



**HAL**  
open science

# Analyse de la diversité génétique chez l'abricotier (*Prunus armeniaca* L.) à l'aide de marqueurs microsatellites

Adrien Lasnier

► **To cite this version:**

Adrien Lasnier. Analyse de la diversité génétique chez l'abricotier (*Prunus armeniaca* L.) à l'aide de marqueurs microsatellites. Sciences agricoles. 2013. dumas-00976659

**HAL Id: dumas-00976659**

**<https://dumas.ccsd.cnrs.fr/dumas-00976659>**

Submitted on 10 Apr 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.



**AGROCAMPUS OUEST CFR  
d'Angers**  
2 rue André Le Nôtre  
F-49045 Angers cedex 01



**BioGEVES**  
Domaine du Magneraud  
B.P. 52  
17700 Surgères



**INRA GAFL UR 1052**  
Domaine St Maurice  
BP 94  
84143 Montfavet cedex

## Mémoire de Fin d'Études

### **Diplôme d'Ingénieur de l'Institut Supérieur des Sciences Agronomiques, Agroalimentaires, Horticoles et du Paysage \***

Année universitaire : 2012-2013

Spécialisation ou option : GDV

## **Analyse de la diversité génétique chez l'abricotier (*Prunus armeniaca* L.) à l'aide de marqueurs microsatellites**

Par : **Adrien LASNIER**

*Volet à renseigner par l'enseignant responsable de l'option/spécialisation \**

Bon pour dépôt (version définitive)

Date : .../.../... Signature :

Autorisation de diffusion : Oui  Non

**Devant le jury :**

Soutenu à Angers le \* : 19/04/2013

Sous la présidence de \* : Josiane LE CORFF

Maître de stage \* : Jean-Marc AUDERGON & Anne BERNOLE

Enseignant référent : Claudie LAMBERT

Autres membres du jury : Charles-Éric DUREL, Béatrice TEULAT-MERAH, Arnaud REMAY

*\* 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 à remercier le laboratoire BioGEVES et l'équipe G2IP de l'INRA d'Avignon pour leur accueil. Plus particulièrement, je remercie Bénédicte Jaudeau, Arnaud Remay, Anne Bernole, Jean-Marc Audergon et Guillaume Roch pour l'aide qu'ils m'ont apportée et la patience dont ils ont fait preuve à mon égard. Je remercie également Claudie Lambert pour ses conseils et ses encouragements.

## Glossaire

ADN	Acide DéoxyriboNucléique
AFC	Analyse Factorielle des Correspondances
CTIFL	Centre Technique Interprofessionnel des Fruits et Légumes
DGAP	Département de Génétique et d'Amélioration des Plantes
DHS	Distinction, Homogénéité, Stabilité
dNTP	désoxyribonucléotide
FAOSTAT	Food and Agriculture Organization Statistic
GAFL	Génétique et Amélioration des Fruits et Légumes
GEVES	Groupe d'Étude et de Contrôle des Variétés et des Semences
G2IP	Génétique Intégrative et Innovation chez les <i>Prunus</i>
GNIS	Groupement National Interprofessionnel des Semences
INRA	Institut National de la Recherche Agronomique
OCVV	Office Communautaire des Variétés Végétales
PCR	Polymerase Chain Reaction
PPV	Plum Pox Virus
RNase	RiboNucléase
SEV	Secteur d'Étude des Variétés
SNES	Station Nationale d'Essais de Semences
SNP	Single Nucleotid Polymorphism
SOC	Service Officiel de Contrôle et de Certification
SSR	Single Sequence Repeats
UPOV	Union internationale pour la Protection des Obtentions Végétales

