifie que la valeur de Fst pour cette paire est négative et que le calcul n’est donc pas possible.

<table>
<thead>
<tr>
<th></th>
<th>Canyon</th>
<th>Big Almaty</th>
<th>Medeu</th>
<th>Turgen</th>
<th>Canyon</th>
<th>Aksu</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canyon Belbulak</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Canyon Kaskelen</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Canyon Zhabagly</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Belbulak Lake</td>
<td>7.79</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Big Almaty Lake</td>
<td></td>
<td>5.14</td>
<td>7.59</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lake Esik Lake</td>
<td>9.76</td>
<td></td>
<td>5.79</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lake Ak-Kain</td>
<td>10.58</td>
<td>13.38</td>
<td>5.01</td>
<td>77.99</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medeu Valley</td>
<td>1.8</td>
<td>3.2</td>
<td>1.93</td>
<td>2.91</td>
<td>1.72</td>
<td></td>
</tr>
<tr>
<td>Medeu Valley</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Turgen Valley</td>
<td>8.1</td>
<td>151.23</td>
<td>6.67</td>
<td>11.86</td>
<td>27.51</td>
<td>3.12</td>
</tr>
<tr>
<td>Turgen Valley</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>USA</td>
<td>11.74</td>
<td>11.99</td>
<td>4.01</td>
<td>10.32</td>
<td>15.77</td>
<td>2.3</td>
</tr>
<tr>
<td>USA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Tableau 7 :** Flux de gènes (Nm) entre les 6 populations détectées par STRUCTURE (données issues des SSR). Les couleurs renvoient aux couleurs utilisées sur la Figure 13 pour chaque population

<table>
<thead>
<tr>
<th></th>
<th>jaune</th>
<th>violet</th>
<th>bleu</th>
<th>rose</th>
<th>vert</th>
<th>orange</th>
</tr>
</thead>
<tbody>
<tr>
<td>jaune</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>violet</td>
<td>2.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bleu</td>
<td>2.43</td>
<td>3.64</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rose</td>
<td>1.32</td>
<td>2.81</td>
<td>5.94</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vert</td>
<td>1.27</td>
<td>2.64</td>
<td>3.04</td>
<td>3.77</td>
<td></td>
<td></td>
</tr>
<tr>
<td>orange</td>
<td>1.84</td>
<td>2.79</td>
<td>4.6</td>
<td>3.17</td>
<td>2.59</td>
<td></td>
</tr>
</tbody>
</table>

**Tableau 8 :** Diversité génétique entre les pays (données issues des SSR). He=hétérozygotie attendue, Ho=hétérozygotie observée, Ar=richesse allélique, Apr=richesse allélique privée.

<table>
<thead>
<tr>
<th></th>
<th>He</th>
<th>Ho</th>
<th>Ar</th>
<th>Apr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Azerbaïdjan</td>
<td>0.6209</td>
<td>0.5865</td>
<td>3.1204</td>
<td>0.22159</td>
</tr>
<tr>
<td>Cultivars de référence</td>
<td>0.6249</td>
<td>0.5694</td>
<td>3.1470</td>
<td>0.30084</td>
</tr>
<tr>
<td>Chine</td>
<td>0.8</td>
<td>0.6566</td>
<td>4.0862</td>
<td>0.84307</td>
</tr>
<tr>
<td>Kazakhstan</td>
<td>0.7903</td>
<td>0.687</td>
<td>3.8760</td>
<td>0.61964</td>
</tr>
<tr>
<td>Ouzbékistan</td>
<td>0.7571</td>
<td>0.7392</td>
<td>4.0149</td>
<td>0.68371</td>
</tr>
<tr>
<td>Turquie</td>
<td>0.6278</td>
<td>0.6075</td>
<td>3.0950</td>
<td>0.189993</td>
</tr>
<tr>
<td>USA</td>
<td>0.7542</td>
<td>0.6598</td>
<td>3.7685</td>
<td>0.49492</td>
</tr>
</tbody>
</table>
La même analyse a été réalisée avec le logiciel GENELAND. Les résultats sont différents puisqu'elles 4 populations sont détectées, mais les mêmes tendances ressortent. De plus, les résultats des deux logiciels restent similaires par la suite. Nous ne présenterons donc que les résultats obtenus avec STRUCTURE.

Ces analyses de structuration issues de 21 loci microsatellites apportent des informations différentes de celles obtenues par phénotypage du débourrement. Nous observons des populations naturelles bien distinctes suivant leur structuration géographique, ce qui ne fut pas le cas lors de l’analyse phénotypique. Ceci s’explique soit par un manque de variabilité du caractère « débourrement », soit par le fait que la majorité des individus génotypés sont issus des pieds-mères (feuilles) alors que ceux qui ont été phénotypés sont des semis d’une partie seulement de ces pieds-mères. Les mêmes analyses ont donc été répétées uniquement sur les individus communs au phénotypage et aux marqueurs SSR. Cependant, une seule population est obtenue, ce qui pourrait s’expliquer par le faible nombre d’individus restants (87 au lieu de 301 pour les SSR et 338 pour le phénotypage).

**Distance génétique et flux de gènes entre populations**

Le calcul du nombre de migrants par génération (valeur Nm=(1-Fst)/4Fst) permet d’estimer les flux de gènes entre les populations. Ainsi, le flux de gènes entre pays varie de 1,47 à 25,51 (Tableau 5). Le matériel turc présente un flux de gènes non négligeable avec le matériel azéri ; les matériaux chinois, kazakh et ouzbek présentent également des échanges importants entre eux ; et le matériel issu des collections américaines, des échanges moyens avec tous les autres pays. Il est à remarquer ici que ce matériel maintenu par l’ARS-USDA (Davis, CA, USA) fut collecté en Asie centrale mais dans sa partie sud (Turkmenistan, Pakistan etc.), ce qui pourrait désigner un groupe d’abricotiers distincts de ceux d’Asie centrale pour sa partie nord (zone du massif montagneux du Tian Shan). Le calcul de la distance génétique de Nei (1978) amène à des résultats similaires et n’est donc pas présenté ici.

Le Tableau 6 présente les calculs de Nm entre les différents sites d’échantillonnage. La vallée de Turgen échange peu de gènes avec les autres populations, certainement du fait de sa situation excentrée (Figure 6). Au contraire, Ak-Kain et la vallée de Medeu ont des échanges importants, ce qui est lié au fait qu’Ak-Kain se trouve à l’entrée de la vallée. Par contre, les flux de gènes particulièrement importants entre le canyon de Kaskelen, le lac Esik et la vallée de Medeu ne peuvent pas être reliés à une proximité géographique puisqu’une moyenne de 56 km les sépare. De même, Aksu Zhabagly présente des flux moyens avec l’ensemble des populations d’Almaty (à part Big Almaty et Turgen) bien que ce site soit situé à l’extrême ouest du pays (529 km d’Almaty, est du pays). Il serait intéressant de tester l’hypothèse que les allèles partagés entre toutes ces populations appartiennent à une population ancestrale d’abricotiers, spécifique aux abricotiers sauvages de la région du Tian Shan. En comparaison, les flux de gènes entre wine-yard (variétés anciennes) et les populations sauvages sont compris entre 2,17 et 3,78, ce qui semble indiquer un flux de gènes relativement faible entre les compartiments « variétés anciennes » et sauvage. Durant l’été 2014, un échantillonnage un peu plus intensif de ces variétés anciennes kazakhs (plantation le long des routes, jardins privés etc.) a été réalisé afin de vérifier ce résultat. Les flux de gènes entre les différents groupes créés par STRUCTURE (K=6) sont, quant à eux, moyens (Tableau 7), sauf avec le groupe jaune où le flux est relativement plus faible.

L’hétérozygotie attendue (He) pour le groupe jaune (variétés anciennes du Caucase et d’Asie centrale) est plus faible que celle des autres groupes (0,66 contre 0,75). Au niveau de
### Tableau 9 : Estimateurs de diversité globaux résumés par marqueur SSR. He=hétérozygotie attendue, Ho=hétérozygotie observée, Na= nombre d’allèles présents.

<table>
<thead>
<tr>
<th>Marqueur</th>
<th>He</th>
<th>Ho</th>
<th>Na</th>
<th>Fis</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMPA100_1</td>
<td>0.8576</td>
<td>0.7711</td>
<td>12</td>
<td>0.103</td>
</tr>
<tr>
<td>AMPA101_1</td>
<td>0.841</td>
<td>0.7264</td>
<td>18</td>
<td>0.139</td>
</tr>
<tr>
<td>AMPA109_1</td>
<td>0.6937</td>
<td>0.2772</td>
<td>16</td>
<td>0.602</td>
</tr>
<tr>
<td>AMPA116_1</td>
<td>0.8024</td>
<td>0.6765</td>
<td>13</td>
<td>0.159</td>
</tr>
<tr>
<td>AMPA119_1</td>
<td>0.6457</td>
<td>0.5377</td>
<td>12</td>
<td>0.17</td>
</tr>
<tr>
<td>BPPCT001_1</td>
<td>0.5624</td>
<td>0.4439</td>
<td>4</td>
<td>0.213</td>
</tr>
<tr>
<td>BPPCT004_1</td>
<td>0.8756</td>
<td>0.7304</td>
<td>18</td>
<td>0.168</td>
</tr>
<tr>
<td>BPPCT008_1</td>
<td>0.8043</td>
<td>0.6616</td>
<td>18</td>
<td>0.18</td>
</tr>
<tr>
<td>BPPCT017_1</td>
<td>0.6742</td>
<td>0.6289</td>
<td>17</td>
<td>0.07</td>
</tr>
<tr>
<td>BPPCT025_1</td>
<td>0.7093</td>
<td>0.5911</td>
<td>6</td>
<td>0.169</td>
</tr>
<tr>
<td>BPPCT030_1</td>
<td>0.8223</td>
<td>0.7415</td>
<td>10</td>
<td>0.101</td>
</tr>
<tr>
<td>BPPCT038_1</td>
<td>0.8698</td>
<td>0.802</td>
<td>19</td>
<td>0.08</td>
</tr>
<tr>
<td>BPPCT040_1</td>
<td>0.7177</td>
<td>0.5902</td>
<td>13</td>
<td>0.18</td>
</tr>
<tr>
<td>CCPCT030_1</td>
<td>0.915</td>
<td>0.7869</td>
<td>19</td>
<td>0.143</td>
</tr>
<tr>
<td>CPPCT006_1</td>
<td>0.8671</td>
<td>0.8049</td>
<td>14</td>
<td>0.074</td>
</tr>
<tr>
<td>CPPCT022_1</td>
<td>0.91</td>
<td>0.801</td>
<td>21</td>
<td>0.122</td>
</tr>
<tr>
<td>CPPCT033_1</td>
<td>0.8264</td>
<td>0.5678</td>
<td>15</td>
<td>0.315</td>
</tr>
<tr>
<td>PGS1_21_1</td>
<td>0.9021</td>
<td>0.8069</td>
<td>25</td>
<td>0.108</td>
</tr>
<tr>
<td>UDP97-402_1</td>
<td>0.871</td>
<td>0.4315</td>
<td>15</td>
<td>0.506</td>
</tr>
<tr>
<td>UDP98-409_1</td>
<td>0.8942</td>
<td>0.803</td>
<td>19</td>
<td>0.105</td>
</tr>
<tr>
<td>UDP98-412_1</td>
<td>0.8769</td>
<td>0.7427</td>
<td>15</td>
<td>0.155</td>
</tr>
<tr>
<td>Moyenne</td>
<td>0.81</td>
<td>0.66</td>
<td>15.19</td>
<td>0.18</td>
</tr>
</tbody>
</table>

### Tableau 10 : Estimateurs de diversité de Weir et Cockerman calculés par GENEPOP sur les populations détectées par STRUCTURE (données SSR). Ici, aucun locus ne présente d’estimateur négatif. Pour chaque estimateur, les valeurs associées à des lettres différentes sont significativement différentes au seuil de 5%.