# Table des matières

<b>1.Introduction.....</b>	<b>4</b>
<b>2.Présentation des structures d'accueil.....</b>	<b>5</b>
2.1 <i>Présentation de BioGEVES.....</i>	5
2.2 <i>Présentation de l'unité GAFL de l'INRA d'Avignon.....</i>	6
<b>3.Synthèse bibliographique.....</b>	<b>6</b>
3.1 <i>L'abricotier.....</i>	6
3.1.1 Taxonomie et biologie de l'espèce.....	6
3.1.2 Origine et diversité de l'espèce.....	7
3.1.3 Importance économique.....	8
3.1.4 Objectifs de la sélection variétale.....	9
3.2 <i>Diversité et structure génétique.....</i>	10
3.2.1 Utilisation de marqueurs microsatellites.....	10
3.2.1.1 Définition et applications.....	10
3.2.1.2 Identification et développement de microsatellites chez les Prunus.....	11
3.2.1.3 Synténie et transférabilité des microsatellites au sein du genre Prunus.....	11
3.2.2 Inférence de la structure génétique.....	11
3.2.2.1 Définition .....	11
3.2.2.2 Influence du mode de reproduction.....	12
3.2.2.3 Applications de la structure génétique.....	12
<b>4.Matériels et méthodes.....</b>	<b>13</b>
4.1 <i>Matériel végétal.....</i>	13
4.2 <i>Extraction et amplification de l'ADN.....</i>	15
4.3 <i>Analyse des produits de PCR.....</i>	20
4.4 <i>Analyses statistiques des données.....</i>	20
4.4.1 Harmonisation des données.....	20
4.4.2 Paramètres de diversité génétique.....	21
4.4.3 Analyse factorielle des correspondances.....	22
4.4.4 Analyse Bayésienne de regroupement.....	22
<b>5.Résultats et discussion.....</b>	<b>22</b>
5.1 <i>Analyse de la diversité observée.....</i>	22
5.1.1 Diversité allélique des accessions patrimoniales et issues de prospections.....	22
5.1.2 Diversité allélique des accessions issues d'amélioration génétique.....	24
5.1.3 Comparaison entre diversité patrimoniale et diversité hybride, et comparaison avec une étude antérieure.....	26
5.1.4 Diversité allélique patrimoniale intra et inter groupes géographiques.....	27
5.1.5 Pertinence des marqueurs utilisés et positionnement dans le génome.....	29
5.1.6 Diversité variétale.....	32
5.2 <i>Structure phylogéographique des accessions patrimoniales.....</i>	33
5.2.1 Analyses factorielles des correspondances .....	33
5.2.1.1 AFC concernant les accessions patrimoniales.....	33
5.2.1.2 Comparaison avec les accessions issues d'amélioration.....	33
5.2.2 Analyse Bayésienne de regroupement.....	38
<b>6.Conclusion.....</b>	<b>45</b>
<b>7.Bibliographie.....</b>	<b>46</b>
7.1 <i>Liste bibliographique.....</i>	46
7.2 <i>Liste sitographique.....</i>	53

# 1.Introduction

L'abricotier commun (*Prunus armeniaca* L.) appartient au genre *Prunus* L. de la famille des *Rosaceae*. *Prunus armeniaca* est une espèce diploïde qui possède 8 paires de chromosomes ( $2n=16$ ). Elle est originaire de Chine, où elle est cultivée depuis 2000 ans, et a été introduite autour du bassin méditerranéen vers le premier siècle avant J.C, par l'intermédiaire de l'Iran et de l'Arménie. Aujourd'hui, le Bassin Méditerranéen représente plus de 50 % de la production mondiale d'abricots. La domestication de l'abricotier s'est accompagnée d'une réduction de la diversité génétique de l'espèce, selon un gradient Est-Ouest. Après une longue période caractérisée par la seule valorisation des ressources génétiques locales, dans les années 1970 ont été engagés des programmes d'amélioration qui ont permis de développer de nouvelles variétés. Ces programmes se sont intensifiés au cours des 20 dernières années et aboutissent à des flux variétaux de l'ordre de 20 variétés par an, posant par là même la question de leur identification.

L'analyse de la structuration phylogéographique de l'abricotier a été étudiée à l'échelle du Bassin Méditerranéen (Hagen et al. 2002, Khadari et al., 2006, Bourguiba et al., 2010, Bourguiba et al., 2012a,b). Elle fait apparaître une véritable structure phylogéographique organisée en 3 phylums fortement ancrée sur l'origine géographique des accessions. Cette analyse restait malheureusement circonscrite au Bassin méditerranéen au sens large et demandait à être complétée par l'incorporation d'accessions en provenance d'Amérique, d'Europe continentale, d'Asie centrale et d'Asie de l'Est.

Dans le présent travail, nous nous proposons, en utilisant des profils de marqueurs microsatellites cartographiés sur la carte de référence des *Prunus* :

- de géotyper les principales variétés cultivées dans une logique de caractérisation variétale et d'en étudier les limites.
- d'examiner la structure génétique de l'espèce abricotier sur une base géographique et génétique large par le moyen d'analyses Bayésiennes de regroupement. Il s'agit plus particulièrement d'examiner la structuration de la diversité des ressources patrimoniales de l'espèce et d'évaluer les conséquences associées :
  - sur le maintien de la diversité
  - sur la caractérisation des accessions
  - sur la conduite des programmes d'innovation variétale.

Afin de répondre à ces objectifs, nous présenterons tout d'abord les structures d'accueil, puis une synthèse bibliographique rendant compte des connaissances sur le sujet. Par la suite, nous détaillerons les matériels et méthodes utilisées, et nous présenterons les résultats obtenus accompagnés d'une discussion.

## 2.Présentation des structures d'accueil

Ce stage présente la particularité de s'être déroulé en deux temps sur deux structures différentes. Les expérimentations ont été réalisées au laboratoire BioGEVES, situé sur le domaine du Magneraud du centre INRA Poitou-Charentes (Surgères, Charente-Maritime). L'analyse des données s'est déroulée au sein de l'unité Génétique et Amélioration des Fruits et Légumes (GAFL) de l'Institut National de la Recherche Agronomique (INRA), sur le domaine Saint Maurice du centre INRA Provence-Alpes-Côte d'Azur (Avignon, Vaucluse).

### 2.1 Présentation de BioGEVES

BioGEVES, présent sur deux sites au niveau national (Surgères et Beaucouzé), constitue le laboratoire de marquage génétique et de biochimie du Groupe d'Etude et de Contrôle des Variétés et des Semences (GEVES). Le GEVES est un groupement d'intérêt public (GIP) dont les administrateurs sont l'INRA, le Groupement National Interprofessionnel des Semences (GNIS) et le ministère en charge de l'agriculture. Le GEVES a pour mission officielle de mener les études nécessaires à :

- **L'inscription des variétés** végétales nouvelles au catalogue officiel
- **La protection** juridique du droit des obtenteurs
- **La certification des semences** avant leur commercialisation, dans le cas des espèces soumises à une certification réglementaire

Le laboratoire BioGEVES a entre autres pour mission de réaliser des analyses moléculaires et biochimiques sur les variétés et les semences. Plus précisément, il est chargé de :

- Caractériser des variétés
- Contrôler l'identité, la pureté variétale
- Contrôler la présence de gènes de résistance
- Évaluer des caractères technologiques
- Détecter des séquences spécifiques (OGM, pathogènes, gènes d'intérêt)

Les analyses réalisées sont destinées principalement au Secteur d'Etude des Variétés (SEV) et à la Station Nationale d'Essais de Semences (SNES), dans le cadre des missions globales du GEVES. Les autres destinataires sont le Service Officiel de Contrôle et de Certification (SOC), la Direction Générale de l'Alimentation (DGAL), des cabinets d'expertise et les acteurs de la filière (coopératives, entreprises semencières...).

## 2.2 Présentation de l'unité GAFL de l'INRA d'Avignon

L'INRA est un établissement public à caractère scientifique et technologique, sous la double tutelle des ministères de l'Agriculture et de la Recherche. Le département Biologie et Amélioration des Plantes (BAP) - anciennement Département de Génétique et d'Amélioration des Plantes (DGAP) - est constitué de 40 unités couvrant l'ensemble du territoire. Parmi elles figurent des unités de recherche, des unités expérimentales, disposant de réseaux d'expérimentation de plein champs, et des plates-formes technologiques.

L'unité de recherche Génétique et Amélioration des Fruits et Légumes (GAFL) de Montfavet travaille sur des espèces maraîchères et fruitières : aubergine, laitue, melon, piment, pomme de terre, tomate, pêche, abricot et leurs porte-greffes. Les axes de recherche abordés sont la diversité génétique, la résistance aux bioagresseurs et l'innovation variétale. L'unité GAFL est constituée de six équipes dont les missions et les espèces étudiées diffèrent. L'équipe Génétique Intégrative et Innovation chez les Prunus (G2IP) regroupe l'ensemble des travaux de recherche et d'innovation variétale sur le pêcher, l'abricotier et leurs porte-greffes. Cela permet d'intégrer les résultats des programmes de recherches menés sur la qualité et les résistances aux bioagresseurs aux travaux d'innovation variétale.

## 3.Synthèse bibliographique

### 3.1 L'abricotier

#### 3.1.1 Taxonomie et biologie de l'espèce

L'abricotier (*Prunus armeniaca* L.) appartient à la famille des *Rosaceae*, à la sous-famille des *Prunoideae*, au genre *Prunus* L., au sous-genre *Prunophora* (Neck) Focke et à la section *Armeniaca* (Lam.) Koch (Rehder, 1949).