<table>
<thead>
<tr>
<th>Estimateur</th>
<th>Valeur (groupes par pays)</th>
<th>Valeur (K=2)</th>
<th>Valeur (K=6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FIS</td>
<td>0.1358 a</td>
<td>0.1403 a</td>
<td>0.1108 b</td>
</tr>
<tr>
<td>FST</td>
<td>0.0771 a</td>
<td>0.1079 b</td>
<td>0.1043 b</td>
</tr>
<tr>
<td>FIT</td>
<td>0.2024 a</td>
<td>0.2331 b</td>
<td>0.2035 a</td>
</tr>
</tbody>
</table>

### Tableau 11 : Estimateurs de diversité de Weir et Cockerman calculés par GENEPOP sur les populations détectées au sein des kazakhs sauvages (données SSR) . Ici, de nombreux loci présentent un ou deux estimateurs négatifs. Pour chaque estimateur, les valeurs associées à des lettres différentes sont significativement différentes au seuil de 5%.

<table>
<thead>
<tr>
<th>Estimateur</th>
<th>Valeur</th>
<th>Valeur (Big Almaty Lake)</th>
<th>Valeur (Turgen Valley)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FIS</td>
<td>0.0502 a</td>
<td>0.0302 a</td>
<td>-0.2089 b</td>
</tr>
<tr>
<td>FST</td>
<td>0.0464 a</td>
<td>0.0669 a</td>
<td>0.1883 b</td>
</tr>
<tr>
<td>FIT</td>
<td>0.0943 a</td>
<td>0.0951 a</td>
<td>0.0187 a</td>
</tr>
</tbody>
</table>
la richesse allélique, les deux populations majoritairement chinoises montrent une richesse allélique plus élevée (5,89 contre 4,89 pour G=12). En effet, ces deux groupes associés au cluster rose contiennent en moyenne 94% des allèles présents tous individus confondus. Au sein des kazakhs sauvages et des variétés anciennes de wine yard, l’hétérozygotie attendue et la richesse allélique sont proches pour toutes les populations (moyenne de 0,68 et 2,31 pour G=3 respectivement). Enfin, cette analyse (Tableau 8) montre clairement que les accessions turques et azéries ainsi que les cultivars de référence ont une diversité plus faible que tous les autres abricotiers issus d’Asie centrale (partie nord, zone du Tian Shan), et que les accessions en collection aux Etats-Unis (origine Asie centrale, partie sud) sont intermédiaires.

**Différenciation des populations**

Les estimateurs de diversité calculés sur les marqueurs SSR dans les différentes subdivisions sont résumés dans les Tableau 10 et Tableau 11. Les Fst pairés ne sont pas présentés car ils montrent les mêmes relations entre les populations que le calcul de Nm. Il apparaît un excès d’homozygotes dans toutes les populations, excepté au sein de la population de la vallée de Turgen. Afin d'étudier le lien entre la distance génétique et la distance géographique des populations, un test de Mantel a été réalisé. Il en ressort qu'il y a un lien entre ces variables pour les populations détectées par STRUCTURE (K=6) sur les SSR (p=0,01), mais pas pour les populations sauvages kazakhs (p=0,97), ni entre les pays (p=0,37).

**Phylogénie**

L’analyse phylogénétique réalisée (Figure 14) révèle que les groupes détectés par STRUCTURE sont globalement liés à des relations de parenté. Certains clusters appartiennent à plusieurs groupes phylogénétiques qui reflètent le type d’accessions (d’un côté les variétés locales et de l’autre les autres en ce qui concerne le groupe vert par exemple). Il y a ainsi une séparation nette et relativement ancienne entre les compartiments cultivé et sauvage.

**Synthèse des résultats obtenus lors de l’analyse de la diversité génétique au niveau de 21 loci microsatellites**

L’étude des microsatellites a permis de montrer que : 1) les accessions azéries et turques sont très proches du matériel cultivé, moins diverses et significativement distinctes du matériel sauvage, 2) le centre d’origine de l’abricotier semble être en Chine du nord qui présente une forte diversité, 3) le compartiment sauvage kazakh est peu structuré, 4) il y a un excès d’homozygotes généralisé, 5) le compartiment cultivé serait en grande partie issu d’une séparation ancienne avec le compartiment sauvage.

Bien que ces 21 loci microsatellites soient distribués sur l’ensemble du génome abricotier (Annexe XI), ils ne reflètent qu’une partie de la diversité génétique de l’abricotier. Nous avons donc réalisé une étude similaire à partir de séquences générées dans le cadre du projet ANR CHEX ABRIWG. Le génome de 77 abricotiers sauvages issus de 7 populations naturelles (Turgen, lac d’Almaty, Aksu-Zhabagly, Shary Chelek, Issik-kul Ortobyrosan, Chuy river Boom Canyon, à raison de 7 à 10 individus par population), essentiellement kazakhs et kirghizes, et quelques variétés locales, anciennes, chinoises et ouzbeks a été reséquencé. Les séquences ont été filtrées selon le protocole décrit dans la partie « matériel et méthodes ». La quantité de reads conservés à chaque étape de l’analyse de séquences et le nombre de SNPs obtenus sont résumés dans la Figure 15. Au final, un total de 207 537 SNPs, de position connue, a été détecté. Ils sont répartis sur l’ensemble du génome et proportionnellement à la taille des pseudo-molécules (ou « scaffolds » qui correspondent aux 8 chromosomes de l’abricotier) (Figure 16).
Figure 14 : Traitement des données de séquençage avant leur utilisation. En bleu sont indiquées les quantités de données conservées après chaque étape, en vert sont indiqués les logiciels utilisés lorsqu’il y a lieu. Dans le cas contraire, des scripts spécifiques ont été utilisés.

Figure 15 : Arbre phylogénétique construit selon la méthode UPGMA avec le logiciel POPTREE à partir des données de SSR. Les individus en rouge correspondent au groupe rose (4), ceux en noir au groupe jaune (3) et les autres couleurs correspondent à celles utilisées pour les groupes détectés par STRUCTURE.

Figure 16 : Répartition des SNPs le long du génome et correspondance avec la taille de chaque scaffold (pseudomolécule).
Diversité génétique globale à partir des données de SNPs (Single Nucleotide Polymorphism) répartis sur l’ensemble du génome

Dans cette partie, nous avons travaillé à partir d’un jeu de 2 076 SNPs sélectionnés parmi les 207 537 SNPs initiaux (Figure 15). En effet, des essais sur des jeux de données moins importants ont montré une perte d’information et avec des données plus conséquentes, aucun gain n’est constaté. De plus, cela permet de limiter le temps de calcul à 4 jours au lieu de quelques semaines. Ces SNPs ont été sélectionnés au hasard selon la règle d’un SNP conservé tous les 100 SNPs en parcourant le génome linéairement.

Estimation de la diversité génétique des abricotiers sauvages d’Asie centrale

La valeur moyenne de théta estimée sur l’ensemble des 207 537 SNPs disponibles et sur les 77 accessions séquencées est de 7,65.10⁻³ avec un écart-type de 1,37.10⁻⁴. Une comparaison de moyennes a révélé une absence de différence significative du niveau de diversité, que ce soit en fonction du pays, du type d’accession ou de la population. Le taux d’erreur moyen est de 1,30.10⁻³, ce qui correspond à un nucléotide toutes les 1000 bases et est faible en regard du taux d’erreur indiqué par la société Illumina (autour de 1%).

En ce qui concerne le Fis calculé pour chaque SNP sur l’ensemble du génome, il est négatif avec une moyenne de -0,318 (Tableau 12). Il existe cependant des régions pour lesquelles le Fis est positif (Annexe XII)

Structuration des populations naturelles d’abricotiers sur la base de 2 076 SNPs

Aucune structuration n’a pu être détectée par le logiciel TESS. Concernant STRUCTURE, une structuration en 3 populations (Figure 17) a été détectée (sans admixture et avec les fréquences alléliques corrélées, avec et sans l’information sur le site d’échantillonnage). Dans tous les autres cas, aucune structuration n’est détectée. Les populations ainsi formées sont les suivantes : Shary Shelek (Kyrgyzstan) (sauf un individu, kr_94_1, seule variété ancienne de ce lieu d’échantillonnage) (jaune, groupe 3), des variétés anciennes (kazakhs, ouzbeks et chinoises) et l’accession sauvage kr_184_1 (sans doute issue d’une contamination du compartiment sauvage par replantation ou flux de gêne important) (bleu, 2) et tous les autres individus sauvages collectés dans les populations naturelles kazakhs et kirghizes complétés par quelques individus sauvages chinois et des variétés anciennes (orange, cluster 1). Aucune structure intra-population n’a été détectée, bien que quelques runs semblent séparer les individus d’Issik-kul du reste des individus sauvages. Cette absence de structure, mis à part le cas spécifique de Sary Chelek, devrait permettre de mener à court terme des études d’association entre la variabilité génétique du matériel sauvage et la résistance à la sharka.

Les flux de gènes observés (Tableau 13) sont assez importants par rapport à ceux obtenus avec les marqueurs SSR et confirment que la population de Shary Chelek est à part, tandis que les deux autres clusters sont plus proches.

Distance génétique et flux de gènes entre populations

Les taux d’hétérozygote observée et la richesse allélique sont comparables pour les trois populations (Tableau 14). L’hétérozygote observée est plus élevée qu’attenda, ce qui montre un excès d’hétérozygotes et donc d’outbreeding dans ces populations naturelles. Ceci est en contradiction avec les résultats obtenus à partir de profils microsatellites.

Différenciation des populations

La valeur de Fst obtenue ici (Tableau 15) confirme l’excès d’hétérozygotes observé précédemment. Lors de la comparaison des valeurs de Fis aux loci proches des marqueurs SSR utilisés (Annexe XI), aucune correspondance ne peut être établie. La valeur de Fst,
Tableau 12 : Fis moyen par scaffold (pseudo-molécule) calculé sur l'ensemble des SNPs. Les lettres indiquent la significativité de la différence à 5% des valeurs.

<table>
<thead>
<tr>
<th>Scaffold</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>Génome entier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fis moyen</td>
<td>0.245 a</td>
<td>0.526 b</td>
<td>0.243 a</td>
<td>0.542 b</td>
<td>0.248 b</td>
<td>0.250 a</td>
<td>0.239 ac</td>
<td>0.253 ad</td>
<td>-0.318</td>
</tr>
</tbody>
</table>

Figure 17 : Structuration en 3 populations détectée par STRUCTURE à partir des données de génotypage de 2 076 SNPs (pas d’admixture, fréquences alléliques corrélées).

Tableau 13 : Flux de gènes (Nm) entre les 3 populations détectées par STRUCTURE (données SNP). Les couleurs renvoient aux couleurs utilisées sur la Figure 13 pour chaque population.

<table>
<thead>
<tr>
<th></th>
<th>orange</th>
<th>bleu</th>
<th>jaune</th>
</tr>
</thead>
<tbody>
<tr>
<td>orange</td>
<td>20.39</td>
<td></td>
<td></td>
</tr>
<tr>
<td>bleu</td>
<td>11.11</td>
<td>11.25</td>
<td></td>
</tr>
<tr>
<td>jaune</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Tableau 14 : Diversité génétique entre les 3 populations détectées par STRUCTURE (données issues des SNPs). He=hétérozygotie attendue, Ho=hétérozygotie observée, Ar=richesse allélique, Apr=richesse allélique privée.

<table>
<thead>
<tr>
<th></th>
<th>He</th>
<th>Ho</th>
<th>Ar</th>
<th>Apr</th>
</tr>
</thead>
<tbody>
<tr>
<td>orange</td>
<td>0.2203</td>
<td>0.3393</td>
<td>1.50711</td>
<td>0.0637743</td>
</tr>
<tr>
<td>bleu</td>
<td>0.2238</td>
<td>0.3449</td>
<td>1.54459</td>
<td>0.0895526</td>
</tr>
<tr>
<td>jaune</td>
<td>0.215</td>
<td>0.3371</td>
<td>1.50202</td>
<td>0.0642369</td>
</tr>
</tbody>
</table>
quant à elle, indique que seul 1,7% de la diversité génétique observée entre les populations est lié à la dérive génétique, ce qui est bien en deçà des valeurs obtenues avec les marqueurs SSR. Les valeurs pairées apportent la même information que le calcul de Nm. Toutefois, 315 SNPs sur les 207 537 initiaux présentent un Fst supérieur à 0,35.