Sous le terme générique « abricotier » on retrouve différentes espèces (Rehder, 1940 ; Mehlenbacher et al, 1991) : *Prunus sibirica* L. (abricotier de Sibérie), *Prunus mume* Sieb. et Zucc. (abricotier japonais), *Prunus brigantia* Vill. (prunier des Alpes), *Prunus holosericea* Batal. (abricotier tibétain), *Prunus mandshurica* (Maxim.) Koehne (abricotier de Mandchourie), et *Prunus dasycarpa* Ehr. (abricotier noir).

L'abricotier est une espèce diploïde comportant huit paires de chromosomes ( $2n = 16$ ) (Hormaza, 2002) et un génome de petite taille (294 Mbp/n) (Arumuganathan et Earle, 1991), fortement synténique avec le génome du pêcher aujourd'hui séquencé (Cipriani et al., 1999 ; Direlewanger et al., 2002 ; Dondini et al., 2007), elle est toutefois beaucoup plus hétérozygote.

Espèce adaptée aux climats tempérés, elle est présente dans des régions très différentes, des oasis sahariennes aux grands lacs canadiens, ce qui traduit un potentiel adaptatif particulièrement large, et des besoins en froid très variés (forts en Europe centrale et au Canada, particulièrement faibles dans les pays du Maghreb). Les variétés traditionnellement cultivées présentent généralement une aire d'adaptation pédoclimatique très limitée (Lichou et Jay, 2012 ; Bourguiba, 2012 ; Bourguiba et al., 2012a,b).

Espèce forestière en Chine, dans certaines régions d'Asie centrale, voire dans les oasis du Maghreb, où elle sert d'ombrage, c'est aussi une espèce ornementale au Japon et en Chine. Elle est cultivée pour ses fruits et pour ses noyaux en Asie centrale, elle ne devient un fruit de dessert qu'au niveau du bassin méditerranéen et dans les pays occidentaux.

Son régime de multiplication dépend des régions et des objectifs, il a longtemps reposé sur le semis, d'autant que la multiplication végétative par bouture est particulièrement difficile et le greffage compliqué par la présence de fortes teneurs en polyphénols. Dans les vergers modernes la multiplication se fait principalement par greffage, ce qui permet notamment une conservation des caractères entre les générations et l'établissement de vergers homogènes. Les porte-greffes utilisés sont principalement des abricotiers, de pêchers et des pruniers (Bourguiba, 2012).

### 3.1.2 Origine et diversité de l'espèce

*Prunus armeniaca* est probablement originaire de Chine, où l'on trouve des forêts sauvages d'abricotiers (Bailey et Hough, 1975; Faust et al., 1998). Vavilov (1992) a décrit trois centres d'origine de l'abricotier cultivé : la Chine (incluant le Tibet), l'Asie centrale et le Proche-Orient (pays du Caucase, Iran, Turquie). L'espèce aurait été disséminée depuis la Chine vers le Moyen-Orient en suivant la route de la soie (Vavilov, 1951). L'introduction de l'abricotier en Europe aurait suivie deux itinéraires historiques différents (Mehlenbacher et al., 1991 ; Faust et al., 1998). Le plus ancien a été suivi par les Arabes vers l'an 1000, de la zone Irano-Caucasienne au nord de l'Afrique d'où il a rejoint le sud de l'Europe. Le plus récent, 440 ans plus tard, commence en Asie Centrale pour arriver en Europe de l'Est (notamment en Hongrie) puis en Europe occidentale.

Les variétés d'abricotier peuvent être classées en plusieurs groupes éco-géographiques, qui diffèrent selon les auteurs. Kostina (1969) en a identifiés quatre : groupe d'Asie Centrale, groupe Irano-Caucasien, groupe Européen et groupe Dzungar-Zailij. Deux groupes ont été ajoutés à cette classification par Layne et al. (1996), il s'agit des groupes de Chine du Nord et de Chine de l'Est.

Le groupe d'Asie Centrale est le plus ancien et le plus diversifié, il comporte des variétés auto-incompatibles pour la plupart. Le groupe Dzungar-Zailij comporte principalement des variétés à petits fruits (Mehlenbacher et al., 1991; Layne et al., 1996; Faust et al., 1998), originaires de régions du Kazakhstan et de Chine. Le groupe Irano-Caucasien comporte majoritairement des variétés auto-incompatibles avec des besoins en froid peu importants originaires des pays du Caucase, d'Iran, d'Irak, d'Afrique du Nord, et pour certaines d'Europe du Sud (Hormaza, 2002). Le groupe Européen est le plus récent et le moins diversifié, il comporte principalement des variétés auto-compatibles (Hormaza, 2002 ; Romero et al., 2003).