*Phylogénie des accessions séquencées à partir de 2 076 SNPs*

Les relations phylogénétiques entre les individus (Figure 18) reflètent bien les groupes détectés par STRUCTURE, sauf pour deux individus (ch_123 et kr_188_1). Ainsi, le groupe bleu de STRUCTURE (variétés anciennes) se serait différencié très tôt des autres groupes (sauvages).

*Synthèse des résultats obtenus lors de l’analyse de la diversité génétique au niveau des SNPs*

L’analyse des marqueurs SNPs a montré que : 1) la diversité de l’abricotier est assez élevée, 2) la structuration des populations est faible, 3) il y a un excès d’hétérozygotes, 4) la structuration reflète la phylogénie. Certains résultats sont donc en contradiction avec ceux obtenus grâce aux marqueurs SSR.

**DISCUSSION**

Dans cette étude, notre objectif était d’étudier et comparer la diversité génétique des variétés d’abricotiers cultivées en Europe, des abricotiers sauvages et des variétés cultivées auparavant ou/et actuellement dans différents pays situés dans la zone d’origine probable de l’abricotier, à savoir la Chine et l’Asie centrale (Bourguiba et al., 2012). Pour cela, nous avons mené différentes analyses ciblant la variabilité du débourrement végétatif, la variation allélique à deux loci spécifiques d’intérêt agronomique (résistance à la sharka et auto-incompatibilité) et enfin la diversité génétique globale à partir de 21 loci microsatellites et 2 076 loci SNP.

**Des variétés anciennes avec de faibles besoins en froid**

Dans le cas de la variabilité associée au débourrement végétatif, des études statistiques (ACP, analyse de variance, GENELAND) ont été menées. Globalement, elles permettent de distinguer le matériel cultivé européen et local du matériel sauvage. Les variétés anciennes (traditionnelles) et actuelles sont en effet plus précoces que les sauvages. Cela peut être dû au fait que lors de la domestication et/ou lors de la sélection par les villageois d’individus qu’ils ont ensuite transférés dans les jardins, les arbres les plus précoces ont été préférés pour une raison encore inconnue. Ainsi, les variétés anciennes peuvent constituer des géniteurs potentiellement intéressants pour introgresser des besoins en froid plus faibles dans les programmes de sélection, ce qui serait plus adapté pour la production d’abricots en Europe méridionale. Cependant, les analyses de diversité allélique aux loci d’intérêt, tout au moins dans le cas de la résistance à la sharka, ont aussi démontré une quasi-absence d’allèles favorables dans ces variétés cultivées, anciennes ou récentes. Il faudra donc croiser entre elles des accessions sauvages porteuses de résistance à la sharka et des variétés cultivées présentant des besoins en froid plus faibles et auto-fertiles. Cela permettrait de continuer à produire des abricots dans les régions où le réchauffement climatique empêche la satisfaction des besoins en froid des variétés plus tardives. Dans cette optique, les variétés anciennes ouzbeks sont les plus prometteuses, de même que la variété tunisienne « Bakour », très précoce et auto-fertile.

Lors des analyses de structuration en fonction des variations du caractère « débourrement », une hiérarchisation des populations semble se dégager selon la latitude, ce qui serait en lien avec l’évolution des climats, plus chauds et arides au sud et à l’ouest de
Tableau 15 : Estimateurs de diversité de Weir et Cockerman calculés par GENEPOP sur les populations détectées par STRUCTURE (données SNP).

<table>
<thead>
<tr>
<th>Estimateur</th>
<th>Valeur</th>
<th>Estimateur</th>
<th>Valeur</th>
</tr>
</thead>
<tbody>
<tr>
<td>FIS</td>
<td>-0.318</td>
<td>FIT</td>
<td>-0.4979</td>
</tr>
<tr>
<td>FST</td>
<td>0.0168</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 18 : Arbre phylogénétique construit selon la méthode UPGMA avec le logiciel DARwin à partir des données de SNP. Les individus en rouge correspondent au groupe de Shary Chelek (3) détecté par STRUCTURE, en bleu au groupe des ouzbeks et kazakhs de l'ouest, excepté les individus sauvages d'Aksu-Zhabagly (2) et en noir les autres (1).
l’Asie centrale et du Caucase. On peut en effet supposer que les populations exposées à des climats différents sont éloignés dans l’analyse. Cependant, nous ne possédons pas de données climatiques assez précises pour étudier cet aspect et déterminer le lien entre climat et débourrement.

Le fait que 100% de la variabilité soit expliqué avec les deux variables HF (nombre d’heures de froid pour atteindre la stade 3) et RG (rayonnement solaire nécessaire pour atteindre le stade 3) indiquerait que le débourrement est à la fois contrôlé par le froid (Fan et al., 2010), et par la luminosité ou la longueur de la photopériode. Cependant, les études précédentes (Basler et al., 2012 ; Laube et al., 2014 ; Cooke et al., 2012) s’accordent sur le fait que seul le froid est indispensable à la levée de dormance chez les **Prunus**. Il est cependant nécessaire que la température printanière soit également suffisante (Ruiz et al., 2007) pour lever la dormance. Ainsi, la variable RG pourrait refléter l’importance du réchauffement printanier plutôt que la photopériode.

Lors des analyses de variance, nous avons mis en évidence un effet fort du type d’accession sur la variation des trois variables HF, RG et HJ, qui correspondent au cumul des heures de froid, du rayonnement solaire et des heures de jours. Cela signifie que selon si l’on a une variété ancienne ou un individu sauvage, la réponse aux trois variables est différente, par l’interaction RG-HJ. Au contraire, l’absence de cette interaction chez les accessions sauvages semblerait indiquer qu’à partir du moment où les besoins en froid sont satisfaits, atteindre l’un des deux autres seuils est suffisant pour passer au stade 3, ce qui confirmerait que l’allongement de la photopériode n’est pas indispensable (Basler et al., 2012 ; Laube et al., 2014 ; Cooke et al., 2012).

**Le débourrement, une variable échouant à révéler une structuration au sein des abricotiers sauvages.**

L'analyse de la structuration du matériel abricotier sauvage à partir des données de phénotypage avec GENELAND n’a été réalisée que sur une seule variable phénotypique afin d’alléger les temps de calculs. Il est donc possible que les résultats soient légèrement différents si l’ensemble des variables est conservé. Cela permet néanmoins d’apprécier la structuration des populations, d’autant plus qu’une double approche avec les marqueurs moléculaires (SSR puis SNP) est venue compléter cette analyse préliminaire. A partir des valeurs phénotypiques observées, il ressort une nette séparation entre les individus provenant d’Aksu-Zhabagly (ouest du Kazakhstan) et les autres, cette séparation n’étant pas liée au climat ou à l’altitude car d’autres accessions se trouvent dans des zones aux caractéristiques similaires (la province autonome du Naxçivan (Azerbaïdjan), par exemple, ou les individus ouzbeks, à la différence près qu’il s’agit dans ce cas de matériel sélectionné ou cultivé et non sauvage). Ainsi, si seuls les individus sauvages sont pris en considération, le fait qu’Aksu-Zhabagly soit à part est à mettre en relation avec son climat chaud et aride par rapport aux autres kazakhs sauvages. Cette séparation pourrait aussi venir du phénomène d’isolation par distance entre populations naturelles puisque le parc national d’Aksu-Zhabagly se situe très à l’ouest par rapport aux autres sites d’échantillonnage du Kazakhstan (529 km d’Almaty).

**Un potentiel des populations sauvages pour lutter contre la sharka**

Des études de variabilité à deux loci d’intérêt agronomique, **PPVres** (pour la résistance à la sharka) et **SI** (pour l’auto-incompatibilité), ont été menées. Les comparaisons de moyennes réalisées sur les allèles du marqueur ZP002, lié à **PPVres**, ont montré que les variétés anciennes chinoises et les populations naturelles d’abricotiers du Kazakhstan (plus particulièrement celle du lac Esik) présentent une fréquence plus élevée de l’allèle lié à la résistance. Zhebentayeva et al. (2008) avait obtenu des résultats similaires à partir de
marqueurs AFLP (Amplified Fragment Lenght Polymorphism) et SSR qui indiquaient alors une origine probable de la résistance en Chine ou en Asie centrale. Ainsi, il apparaît très probable que ces populations naturelles constituent des réservoirs intéressants pour la résistance à la sharka, en vue de son introgression dans les variétés cultivées. Cependant, une récente étude de Decroocq et al. (2014) a montré que la présence de l’allèle favorable de ZP002 (ou même des autres marqueurs de PPVres) n’implique pas systématiquement un phénotype résistant. Au travers d’analyses menées sur des populations en ségrégation (F1, F2, rétrocroisement), il semblerait que la résistance au virus de la sharka requiert deux loci distincts, dont PPVres (Decroocq et al, 2014). Des analyses d’association au niveau du génome complet de 76 cultivars d’abricotiers ont été conduites au laboratoire de Bordeaux dans le cadre du projet ANR ABRIWG. Elles confirment l’implication de deux loci distincts, contrôlant la réponse au virus, tous deux étant cartographiés sur le chromosome 1, à 10-20 cM de distance (article en cours de préparation). Ces deux loci correspondent aux QTLs, Meta-PPV1a et Meta-PPV1b, mis en évidence lors d’une méta-analyse par Marandel et al., (2009). Afin de confirmer le statut de réservoir des populations naturelles d’abricotiers, il conviendra donc d’étudier l’haplotype à ce deuxième locus dans le compartiment sauvage et de l’associer à la résistance ou sensibilité au virus de la sharka, en combinaison avec les variations génotypiques au locus PPVres. Ces travaux sont en cours au laboratoire mais requièrent au moins deux ans de phénotypage intensif avant de conclure sur cette hypothèse.

Une mutation « récente » à l’origine de l’auto-compatibilité ?

Chez les abricotiers cultivés, une majorité de l’auto-compatibilité observée est liée à une insertion dans la phase codante du gène SFB, au locus d’auto-incompatibilité SI (Romero et al., 2004). Au début de ce stage, nous avions émis l’hypothèse que cet événement d’insertion aurait eu lieu il y a très longtemps et que l’insertion serait donc présente dans le pool sauvage. 321 accèssions provenant du pool sauvage, de variétés locales et cultivées en Turquie, dans le Caucase et en Asie Centrale (à l’exception des kirghizs, non-disponibles à ce moment) ont été testées avec un marqueur ciblant cette insertion. Finalement, seule quelques variétés anciennes ou locales ainsi qu’un individu conservé dans des collections botaniques américaines se sont révélées positives. Il apparaît donc que cette mutation du gène SFB n’existe pas dans le pool sauvage et qu’elle serait apparue plus tard lors de la domestication et ensuite sélectionnée dans les programmes d’amélioration de l’abricotier cultivé. Sa présence dans certaines variétés locales ou matériel conservé en collection pourrait indiquer que, soit l’auto-compatibilité a été introduite plus récemment dans le pool local, soit qu’elle est présente dans le pool sauvage mais pas dans les échantillons testés durant ce stage. Une autre possibilité évoquée par Delplancke et al. (2011) pourrait être que ce caractère a été transféré spontanément depuis une espèce proche, comme par exemple le pêcher. Cependant, aucune espèce apparentée pouvant se croiser avec l’abricotier n’est auto-compatible, cela est donc peu probable chez cette espèce. Il serait intéressant de récupérer du matériel un peu plus au sud du bassin d’origine, au Tadjikistan, Pakistan et Afghanistan et d’intégrer les espèces proches dans l’étude mais l’accès à ce matériel demeure difficile pour des raisons géopolitiques.

Une proximité des abricotiers issus du Caucase avec le matériel cultivé européen révélée par les marqueurs SSR

A partir d’une partie de ce même matériel végétal abricotier pour laquelle nous disposions déjà des ADN, une analyse des profils de 21 marqueurs microsatellites a été réalisée. Lors de cette étude, nous avons observé que les populations azérie et turque sont apparentées et très proches du matériel cultivé, européen (plus précisément, méditerranéen, incluant « Bakour », « Bergeron », « Canino », « Currot », « Goldrich », « Luizet », « Moniqui »), ce qui
s’explique par le fort lien historique entre ces deux pays, l’Azerbaïdjan et la Turquie étant tous de civilisations turcophones. De plus, ils correspondent, avec l’Iran, à une zone de domestication, très certainement secondaire, de l’espèce (Bourguiba et al., 2012). Pour cette raison, ils ont fortement contribué à la diffusion de l’abricotier dans le pourtour méditerranéen, d’où l’apparentement avec les variétés d’Europe méditerranéenne mais pas avec « SEO ». Ceci confirme également le statut particulier de ce cultivar, plus proche du matériel chinois que des cultivars européens.

Des flux de gènes entre compartiments sauvage et cultivé

Les flux de gènes entre les pays permettent d’évaluer la hauteur des échanges de matériel entre les pays, notamment pour les variétés anciennes et cultivées. En effet, les forts flux de gènes (valeur supérieure à 10) peuvent être associés à des échanges commerciaux de matériel comme cela se fait traditionnellement entre la Turquie et l’Azerbaïdjan. Suivant ce critère, il y aurait également des échanges entre la Chine et le Kazakhstan, entre la Chine et l’Ouzbékistan et le Kazakhstan et l’Ouzbékistan, bien que de moindre envergure. Il est raisonnable de penser que si les données pour le Kirghizstan étaient également disponibles, ce pays apparaîtrait également avec un fort flux de gènes avec les pays voisins.

Au sein du matériel sauvage kazakh, les flux de gènes moyens entre la population naturelle d’Aksu-Zhabagly et celles situées à plus de 500 km à l’ouest peuvent s’expliquer par la replantation d’arbres d’un site à l’autre, bien que cela ne semblait pas être le cas lors de l’échantillonnage. La même remarque s’applique pour le canyon de Kaskelen qui correspond dans ce cas effectivement à des transferts de semis d’autres populations naturelles vers ce site. Les faibles flux de gènes entre le matériel cultivé localement (wine-yard) et le compartiment sauvage confirment les analyses de structuration. Ceci pourrait s’expliquer par le fait que ces arbres ont une origine géographique plus distante et encore inconnue et ne semblent pas provenir ou avoir échangé avec le compartiment sauvage le plus proche (vallée de Turgen). Ils présentent très certainement un caractère apprécié de la population locale (fruits en forme de cœur et très sucrés) qui les a donc introduits là.

L’intégration du cultivar nord-américain « SEO » à un groupe chinois semble confirmer que les variétés nord-américaines sont issues de croisement entre cultivars européens et matériel chinois (Zebhentayieva et al., 2008 ; Li et al., 2014) mais dans ce cas, de Chine Occidentale (Xinjiang), aux limites de l’Asie centrale, et non du nord de la Chine comme l’ont proposé Zebhentayieva et al. De plus, la portion de génome cultivé observée chez les individus chinois pour K=2 avec le logiciel STRUCTURE semble confirmer que les flux de gènes entre compartiment sauvage et cultivé sont plus importants en Chine qu’ailleurs en Asie centrale. Dans tous les cas et tous les pays, le matériel cultivé forme un compartiment à part, distinct des variétés anciennes et des individus sauvages.

Des populations sauvages isolées par distance ?

Lors de l’analyse de structuration des populations naturelles, nous avons remarqué une répartition reflétant plus ou moins bien la géographie des différents sites d’échantillonnage, ce qui semble indiquer un isolement par la distance (IBD) de la plupart de ces populations sauvages. Cependant, le test de Mantel dément cette conclusion puisque la variabilité génétique n’est pas structurée selon la localisation géographique, ce qui va à l’encontre du phénomène d’isolation par distance (Diniz-Filho, 2013). De plus, seules deux populations présentent une sous-structuration qui ne peut s’expliquer par une IBD : la vallée de Turgen et celle du grand lac d’Almaty (Big Almaty lake). Dans le premier cas, cette sous-structuration a pu être reliée à une distribution altitudinale mais pas dans le cas du grand lac d’Almaty. Ceci pourrait être dû à des replantations ou à une situation à mi-chemin entre deux populations.
ancestrales. Un échantillonnage plus intensif a été réalisé en 2014 et a montré l’existence de deux vallées distinctes au niveau de ce site d’échantillonnage. Une vingtaine d’individus ont donc été prélevés dans chaque vallée, ils seront testés avec les mêmes marqueurs SSR afin de vérifier l’existence d’une ou de deux populations distinctes entre les deux vallées.

Des études de différentiation ont également été réalisées à partir des données des marqueurs microsatellites. La faible hétérozygotie attendue chez les abricotiers azéris, turques et cultivés (européens) par rapport aux autres accessions d’Asie centrale démontre ici un goulot d’étranglement, lors de la domestication, comme observé précédemment chez le pommier (Cornille et al., 2012). De plus, la forte richesse allélique des populations chinoises tend à confirmer que la Chine est le centre d’origine de l’abricotier (Li et al., 2014), ou tout au moins participe de manière significative à sa diversité. Cependant, ces données sont issues d’un d’échantillonnage très disparate entre les pays, avec très peu d’individus collectés en Chine (province du Xinjiang) pour cause de refus des autorités locales. Il conviendra donc à l’avenir à rééquilibrer cet échantillonnage ou, à défaut, de relativiser ces conclusions. Enfin, il est à noter l’absence de goulot d’étranglement parmi les populations naturelles d’abricotiers sauvages, quelle que soit leur localisation.

**Des populations sauvages peu structurées**

L’excès d’homozygotes observé (Fis positif) avec les marqueurs SSR peut paraître surprenant puisque les abricotiers sauvages sont auto-incompatibles. Cependant, les logiciels utilisés se basent sur l’hypothèse que la structuration des populations entraîne un effet Wahlund (excès d’homozygotes par rapport à l’équilibre de Hardy-Weinberg) (Francois et Durand 2010). Il est donc logique d’observer ici un excès d’huigotie. Cependant, les valeurs de Fis sont plus élevées lorsque le compartiment cultivé est inclus. Ainsi, cela reflète le fait que le compartiment cultivé contient des individus auto-compatibles et est propagé par greffage, ce qui n’est pas le cas des individus sauvages, qui présentent un Fis proche de 0 ou négatif. En effet, Bourguiba et al. (2012) avait obtenu un Fis positif au sein du matériel abricotier cultivé, ce qui va également dans ce sens.

Cornille et al. (2012) indique qu’un Fis faible est un indicateur d’une faible structuration des populations. En effet, les populations kazakhs (celles pour lesquelles nous disposions de plus d’individus et d’allèles) ne forment pas un ensemble très structuré parmi toutes les populations échantillonnées, à part la structuration en lien avec leur distribution géographique. Ce lien avec la situation géographique avait déjà été mis en évidence par Li et al. (2010) sur le soja. D’un autre côté, le fait que la population du lac de Big Almaty présente deux groupes distincts mais non-associés à une structuration géographique ou altitudinale et que ceci soit associé à la très faible valeur de Fis observée, laisse penser que cette région contenait autrefois deux populations bien distinctes mais qui se mélangent pour finir par ne former qu’une seule et même population. Si l’hypothèse de deux populations distinctes entre les deux vallées échantillonnées au cours de l’été se confirme, cela montrerait l’existence d’un point de contact entre ces deux vallées qui restent malgré tout isolées. Ces hypothèses semblent soutenues par l’absence de distinction des populations de Big Almaty lake par rapport aux autres populations avec les marqueurs SNPs. Au contraire, dans la vallée de Turgen, les deux populations restent bien séparées selon l’altitude et présentent un fort Fis négatif, basé sur les données de profil microsatellites. Dans ce cas, nous pouvons poser l’hypothèse que ces populations se sont adaptées à deux microclimats distincts présents dans la vallée de Turgen, mais il faudrait croiser ces résultats avec des données pédo-climatiques, pas toujours disponibles dans ces régions. De plus, cette structuration en deux sous-populations n’est pas mise en évidence avec les marqueurs SNP. Toutefois, cette absence de subdivision en deux
sous-populations à partir des données de reséquençage ne signifie pas qu’elle n’existe pas. En effet, Li et al. (2010) et Van Inghelandt (2010) ont mis en évidence que le pouvoir de détection de la structuration à partir de SNP est inférieur à celui issu des SSR. Ceci est probablement à mettre en lien avec le nombre d’allèles qui peut être observé par locus : 4 pour les SNPs (A,C,T,G) et l’infini pour les microsatellites (taille d’allèles). De même, si nous avions eu les capacités d’utiliser un plus grand nombre de SNP, nous aurions peut-être observé cette même division en deux groupes. Ceci sera testé à l’avenir, à partir du jeu de 207 537 SNPs obtenus.

Li et al. (2014) ont étudié la diversité de l’abricot sibérien (*Prunus sibirica* L.) sauvage à partir des profils de 30 marqueurs ISSR (Inter Simple Sequence Repeat), SSR et SRAP (Sequence Related Amplified Polymorphism). Ils ont obtenu un Fst de 0,15 avec les marqueurs SSR, ce qui est sensiblement supérieur à celui que nous avons observé (0,1107) chez *Prunus armeniaca* (sp vulgaris). Ceci indique une diversité moins importante chez *Prunus armeniaca*, d’autant plus que, sans les populations turque et azérie ou en ne conservant que les populations kazakhs, le Fst est drastiquement réduit (0,0335). Ainsi les populations sauvages d’Asie centrale de l’espèce *Prunus armeniaca* semblent peu différenciées. En comparaison, *P. sibirica* est une espèce exclusivement sauvage, au contraire de *P. armeniaca*, ce qui pourrait également expliquer ces différences de différenciation. De plus, nous avons également montré que toute la diversité attendue de *P. armeniaca* n’est pas présente dans cette étude comme semblent l’indiquer les données obtenues sur matériel chinois ou sur individus originaires de la partie sud d’Asie centrale (Turkmenistan, Pakistan). En effet, ces accessions présentent des allèles non détectés dans les populations naturelles du nord de l’Asie centrale (Kazakhstan, Ouzbékistan et partie nord du Kirghizstan). De plus, le statut à part de la population naturelle de Sary Chelek, située plus au sud du Kirghizstan et qui ne groupe pas avec les autres populations kazakhs et kirghizes confirmerait l’existence d’un pool génétique particulier plus au sud. Ce pool se serait partiellement mélangé avec le matériel du nord de l’Asie centrale au niveau de la province du Xinjiang, en Chine comme le montrent les analyses de structuration dans cette région. Ceci est assez probable puisque le Xinjiang partage une frontière commune avec le Pakistan et une histoire culturelle non-négligeable, notamment au niveau du district de Kasgar (sud-ouest du Xinjiang). Il est donc raisonnable de penser que le Fst obtenu pour l’espèce *Prunus armeniaca* sp vulgaris est ici sous-estimé. Néanmoins, cette faible structuration génétique indique que l’abricotier présente une forte capacité de dispersion comme ceci fut observé chez le pommier (Cornille et al., 2013a).

D’autres études menées sur *P. armeniaca* à partir de profils SSR ont également conduit à calculer le Fst : 0,137 dans la vallée de Yili (Chine, province du Xinjiang) (Tian Ming et al., 2007), 0,086 pour des accessions issues essentiellement d’Europe, des USA et du Maghreb (Lasnier 2013), 0,5768 sur des variétés cultivées (Maghuly et al., 2005), 0,111 sur des cultivars provenant du bassin méditerranéen (Bourguiba et al., 2012). Les résultats obtenus sont extrêmement variables, même si notre étude semble aller dans le sens des résultats de Bourguiba et al. (2012) et Tian Ming et al. (2007). Il est à noter que les résultats présentés par Maghuly et al (2005) sont en partie faussés par l’ajout dans l’analyse d’individus interspécifiques prunier x abricotier (*Prunus cerasifera* et *Prunus salicina*) et d’une espèce apparentée (*Prunus brigantiana*).