Les variétés d'abricotier ainsi classées présentent un gradient décroissant de diversité génétique au sein du Bassin Méditerranéen partant de l'Est (Turquie-Iran) pour se diriger vers le Sud-Ouest (Maroc) (Hagen et al., 2002). Les récents travaux conduits par Bourguiba et al. (2012b) ont ainsi permis d'estimer à 45% la perte de variabilité allélique sur ce gradient étudié au sein des variétés patrimoniales. Ils ont aussi permis de différencier 3 groupes phylogéographiques en dehors de la zone de diversification irano-caucasienne, correspondant à un phylum Central Européen, un phylum Sud Méditerranéen et un phylum transversal caractérisé par une adaptabilité large mais des qualités organoleptiques limitées.



### 3.1.3 Importance économique

L'abricotier est une espèce fruitière cultivée dans le monde entier. L'abricot est le troisième fruit à noyau le plus produit, après la pêche et la prune (Bourguiba, 2012), avec une production mondiale de l'ordre de 3.6 millions de tonnes. Parmi l'ensemble des fruits tempérés, il représente la septième production mondiale (Lichou et Jay, 2012).

La majorité de la production est assurée par un petit nombre de pays. Ainsi, les 15 premiers pays producteurs représentent 80 % de la production (Tableau1). Le bassin méditerranéen est une zone fortement productrice, avec près de la moitié de la production mondiale, grâce à des pays comme la Turquie (premier pays producteur, 17 % de la production mondiale), l'Italie, l'Algérie, la France, le Maroc ou l'Egypte.

**Tableau 1 : Les principaux pays producteurs d'abricot (en tonnes, moyenne 2008-2010). En gras, les pays du bassin méditerranéen. (source FAOSTAT, 2010)**

<b>Pays</b>	<b>Production</b>
Turquie	640690
Iran	419716
Ouzbékistan	294000
<b>Italie</b>	224502
Pakistan	217291
<b>Algérie</b>	204972
<b>France</b>	164975
<b>Maroc</b>	126404
Japon	109533
<b>Egypte</b>	103949
<b>Syrie</b>	97838
<b>Espagne</b>	93736
Chine	87566
Ukraine	79833
<b>Grèce</b>	72162
Etats-Unis	65532
Afghanistan	61152
Arménie	57134
Russie	56167
Afrique du Sud	50721

L'innovation variétale s'est dynamisée de manière importante depuis une vingtaine d'années. Ainsi, en France, on est passé de 14 nouvelles variétés inscrites au catalogue officiel durant la décennie 1980, à 31 durant la décennie 1990 et à 55 durant la décennie 2000 (Figure 1).

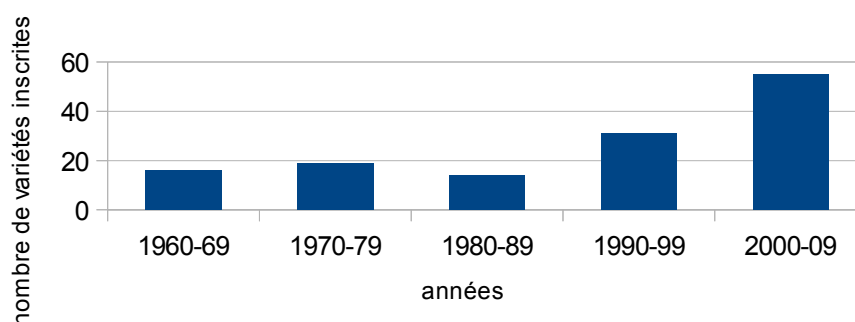


Figure 1: Évolution du nombre de variétés inscrites au catalogue officiel français

### 3.1.4 Objectifs de la sélection variétale

Traditionnellement, les variétés d'abricotier cultivées dans une région donnée sont peu nombreuses et fortement spécifiques des conditions climatiques et écologiques (Layne et *al.*, 1996 ; Lichou et Jay, 2012). Ainsi la plupart des programmes de sélection en Europe étaient basés sur l'utilisation de variétés locales (Audergon, 1995), qui présentent une base génétique étroite. Ils visaient un élargissement des gammes variétales.