Une faible structuration des populations observée avec les marqueurs SNP

Lors de la recherche de structuration au sein des individus à partir des marqueurs SNP, un seul modèle a pu détecter une structuration. Il est donc raisonnable de penser que la
différenciation des populations est faible. En effet, la faible valeur de Fst obtenue (0,0168) va également dans ce sens. Ceci constitue un avantage non-négligeable pour les futures analyses d’association au niveau du génome complet d’abricotiers sauvages, notamment pour la résistance à la sharka. Toutefois, les valeurs élevées de Fst obtenues sur 315 des 2 076 SNPs pourraient indiquer la présence de gènes sous sélection divergente (Sousa et al., 2013).

L’analyse de la diversité des populations détectées par STRUCTURE avec les marqueurs SNP a montré un excès d’hétérozygotes par rapport à l’équilibre de Hardy-Weinberg. Cela peut-être relié au fait que la mutation entrainant l’auto-compatibilité n’a pu être détectée dans les populations sauvages, le phénomène d’auto-incompatibilité favorisant les hétérozygotes par rapport aux homozygotes. De plus ces populations sont plus soumises à la sélection naturelle qui élimine les individus les plus homozygotes, conduisant à un excès d’hétérozygotes dans les populations plus âgées (Stoeckel et al., 2006). Les populations ont montré une faible différenciation, ce qui rejette les résultats obtenus avec les marqueurs SSR qui montrent une faible structuration des populations.

Une diversité importante révélée par les SNP

La diversité nucléotidique moyenne (π), qui constitue un estimateur de Θ, est de 1,5.10^{-3} chez le pêcher (Verde et al, 2013), il semble donc que la diversité présente chez l’abricotier sauvage est bien plus importante puisque la valeur calculée ici est de 7,65.10^{-3}. Cependant, il s’agit, dans le cas du pêcher, uniquement du compartiment cultivé. Lam et al (2010) ont montré une diversité plus faible que chez le soja sauvage (1,9.10^{-3}) mais dont la forme sauvage a disparu, cette observation peut être vue comme similaire à celle du soja. Cependant, une étude réalisée sur les abricotiers cultivés dans le cadre du projet ANR ABRIWG, a obtenu une valeur de Θ de 7,97.10^{-3} (article en cours de rédaction), similaire à celle que nous avons obtenue sur le compartiment sauvage. L’observation de Lam et al. (2010) n’est donc pas valable dans le cas de l’abricotier. Toutefois, le niveau de diversité de l’abricotier reste élevé en regard d’autres espèces végétales, comme par exemple 3,7.10^{-3} pour les espèces du genre Malus (Micheletti et al., 2011) et 5,9.10^{-3} pour Arabidopsis thaliana (Schmidt et al., 2005). Cette forte diversité pourrait s’expliquer par la diversité des environnements dans lesquels pousse l’abricotier (altitude, climat), mais également par son allogamie. Ainsi, l’abricotier aurait subi un goulot d’étranglement plutôt faible, comme semble le démontrer également Tian Ming et al. (2007), du fait de son mode de reproduction à l’état sauvage. Chez le pommier, Cornille et al. (2012) le met en lien avec son mode de propagation par greffage pour le compartiment cultivé et sexué pour le compartiment sauvage.

Une séparation précoce du compartiment cultivé

Lors des analyses phylogénétiques réalisées à partir des marqueurs SSR et SNP, nous avons pu mettre en évidence une séparation nette entre les individus sauvages et les individus cultivés (variétés locales et anciennes, cultivars). Cette séparation aurait eu lieu très tôt dans l’histoire de l’abricotier et les individus domestiqués présenteraient des caractères très particuliers et prisés qui ont été conservés au cours du temps. Ceci est également confirmé par les faibles flux de gènes observés entre le compartiment cultivé et les populations sauvages. Il serait intéressant de dater cette séparation afin d’estimer quand a eu lieu la domestication.

Des interrogations qui demeurent...

Durant ce stage, l’analyse de la diversité chez l’abricotier dans son bassin d’origine a été réalisée selon différentes approches complémentaires, à partir de données phénotypiques ou génotypiques, ciblées ou au niveau du génome complet. Certains résultats sont apparus contradictoires et des analyses supplémentaires devront être conduites afin de répondre à ces
interrogations. En effet, alors que l’analyse génotypique au niveau pangénomique montre une diversité élevée chez les abricotiers issus de populations naturelles, nous avons observé peu ou pas de variabilité phénotypique en lien avec le débourrement végétatif dans le matériel sauvage. De manière similaire, Glémin et al. (2009) dans leur revue sur les herbacées relèvent une observation de Darwin encore inexpliquée à ce jour : la variation moléculaire dans le compartiment cultivé est plus faible que dans le compartiment sauvage, alors qu’au niveau phénotypique, l’inverse est observé. Dans notre cas, ceci peut signifier la fixation d’allèles au(x) locus/loci contrôlant ce caractère dans les populations naturelles, sous l’effet de leur adaptation aux conditions naturelles environnantes : hivers longs et très froids –jusqu’à -35/-40°C–, quasi-absence de saison printanière et donc évitement des gelées de fin de saison hivernale par un débourrement tardif. En contrepartie, l’Homme a sélectionné un sous-ensemble d’individus (ou même de mutants apparus tout au long de la domestication) présentant un maximum de variabilité phénotypique afin de pouvoir adapter cette culture à des conditions contrastées : hiver doux et absence de gelées de printemps au sud de l’Europe, hiver froid et conditions climatiques similaires à celles d’Asie centrale (même si moins extrêmes) en Europe de l’est. Plus récemment, les sélectionneurs nord-américains ont ré-introduit encore plus de variabilité phénotypique au matériel européen par croisement avec des individus provenant d’Asie centrale, afin de pouvoir adapter la culture de l’abricotier européen aux hivers rigoureux du continent nord-américain (voir le cas du cultivar « SEO »). En conséquence, il apparaît que le matériel cultivé ou sélectionné (variétés traditionnelles, anciennes) présente plus de variabilité phénotypique en lien avec le débourrement que les accessions issues de populations naturelles.

Lors des analyses de structuration réalisées sur le phénotype, nous avions déjà obtenu une séparation en deux groupes de la population du lac de Big Almaty. Le même phénomène a été observé avec les marqueurs SSR. Toutefois, la répartition des individus dans ces deux sous-groupes ne se fait pas de la même manière. Il faudrait donc étudier ce point plus en profondeur avec les individus récoltés pendant l’été.

Les analyses réalisées sur les marqueurs SSR et SNPs amènent à des résultats contradictoires concernant le taux d’hétérozygotes chez les abricotiers sauvages. Lasnier (2013) a, à partir des mêmes marqueurs microsatellites, également obtenu un excès d’homozygotes au sein des abricotiers cultivés. A contrario, l’analyse à partir des marqueurs SNP montre un excès d’hétérozygotes. Cette contradiction pourrait s’expliquer par le fait que les SNP sont des marqueurs bi-alléliques mettant potentiellement en évidence le polymorphisme pour chaque position nucléotidique et donc, au final, un plus grand nombre d’hétérozygotes. A l’inverse, il se pourrait qu’un certain nombre des marqueurs SSR utilisés dans ces études génèrent des allèles nuls, ce qui conduit à sur-estimer le nombre réel d’homozygotes. Cependant, cela ne semble pas être le cas, puisque tous les marqueurs douteux ont été éliminés du jeu de données initial. Une autre explication pourrait être une sur-estimation des loci hétérozygotes lors du reséquençage, par alignement de paralogues (ou de gènes dupliqués) et donc la mise en évidence de variation nucléotidique erronée conduisant à des SNPs faux-positifs. Un biais similaire fut constaté lors de l’analyse de séquences issues de 76 variétés d’abricotier cultivé, dans le cadre du projet ANR ABRIWG (article en préparation). Cet artéfact fut corrigé en ajoutant un filtre supplémentaire à la sélection des SNPs, filtre basé sur l’élimination des loci présentant une rareté allélique (Minor Allele Frequency, MAF<2%) et une déviation significative de l’hypothèse de départ, c’est-à-dire non issus de reproduction aléatoire (p-value<1%). Il sera donc intéressant de filtrer de manière similaire les SNPs issus de reséquençage de matériel abricotier sauvage, recalculer le Fis et vérifier ainsi si les populations naturelles présentent réellement ou pas un excès d’hétérozygotes.
A ce jour, relativement peu d’études ont utilisé l’approche combinée des marqueurs SSR et SNP. Nous arrivons ici à la même conclusion que celle de Li et al. (2010) ; Varshney et al. (2007) et Kota et al. (2001), à savoir que les marqueurs SSR semblent plus adaptés pour les études de diversité génétique, à moins peut-être de disposer d’une puissance de calcul suffisante à l’exploitation d’un très grand nombre de SNPs et à la condition que les marqueurs microsatellites ne génèrent pas (ou très peu) d’allèles nuls.

CONCLUSION ET PERSPECTIVES

Ce travail a permis de montrer que le phénomène d’auto-compatibilité présent dans le compartiment cultivé est apparu après la domestication de l’abricotier et que la diversité présente dans le compartiment sauvage est fort bien que peu structurée. Elle sera cependant intéressante à exploiter, notamment pour limiter les besoins en froid et introgresser la résistance à la sharka. Ainsi, la recherche de diversité sera plus efficace si elle est réalisée en Chine, au Kazakhstan ou en Ouzbékistan et la recherche d’association entre variation phénotypique et génotypique pourra être facilitée par cette faible structuration. Pour aller plus loin dans ce travail, il est nécessaire d’ajouter de nouvelles accessions à celle étudiées ici et de disposer d’une puissance de calcul plus importante. Cela permettra de repérer plus précisément les réservoirs de diversité en fonction du caractère d’intérêt.

BIBLIOGRAPHIE

Ressources en ligne


Publications


Decrooq S., Chague A., Lannibert P., Roch G., Audergon J-M., Geuna F., Chiozzotto R., Bassi D., Dondini L., Tartarini S., Salava J., Kriska B., Palmisano F., Karayiannis I., Decrooq V., (2014), Selecting with markers linked to the PPVs major QTL is not sufficient to predict resistance to Plum Pox Virus (PPV) in apricot, Tree Genetics and Genomes (en cours de publication)


François O., Durand E., (2010), Spatially explicit Bayesian clustering models in population genetics, Molecular Ecology Resources, 10, pp. 773-784.


Maghuly F., Fernandez E.B., Ruthner S., Pedryc A., Launer M., (2005), Microsatellite variability in apricots (Prunus armeniaca L.) reflects their geographic origin and breeding history, Tree Genetics and Genomes, 1, pp. 151-165.


ANNEXES
| Distance entre les lieux de prélèvement (km) - Partie 1. | Annexe I |
Annexe I : Distance entre les lieux de prélèvement (km)—Partie 2.
| Annexe I : Distance entre les lieux de prélèvement (km)-Partie 3 (fin). |
Annexe 11 : Liste des accèsions utilisées dans chaque partie - Phénotypage. Lecture des noms d'accession : les deux premières lettres indiquent le pays de récolte, le premier chiffre identifie le numéro de prélèvement, le deuxième chiffre indique que plusieurs échantillons ont été prélevés pour un prélèvement. Landrace renvoie aux variétés anciennes et arraché aux variétés locales.
Annexe II : Liste des accessions utilisées dans chaque partie - Marqueur SFB - partie 1.

Lecture des noms d’accession : les deux premières lettres indiquent le pays de récolte, le premier chiffre identifie le numéro de prélèvement, le deuxième chiffre indique que plusieurs échantillons ont été prélevés pour un prélèvement. Landrace renvoie aux variétés anciennes et amirchi aux variétés locales.
<table>
<thead>
<tr>
<th>Accession Pays</th>
<th>Population</th>
<th>Type</th>
<th>Accession Pays</th>
<th>Population</th>
<th>Type</th>
<th>Accession Pays</th>
<th>Population</th>
<th>Type</th>
<th>Accession Pays</th>
<th>Population</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>kz_11_2</td>
<td>Kazakhstan</td>
<td>Landrace</td>
<td>kz_40_2</td>
<td>Kazakhstan</td>
<td>Lake Esik</td>
<td>Sauvage</td>
<td>kz_87_2</td>
<td>Kazakhstan</td>
<td>Turgen valley</td>
<td>Sauvage</td>
<td>Turk_60</td>
</tr>
<tr>
<td>kz_13_3</td>
<td>Kazakhstan</td>
<td>Almaty</td>
<td>kz_40_4</td>
<td>Kazakhstan</td>
<td>Lake Esik</td>
<td>Sauvage</td>
<td>kz_87_3</td>
<td>Kazakhstan</td>
<td>Turgen valley</td>
<td>Sauvage</td>
<td>Turk_68</td>
</tr>
<tr>
<td>kz_11_4</td>
<td>Kazakhstan</td>
<td>Almaty</td>
<td>kz_9_3</td>
<td>Kazakhstan</td>
<td>Big Almaty lake</td>
<td>Sauvage</td>
<td>kz_8_1</td>
<td>Kazakhstan</td>
<td>Almaty-Pomological garden</td>
<td>Sauvage</td>
<td>Turk_60</td>
</tr>
<tr>
<td>kz_13_1</td>
<td>Kazakhstan</td>
<td>Turgen Valley</td>
<td>sauvage</td>
<td>kz_3_3</td>
<td>Kazakhstan</td>
<td>Almaty-Pomological garden</td>
<td>melange</td>
<td>kz_9_2</td>
<td>Kazakhstan</td>
<td>Almaty-Pomological garden</td>
<td>Sauvage</td>
</tr>
<tr>
<td>kz_13_2</td>
<td>Kazakhstan</td>
<td>Turgen Valley</td>
<td>sauvage</td>
<td>kz_3_3</td>
<td>Kazakhstan</td>
<td>Almaty-Pomological garden</td>
<td>melange</td>
<td>kz_9_3</td>
<td>Kazakhstan</td>
<td>Almaty-Pomological garden</td>
<td>Sauvage</td>
</tr>
<tr>
<td>kz_11_4</td>
<td>Kazakhstan</td>
<td>Turgen Valley</td>
<td>sauvage</td>
<td>kz_3_3</td>
<td>Kazakhstan</td>
<td>Almaty-Pomological garden</td>
<td>melange</td>
<td>kz_9_1</td>
<td>Kazakhstan</td>
<td>Canyon Kastelen</td>
<td>Sauvage</td>
</tr>
<tr>
<td>kz_11_1</td>
<td>Kazakhstan</td>
<td>Turgen Valley</td>
<td>sauvage</td>
<td>kz_3_3</td>
<td>Kazakhstan</td>
<td>Almaty-Pomological garden</td>
<td>melange</td>
<td>kz_9_1</td>
<td>Kazakhstan</td>
<td>Canyon Kastelen</td>
<td>Sauvage</td>
</tr>
<tr>
<td>kz_11_1</td>
<td>Kazakhstan</td>
<td>Turgen Valley</td>
<td>sauvage</td>
<td>kz_3_3</td>
<td>Kazakhstan</td>
<td>Almaty-Pomological garden</td>
<td>melange</td>
<td>kz_9_1</td>
<td>Kazakhstan</td>
<td>Canyon Kastelen</td>
<td>Sauvage</td>
</tr>
<tr>
<td>kz_11_5</td>
<td>Kazakhstan</td>
<td>Canyon Kastelen</td>
<td>sauvage</td>
<td>kz_3_4</td>
<td>Kazakhstan</td>
<td>Almaty-Pomological garden</td>
<td>melange</td>
<td>kz_9_2</td>
<td>Kazakhstan</td>
<td>Almaty</td>
<td>Marché</td>
</tr>
<tr>
<td>kz_11_5</td>
<td>Kazakhstan</td>
<td>Canyon Kastelen</td>
<td>sauvage</td>
<td>kz_3_4</td>
<td>Kazakhstan</td>
<td>Almaty-Pomological garden</td>
<td>melange</td>
<td>kz_9_2</td>
<td>Kazakhstan</td>
<td>Almaty</td>
<td>Marché</td>
</tr>
<tr>
<td>kz_11_5</td>
<td>Kazakhstan</td>
<td>Cannon Kastelen</td>
<td>sauvage</td>
<td>kz_3_4</td>
<td>Kazakhstan</td>
<td>Almaty-Pomological garden</td>
<td>melange</td>
<td>kz_9_2</td>
<td>Kazakhstan</td>
<td>Almaty</td>
<td>Marché</td>
</tr>
<tr>
<td>kz_11_5</td>
<td>Kazakhstan</td>
<td>Cannon Kastelen</td>
<td>sauvage</td>
<td>kz_3_4</td>
<td>Kazakhstan</td>
<td>Almaty-Pomological garden</td>
<td>melange</td>
<td>kz_9_2</td>
<td>Kazakhstan</td>
<td>Almaty</td>
<td>Marché</td>
</tr>
<tr>
<td>kz_11_5</td>
<td>Kazakhstan</td>
<td>Cannon Kastelen</td>
<td>sauvage</td>
<td>kz_3_4</td>
<td>Kazakhstan</td>
<td>Almaty-Pomological garden</td>
<td>melange</td>
<td>kz_9_2</td>
<td>Kazakhstan</td>
<td>Almaty</td>
<td>Marché</td>
</tr>
<tr>
<td>kz_11_5</td>
<td>Kazakhstan</td>
<td>Cannon Kastelen</td>
<td>sauvage</td>
<td>kz_3_4</td>
<td>Kazakhstan</td>
<td>Almaty-Pomological garden</td>
<td>melange</td>
<td>kz_9_2</td>
<td>Kazakhstan</td>
<td>Almaty</td>
<td>Marché</td>
</tr>
<tr>
<td>kz_11_5</td>
<td>Kazakhstan</td>
<td>Cannon Kastelen</td>
<td>sauvage</td>
<td>kz_3_4</td>
<td>Kazakhstan</td>
<td>Almaty-Pomological garden</td>
<td>melange</td>
<td>kz_9_2</td>
<td>Kazakhstan</td>
<td>Almaty</td>
<td>Marché</td>
</tr>
<tr>
<td>kz_11_5</td>
<td>Kazakhstan</td>
<td>Cannon Kastelen</td>
<td>sauvage</td>
<td>kz_3_4</td>
<td>Kazakhstan</td>
<td>Almaty-Pomological garden</td>
<td>melange</td>
<td>kz_9_2</td>
<td>Kazakhstan</td>
<td>Almaty</td>
<td>Marché</td>
</tr>
<tr>
<td>kz_11_5</td>
<td>Kazakhstan</td>
<td>Cannon Kastelen</td>
<td>sauvage</td>
<td>kz_3_4</td>
<td>Kazakhstan</td>
<td>Almaty-Pomological garden</td>
<td>melange</td>
<td>kz_9_2</td>
<td>Kazakhstan</td>
<td>Almaty</td>
<td>Marché</td>
</tr>
<tr>
<td>kz_11_5</td>
<td>Kazakhstan</td>
<td>Cannon Kastelen</td>
<td>sauvage</td>
<td>kz_3_4</td>
<td>Kazakhstan</td>
<td>Almaty-Pomological garden</td>
<td>melange</td>
<td>kz_9_2</td>
<td>Kazakhstan</td>
<td>Almaty</td>
<td>Marché</td>
</tr>
</tbody>
</table>

**Annexe II : Liste des accessions utilisées dans chaque partie - Marqueur SFB - partie 2.** Lecture des noms d’accession : les deux premières lettres indiquent le pays de récolte, le premier chiffre identifie le numéro de prélèvement, le deuxième chiffre indique que plusieurs échantillons ont été prélés pour un prélèvement. Landrace renvoie aux variétés anciennes et arraché aux variétés locales.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Chychkan river</td>
<td>Sauvage</td>
<td>Chychkan river</td>
<td>Sauvage</td>
<td>Chychkan river</td>
<td>Sauvage</td>
<td>Chychkan river</td>
<td>Sauvage</td>
<td>Chychkan river</td>
<td>Sauvage</td>
<td>Chychkan river</td>
<td>Sauvage</td>
<td>Chychkan river</td>
<td>Sauvage</td>
<td>Chychkan river</td>
<td>Sauvage</td>
<td>Chychkan river</td>
<td>Sauvage</td>
<td>Chychkan river</td>
<td>Sauvage</td>
<td>Chychkan river</td>
<td>Sauvage</td>
</tr>
<tr>
<td>Sarachek</td>
<td>Sauvage</td>
<td>Sarachek</td>
<td>Sauvage</td>
<td>Sarachek</td>
<td>Sauvage</td>
<td>Sarachek</td>
<td>Sauvage</td>
<td>Sarachek</td>
<td>Sauvage</td>
<td>Sarachek</td>
<td>Sauvage</td>
<td>Sarachek</td>
<td>Sauvage</td>
<td>Sarachek</td>
<td>Sauvage</td>
<td>Sarachek</td>
<td>Sauvage</td>
<td>Sarachek</td>
<td>Sauvage</td>
<td>Sarachek</td>
<td>Sauvage</td>
</tr>
<tr>
<td>Luntai</td>
<td>Sauvage</td>
<td>Luntai</td>
<td>Sauvage</td>
<td>Luntai</td>
<td>Sauvage</td>
<td>Luntai</td>
<td>Sauvage</td>
<td>Luntai</td>
<td>Sauvage</td>
<td>Luntai</td>
<td>Sauvage</td>
<td>Luntai</td>
<td>Sauvage</td>
<td>Luntai</td>
<td>Sauvage</td>
<td>Luntai</td>
<td>Sauvage</td>
<td>Luntai</td>
<td>Sauvage</td>
<td>Luntai</td>
<td>Sauvage</td>
</tr>
<tr>
<td>Bakour</td>
<td>Sauvage</td>
<td>Bakour</td>
<td>Sauvage</td>
<td>Bakour</td>
<td>Sauvage</td>
<td>Bakour</td>
<td>Sauvage</td>
<td>Bakour</td>
<td>Sauvage</td>
<td>Bakour</td>
<td>Sauvage</td>
<td>Bakour</td>
<td>Sauvage</td>
<td>Bakour</td>
<td>Sauvage</td>
<td>Bakour</td>
<td>Sauvage</td>
<td>Bakour</td>
<td>Sauvage</td>
<td>Bakour</td>
<td>Sauvage</td>
</tr>
</tbody>
</table>

Annexe 11 : Liste des accessions utilisées dans chaque partie - Marqueur ZP002 - partie 1. Lecture des noms d’accession : les deux premières lettres indiquent le pays de récolte, le premier chiffre identifie le numéro de prélèvement, le deuxième chiffre indique que plusieurs échantillons ont été prélevés pour un prélèvement. Landrace renvoie aux variétés anciennes et amrché aux variétés locales, mélange à un ensemble d’individus sauvages et landraces.
<table>
<thead>
<tr>
<th>Accession Pays</th>
<th>Population Type</th>
<th>Accession Pays</th>
<th>Population Type</th>
<th>Accession Pays</th>
<th>Population Type</th>
<th>Accession Pays</th>
<th>Population Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>197.1 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
<td>197.2 Kazakhstan</td>
<td>Big Almaty Lake</td>
<td>197.3 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
<td>197.4 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
</tr>
<tr>
<td>197.5 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
<td>197.6 Kazakhstan</td>
<td>Big Almaty Lake</td>
<td>197.7 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
<td>197.8 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
</tr>
<tr>
<td>197.9 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
<td>197.10 Kazakhstan</td>
<td>Big Almaty Lake</td>
<td>197.11 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
<td>197.12 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
</tr>
<tr>
<td>197.13 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
<td>197.14 Kazakhstan</td>
<td>Big Almaty Lake</td>
<td>197.15 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
<td>197.16 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
</tr>
<tr>
<td>197.17 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
<td>197.18 Kazakhstan</td>
<td>Big Almaty Lake</td>
<td>197.19 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
<td>197.20 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
</tr>
<tr>
<td>197.21 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
<td>197.22 Kazakhstan</td>
<td>Big Almaty Lake</td>
<td>197.23 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
<td>197.24 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
</tr>
<tr>
<td>197.25 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
<td>197.26 Kazakhstan</td>
<td>Big Almaty Lake</td>
<td>197.27 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
<td>197.28 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
</tr>
<tr>
<td>197.29 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
<td>197.30 Kazakhstan</td>
<td>Big Almaty Lake</td>
<td>197.31 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
<td>197.32 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
</tr>
<tr>
<td>197.33 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
<td>197.34 Kazakhstan</td>
<td>Big Almaty Lake</td>
<td>197.35 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
<td>197.36 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
</tr>
<tr>
<td>197.37 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
<td>197.38 Kazakhstan</td>
<td>Big Almaty Lake</td>
<td>197.39 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
<td>197.40 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
</tr>
<tr>
<td>197.41 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
<td>197.42 Kazakhstan</td>
<td>Big Almaty Lake</td>
<td>197.43 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
<td>197.44 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
</tr>
<tr>
<td>197.45 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
<td>197.46 Kazakhstan</td>
<td>Big Almaty Lake</td>
<td>197.47 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
<td>197.48 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
</tr>
<tr>
<td>197.49 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
<td>197.50 Kazakhstan</td>
<td>Big Almaty Lake</td>
<td>197.51 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
<td>197.52 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
</tr>
<tr>
<td>197.53 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
<td>197.54 Kazakhstan</td>
<td>Big Almaty Lake</td>
<td>197.55 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
<td>197.56 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
</tr>
<tr>
<td>197.57 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
<td>197.58 Kazakhstan</td>
<td>Big Almaty Lake</td>
<td>197.59 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
<td>197.60 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
</tr>
<tr>
<td>197.61 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
<td>197.62 Kazakhstan</td>
<td>Big Almaty Lake</td>
<td>197.63 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
<td>197.64 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
</tr>
<tr>
<td>197.65 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
<td>197.66 Kazakhstan</td>
<td>Big Almaty Lake</td>
<td>197.67 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
<td>197.68 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
</tr>
<tr>
<td>197.69 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
<td>197.70 Kazakhstan</td>
<td>Big Almaty Lake</td>
<td>197.71 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
<td>197.72 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
</tr>
<tr>
<td>197.73 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
<td>197.74 Kazakhstan</td>
<td>Big Almaty Lake</td>
<td>197.75 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
<td>197.76 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
</tr>
<tr>
<td>197.77 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
<td>197.78 Kazakhstan</td>
<td>Big Almaty Lake</td>
<td>197.79 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
<td>197.80 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
</tr>
<tr>
<td>197.81 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
<td>197.82 Kazakhstan</td>
<td>Big Almaty Lake</td>
<td>197.83 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
<td>197.84 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
</tr>
<tr>
<td>197.85 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
<td>197.86 Kazakhstan</td>
<td>Big Almaty Lake</td>
<td>197.87 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
<td>197.88 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
</tr>
<tr>
<td>197.89 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
<td>197.90 Kazakhstan</td>
<td>Big Almaty Lake</td>
<td>197.91 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
<td>197.92 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
</tr>
<tr>
<td>197.93 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
<td>197.94 Kazakhstan</td>
<td>Big Almaty Lake</td>
<td>197.95 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
<td>197.96 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
</tr>
<tr>
<td>197.97 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
<td>197.98 Kazakhstan</td>
<td>Big Almaty Lake</td>
<td>197.99 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
<td>197.100 Kazakhstan</td>
<td>Almaty - Pomological garden</td>
</tr>
</tbody>
</table>

Annexe II : Liste des accès utilisées dans chaque partie - Marqueur ZP002 - partie 2. Lecture des noms d’acccession : les deux premières lettres indiquent le pays de récolte, le premier chiffre identifie le numéro de prélèvement, le deuxième chiffre indique que plusieurs échantillons ont été prélevés pour un prélèvement. Landrace renvoie aux variétés anciennes et amnché aux variétés locales, mélange à un ensemble d’individus sauvages et landraces.
Annexe II : Liste des accessions utilisées dans chaque partie - Marqueurs SSR - partie 1. Lecture des noms d'accession : les deux premières lettres indiquent le pays de récolte, le premier chiffre identifie le numéro de prélèvement, le deuxième chiffre indique que plusieurs échantillons ont été prélevés pour un prélèvement. Landrace renvoie aux variétés anciennes et améliorées aux variétés locales, mélange à un ensemble d'individus sauvages et landraces.
Annexe II : Liste des accessions utilisées dans chaque partie - Marqueurs SNP. Lecture des noms d’accession : les deux premières lettres indiquent le pays de récolte, le premier chiffre identifie le nombre de prélèvement, le deuxième chiffre indique que plusieurs échantillons ont été prélevés pour un prélèvement. Landrace renvoie aux variétés anciennes et amélior aux variétés locales, mélangé à un ensemble d’individus sauvages et landraces.

<table>
<thead>
<tr>
<th>Accession Pays</th>
<th>Population Type Accession Pays Population Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>kz_ma_9</td>
<td>Kazakhstan Almaty market marché</td>
</tr>
<tr>
<td>us_218</td>
<td>USA Pakistan landrace</td>
</tr>
<tr>
<td>us_213</td>
<td>USA Pakistan landrace</td>
</tr>
<tr>
<td>us_215</td>
<td>USA Pakistan landrace</td>
</tr>
<tr>
<td>us_217</td>
<td>USA Pakistan landrace</td>
</tr>
</tbody>
</table>

Annexe II : Liste des accessions utilisées dans chaque partie - Marqueurs SSR - partie 2. Lecture des noms d’accession : les deux premières lettres indiquent le pays de récolte, le premier chiffre identifie le numéro de prélèvement, le deuxième chiffre indique que plusieurs échantillons ont été prélevés pour un prélèvement. Landrace renvoie aux variétés anciennes et amélior aux variétés locales, mélangé à un ensemble d’individus sauvages et landraces.
Annexe II : Amorces ciblant l’insertion de 358 pb dans le gène SFB et conditions PCR correspondantes. 
F=amorce sens, R=amorce anti-sens, pb=paires de bases.

<table>
<thead>
<tr>
<th>Nom de la paire d’amorces</th>
<th>Séquence des amorces (5’-&gt;3’)</th>
<th>Taille du fragment attendu (pb)</th>
<th>Conditions PCR</th>
<th>Référence</th>
</tr>
</thead>
<tbody>
<tr>
<td>SFBc_F</td>
<td>F-TCGACATCCTTAGAAGACTACCCGC</td>
<td>972</td>
<td>94°C-2min, 35x(94°C-1min, 60°C-40s, 72°C-1min), 72°C-20 min</td>
<td>Vilanova et al. 2006</td>
</tr>
<tr>
<td></td>
<td>R-TCAAGAACTTGGTTGGATTTCG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SFBc_Fv1</td>
<td>F-TGAATTTTCGATAGTCTATACCA</td>
<td>633</td>
<td>94°C-2min, 35x(94°C-1min, 50°C-40s, 72°C-40s), 72°C-20 min</td>
<td>Amorce sens développée au laboratoire, Vilanova et al. 2006</td>
</tr>
<tr>
<td></td>
<td>R-TCAAGAACTTGGTTGGATTTCG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SFBc_Fv2</td>
<td>F-TCATAGCACCAGATGCACTTTGC</td>
<td>265</td>
<td>94°C-2min, 35x(94°C-1min, 60°C-40s, 72°C-30s), 72°C-20 min</td>
<td>Amorce sens développée au laboratoire, Vilanova et al. 2006</td>
</tr>
<tr>
<td></td>
<td>R-TCAAGAACTTGGTTGGATTTCG</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Annexe III: Protocole pour l’extraction d’ADN

**Tampon 1** : tris-HCl pH8 à 0,2M, EDTA pH8 à 0,07M, NaCL à 2M et ajout à la dernière minute de sodium metabisulfite à 0,02M.

**Tampon 2** : HATMAB à 2%, NaCl à 1,4M, ADTA pH8 à 0,02M, tris-HCl ph8 à 0,1M.

Broyer environ 5 feuilles dans un sachet ELISA dans 10mL de tampon 1. Placer 600µL dans un tube eppendorf et ajouter 450µL et mélanger. Placer au bain-marie 1h à 65°C. Ajouter 10µL de RNase à 10mg/mL et attendre 10 minutes puis centrifuger à 4°C (10000rpm-20 min).

Transférent 1ml de surnageant dans une autre tube eppendorf avec 900µL de chloroforme:octanol(24 :1). Mélanger jusqu’à la formation d’un mélange blanchâtre, placer sur la glace puis remélanger. Centrifuger 5min à 10000rpm (4°C). Transférent 800µL de surnageant dans une autre tube eppendorf avec 800µL de chloroforme:octanol(24 :1). Mélanger jusqu’à la formation d’un mélange blanchâtre, placer sur la glace puis remélanger. Centrifuger 5min à 10000rpm (4°C).

Transvaser les 800µL de surnageant dans un tube contenant 800µL Isopropanol à -20°C et 400µL d’acétate d’ammonium à 10M et replacer dans la glace. Mélanger et placer à -20°C pendant au moins 25°C en mélangeant régulièrement. Centrifuger 30min à 10000rpm à 4°C. Jeter le surnageant puis ajouter 750µL d’éthanol à 70% glacé, mélanger pour décoller le culot., centrifuger 10 min à 10000rpm, jeter le surnageant et retourner le tube sur du papier absorbant. Sécher le culot pendant 10min au speedvac et diluer le culot dans 100-200µL d’EDTA à 0,1X pour la nuit.

Vérifier la quantité d’ADN présente sur gel d’agarose à 0,8%.
Annexe IV : Sorties graphiques de l'ACP sur les deux premiers axes à partir des données HF (heures de froid nécessaire pour atteindre le stade 3) et RG (Rayonnement solaire nécessaire pour atteindre le stade 3).
Annexe V : Coordonnées du barycentre de chaque population et ses caractéristiques. Dim 1 correspond aux coordonnées sur le premier axe de l’ACP et Dim 2 correspond aux coordonnées sur le deuxième axe de l’ACP. La colonne Pays indique le pays dans lequel la population a été échantillonnée et la colonne GPS (Global Positioning system) contient les informations de localisation géographique et d’altitude (alt) lorsqu’elles sont disponibles (dnd -données non disponibles-dans le cas contraire).

<table>
<thead>
<tr>
<th>Population</th>
<th>Dim1</th>
<th>Dim2</th>
<th>Pays</th>
<th>GPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>proche Sayram</td>
<td>3.91638816</td>
<td>-1.41567883</td>
<td>Kazakhstan</td>
<td>42°14’58N - 70°04’31E alt.1014m</td>
</tr>
<tr>
<td>Chymkent</td>
<td>3.50317616</td>
<td>-1.03805347</td>
<td>Kazakhstan</td>
<td>Dnd</td>
</tr>
<tr>
<td>Chimgan Beldersay</td>
<td>3.0344981</td>
<td>-0.74772422</td>
<td>Ouzbékistan</td>
<td>41°30’57N - 69°57’31E alt.1502m</td>
</tr>
<tr>
<td>Samarcande</td>
<td>2.94145798</td>
<td>-0.48866605</td>
<td>Ouzbékistan</td>
<td>Dnd</td>
</tr>
<tr>
<td>Boukhara</td>
<td>2.39074798</td>
<td>-0.40425476</td>
<td>Ouzbékistan</td>
<td>39°46’39N - 64°24’24E alt.228m</td>
</tr>
<tr>
<td>Urumqi</td>
<td>2.28349943</td>
<td>0.13430387</td>
<td>Chine</td>
<td>Dnd</td>
</tr>
<tr>
<td>Soutjetskoe</td>
<td>1.89854143</td>
<td>0.30547768</td>
<td>Kazakhstan</td>
<td>Dnd</td>
</tr>
<tr>
<td>Naxçivan-Arinc mountain</td>
<td>1.83074651</td>
<td>0.36751844</td>
<td>Azerbaïdjan</td>
<td>39°27’54,27N - 45°35’57,04E</td>
</tr>
<tr>
<td>Mehe2</td>
<td>1.27725914</td>
<td>0.66232072</td>
<td>Chine</td>
<td>43°14’11.52N - 82°44’07.69E</td>
</tr>
<tr>
<td>Naxçivan-Badamli village</td>
<td>1.05058551</td>
<td>0.70243238</td>
<td>Azerbaïdjan</td>
<td>39°26’58,16N - 45°31’43,49E alt.1400m</td>
</tr>
<tr>
<td>Tashkent</td>
<td>1.00482673</td>
<td>0.64105835</td>
<td>Ouzbékistan</td>
<td>Dnd</td>
</tr>
<tr>
<td>Naxçivan-Ordubbad-NusNus</td>
<td>0.92801943</td>
<td>0.57188455</td>
<td>Azerbaïdjan</td>
<td>38°56’46,71N - 46°01’15,02E</td>
</tr>
<tr>
<td>Luntaï</td>
<td>0.86260513</td>
<td>0.70447113</td>
<td>Chine</td>
<td>Dnd</td>
</tr>
<tr>
<td>Aksu Zhabagyly</td>
<td>0.46024415</td>
<td>0.15915745</td>
<td>Kazakhstan</td>
<td>42°24’35N - 70°32’45E</td>
</tr>
<tr>
<td>Yili</td>
<td>-0.11334718</td>
<td>1.05310227</td>
<td>Chine</td>
<td>43°22’42.97N - 83°34’50.33E</td>
</tr>
<tr>
<td>Ak-Kain</td>
<td>-0.48410846</td>
<td>0.80701637</td>
<td>Kazakhstan</td>
<td>43°10’32.48N - 76°58’11.73E alt.1220m</td>
</tr>
<tr>
<td>wine yard2</td>
<td>-0.7570249</td>
<td>1.05938793</td>
<td>Kazakhstan</td>
<td>43°29’26.26N - 77°42’30.82E</td>
</tr>
<tr>
<td>Canyon Kaskelen</td>
<td>-1.23377152</td>
<td>0.30538519</td>
<td>Kazakhstan</td>
<td>43°10’00.77N - 76°36’56E</td>
</tr>
<tr>
<td>Almaty</td>
<td>-1.23544468</td>
<td>0.92783295</td>
<td>Kazakhstan</td>
<td>43°13’36.10N - 76°52’32.35E</td>
</tr>
<tr>
<td>Almaty-Pomological garden</td>
<td>-1.3129843</td>
<td>1.13270203</td>
<td>Kazakhstan</td>
<td>43°16’48.48N - 77°12’05.84E</td>
</tr>
<tr>
<td>Lake Esik</td>
<td>-1.86738593</td>
<td>1.09544393</td>
<td>Kazakhstan</td>
<td>43°15’15.55N - 77°29’11.61E</td>
</tr>
<tr>
<td>Turgen valley</td>
<td>-2.38638571</td>
<td>0.07698984</td>
<td>Kazakhstan</td>
<td>43°20’20.37N - 77°36’44.27E</td>
</tr>
<tr>
<td>Mehe1</td>
<td>-2.73275533</td>
<td>1.02217357</td>
<td>Chine</td>
<td>43°14’45.44N - 82°42’10.06E</td>
</tr>
<tr>
<td>Big Almaty Lake</td>
<td>-3.53108349</td>
<td>-0.17907121</td>
<td>Kazakhstan</td>
<td>43°05’39.15N - 76°57’19.06E</td>
</tr>
<tr>
<td>Medeu valley</td>
<td>-4.02312991</td>
<td>-1.76872651</td>
<td>Kazakhstan</td>
<td>43°09’24.40N - 77°03’41.06E</td>
</tr>
</tbody>
</table>
Annexe VI : Résultats des ANOVAs sur le nombre d'heures de froid (HF), la quantité de rayonnement solaire (RG) et le nombre d'heures de jour (HJ) nécessaires pour atteindre le stade 3 du débourrement en fonction des critères ‘type d'accession’, ‘population d'origine’ et ‘pays d'origine’. *** signifie que la valeur est significative au seuil de 0%.

<table>
<thead>
<tr>
<th>Effet</th>
<th>HF</th>
<th>RG</th>
<th>HJ</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F</td>
<td>p-value</td>
<td>F</td>
</tr>
<tr>
<td>Population</td>
<td>7.6632</td>
<td>&lt;2E-16</td>
<td>10.8795</td>
</tr>
<tr>
<td>Type</td>
<td>0.8454</td>
<td>0.4304</td>
<td>0.7847</td>
</tr>
<tr>
<td>Pays d’origine</td>
<td>0.0204</td>
<td>0.9798</td>
<td>0.2018</td>
</tr>
</tbody>
</table>
Annexe VII : Résultats de la MANOVA sur les notes attribuées tout au long du débourrement en fonction des effets du type d'accession, de la population et du pays d'origine. *** signifie que la valeur est significative au seuil de 0%.

<table>
<thead>
<tr>
<th>Variables qualitatives</th>
<th>ensemble des accessions</th>
<th>Accessions sauvages</th>
<th>Variétés anciennes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>λ de Wilks</td>
<td>p-value</td>
<td>λ de Wilks</td>
</tr>
<tr>
<td>Type</td>
<td>0.57774</td>
<td>&lt;2.20E-16</td>
<td>-</td>
</tr>
<tr>
<td>Population</td>
<td>0.18444</td>
<td>2.51E-14</td>
<td>0.28415</td>
</tr>
<tr>
<td>Origine</td>
<td>0.9016</td>
<td>0.1965</td>
<td>-</td>
</tr>
</tbody>
</table>
Annexe VIII : Résultats de la MANOVA sur les notes attribuées tout au long du débourrement en fonction des effets du nombre d’heures de froid (HF), de la quantité de rayonnement solaire (RG) et du nombre d’heures de jour (HJ) nécessaires pour atteindre le stade 3 du débourrement.*** signifie que la valeur est significative au seuil de 0%, ** est significatif au seuil de 0,1%.

<table>
<thead>
<tr>
<th>Variables quantitatives</th>
<th>ensemble des accessions</th>
<th>Accessions sauvages</th>
<th>Variétés anciennes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>λ de Wilks</td>
<td>p-value</td>
<td>λ de Wilks</td>
</tr>
<tr>
<td>HF</td>
<td>0.05389</td>
<td>&lt;2.20E-16</td>
<td>***</td>
</tr>
<tr>
<td>RG</td>
<td>0.24869</td>
<td>&lt;2.20E-16</td>
<td>***</td>
</tr>
<tr>
<td>HJ</td>
<td>0.1411</td>
<td>&lt;2.20E-16</td>
<td>***</td>
</tr>
<tr>
<td>HF:RG</td>
<td>0.42336</td>
<td>&lt;2.20E-16</td>
<td>***</td>
</tr>
<tr>
<td>HF:HJ</td>
<td>0.72534</td>
<td>&lt;2.20E-16</td>
<td>***</td>
</tr>
<tr>
<td>RG:HJ</td>
<td>0.91231</td>
<td>0.004357</td>
<td>**</td>
</tr>
<tr>
<td>HF:RG:HJ</td>
<td>0.61889</td>
<td>&lt;2.20E-16</td>
<td>***</td>
</tr>
</tbody>
</table>
Annexe IX : Résultats de la MANOVA sur le nombre d’heures de froid la quantité de rayonnement solaire et le nombre d’heures de jour nécessaires pour atteindre le stade 3 du débourrement en fonction des effets type d’accession, population d’origine et pays d’origine. *** signifie que la valeur est significative au seuil de 0%.

<table>
<thead>
<tr>
<th>Variables qualitatives</th>
<th>ensemble des accessions</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>λ de Wilks</td>
<td>p-value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type</td>
<td>0.69726</td>
<td>&lt;2E-16</td>
<td>***</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Population</td>
<td>0.49625</td>
<td>&lt;2E-16</td>
<td>***</td>
<td>0.6288</td>
<td>8.36E-06</td>
<td>***</td>
<td>0.47158</td>
<td>7.70E-08</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Origine</td>
<td>0.98559</td>
<td>0.6035</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.9954</td>
<td>0.8876</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Annexe X : Structuration détectée par GENELAND sur le phénotype (HF=nombre d’heures de froid pour atteindre le stade 3) des individus sauvages. Les individus apparentés sont conservés. La population du lac de Big Almaty se scinde en deux.
Annexe XI : Comparaison des Fis obtenus avec les marqueurs SSR et SNP. Lorsque la position du marqueur SSR n’est pas renseignée, cela signifie que ce marqueur n’apparait pas dans la base de données sur le génome des rosacées (http://www.rosaceae.org/).

<table>
<thead>
<tr>
<th>Marqueur SSR</th>
<th>Scaffold</th>
<th>Position</th>
<th>Fis (kazaeks sauvages)</th>
<th>Position du SNP le plus proche</th>
<th>Fis au SNP le plus proche</th>
</tr>
</thead>
<tbody>
<tr>
<td>BPPCT001</td>
<td>2</td>
<td>16 134 154</td>
<td>0.0196</td>
<td>16 161 717</td>
<td>-0.4323</td>
</tr>
<tr>
<td>UDP98-409</td>
<td>8</td>
<td>-</td>
<td>-0.1189</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>AMPA116</td>
<td>2</td>
<td>25 108 684</td>
<td>0.079</td>
<td>25 108 258</td>
<td>-1</td>
</tr>
<tr>
<td>AMPA100</td>
<td>6</td>
<td>-</td>
<td>0.0295</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>BPPCT004</td>
<td>2</td>
<td>-</td>
<td>-0.0228</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>BPPCT040</td>
<td>4</td>
<td>6 460 443</td>
<td>-0.0625</td>
<td>6 454 079</td>
<td>0</td>
</tr>
<tr>
<td>UDP97-402</td>
<td>4</td>
<td>10 486 180</td>
<td>0.56</td>
<td>10 485 850</td>
<td>-0.9241</td>
</tr>
<tr>
<td>AMPA109</td>
<td>1</td>
<td>39 337 498</td>
<td>0.4518</td>
<td>38 344 006</td>
<td>0</td>
</tr>
<tr>
<td>CPPCT006</td>
<td>8</td>
<td>13 659 021</td>
<td>-0.0752</td>
<td>13 660 657</td>
<td>-0.027</td>
</tr>
<tr>
<td>AMPA119</td>
<td>3</td>
<td>6 357 496</td>
<td>0.0281</td>
<td>6 346 989</td>
<td>0</td>
</tr>
<tr>
<td>AMPA101</td>
<td>3</td>
<td>3 595 324</td>
<td>0.0781</td>
<td>4 592 725</td>
<td>0</td>
</tr>
<tr>
<td>BPPCT017</td>
<td>5</td>
<td>11 174 442</td>
<td>-0.2001</td>
<td>11 192 782</td>
<td>0.3812</td>
</tr>
<tr>
<td>BPPCT008</td>
<td>6</td>
<td>10 280 088</td>
<td>-0.0958</td>
<td>11 292 934</td>
<td>1</td>
</tr>
<tr>
<td>BPPCT038</td>
<td>5</td>
<td>14 658 198</td>
<td>-0.1306</td>
<td>14 649 267</td>
<td>0</td>
</tr>
<tr>
<td>UDP98-412</td>
<td>6</td>
<td>24 753 513</td>
<td>0.1584</td>
<td>24 755 315</td>
<td>0</td>
</tr>
<tr>
<td>CPPCT022</td>
<td>7</td>
<td>10 225 365</td>
<td>-0.009</td>
<td>10 186 462</td>
<td>0.1253</td>
</tr>
<tr>
<td>BPPCT030</td>
<td>2</td>
<td>22 248 484</td>
<td>-0.1012</td>
<td>22 264 778</td>
<td>-1</td>
</tr>
<tr>
<td>BPPCT025</td>
<td>6</td>
<td>21 129 947</td>
<td>-0.0204</td>
<td>21 125 929</td>
<td>-0.617</td>
</tr>
<tr>
<td>CPPCT033</td>
<td>7</td>
<td>16 702 195</td>
<td>0.3245</td>
<td>16 695 993</td>
<td>0</td>
</tr>
<tr>
<td>CCPCT030</td>
<td>6</td>
<td>26 851 012</td>
<td>0.1282</td>
<td>26 851 581</td>
<td>0.1399</td>
</tr>
<tr>
<td>PGS1_21</td>
<td>1</td>
<td>8 157 718</td>
<td>-0.0421</td>
<td>8 157 918</td>
<td>0.0194</td>
</tr>
</tbody>
</table>
Annexe XII : Valeur de Fis pour chaque SNP le long du génome.