Aujourd'hui, les programmes de sélection incorporent plus facilement des variétés issues d'autres groupes éco-géographiques (Hagen et *al.*, 2002). Cet élargissement du matériel génétique utilisé doit permettre de répondre à trois grands objectifs de sélection (Hormaza, 2002) : l'extension et la diversification des gammes variétales, la régularité de production et la qualité des fruits, et la résistance aux stress biotiques.

A cet effet, les nouvelles variétés doivent avoir une **adaptabilité plus large au milieu de culture, caractérisée par une plus grande régularité de production et une productivité satisfaisante.**

Elles doivent aussi présenter une moindre sensibilité aux **bio-agresseurs**. Un focus existe actuellement sur le **virus de la sharka** (plum pox virus, PPV), qui est largement répandu en Europe et qui peut être un facteur limitant de la culture (Vilanova et *al.*, 2003). Le PPV, transmis principalement par pucerons, est le virus le plus pénalisant affectant les productions de fruits à noyaux en Europe. Il provoque des dégradations importantes au niveau des fruits, qui sont alors invendables. L'éradication systématique des plants sources infectés n'a pas permis d'arrêter la propagation de la maladie. Pour résoudre ce problème les programmes de sélection intègrent des sources présentant une résistance à la sharka, pour la plupart originaires d'Amérique du Nord (Vilanova et *al.*, 2003). Les travaux sur la résistance portent aussi sur d'autres cibles comme le dépérissement bactérien, ou le monilia..

Un autre objectif majeur de sélection est l'amélioration des **qualités gustatives du fruit**. Les programmes de sélection s'intéressent ainsi tout particulièrement au rôle joué par les acides et les sucres dans l'expression de la saveur du fruit (Lachkar *et al.*, 2012), et à l'influence de l'éthylène sur la vitesse de maturation des fruits.

Les travaux relatifs à la diversité génétique constituent une étape importante des programmes, en amont de la sélection. Elles permettent d'améliorer l'utilisation des nouveaux génotypes intégrés aux programmes de sélection, et de définir des nouveaux croisements (Sanchez-Perez *et al.*, 2005). La diversité génétique au sein d'une espèce peut être analysée grâce aux marqueurs moléculaires. Les profils génétiques générés au cours d'études de ce type peuvent aussi être très utiles aux programmes de certification, pour identifier et protéger les nouvelles variétés produites.

## ***3.2 Diversité et structure génétique***

### **3.2.1 Utilisation de marqueurs microsatellites**

#### ***3.2.1.1 Définition et applications***

Les marqueurs moléculaires sont utilisés lors des études de diversité génétique car ils ne sont pas modifiés par des variations environnementales, contrairement aux caractères morphologiques (Ricciardi *et al.*, 2002). Les isozymes ont été les premiers marqueurs moléculaires utilisés dans la sélection variétale ( Tanksley et Orton, 1983), notamment chez l'abricotier (Badenes *et al.*, 1996). Les marqueurs moléculaires basés sur l'ADN sont aujourd'hui davantage utilisés car ils permettent d'étudier tout le génome, régions codantes comme régions non codantes, et donc révèlent un plus grand polymorphisme (Hormaza, 2002).

Parmi les marqueurs basés sur l'ADN, la diversité génétique d'une espèce peut être étudiée par les polymorphismes de longueur des fragments de restriction (RFLP), les polymorphismes de longueur de fragments amplifiés (AFLP), les polymorphismes au niveau nucléotidique (SNP) ou les répétitions de séquences simples non codantes (SSR ou microsatellites).

Les marqueurs microsatellites consistent en des séquences d'ADN longues de quelques nucléotides (2 à 6 paires de bases) répétées plusieurs fois en tandem. Ils présentent un polymorphisme très important, qui repose sur la variation du nombre de motifs répétés constituant le microsatellite. Les marqueurs SSR sont co-dominants, facilement amplifiables par PCR et sont répartis de manière homogène sur tout le génome (Powell *et al.*, 1996).

Les marqueurs microsatellites ont différentes applications agronomiques. Ils sont utilisés notamment pour :

- L'établissement de l'empreinte génétique d'un individu ou d'une variété, en vue de son inscription ou sa protection.
- L'évaluation de la diversité génétique d'une espèce ou d'un groupe variétal
- L'établissement de la cartographie des génomes.
- La mise en évidence ou le suivi des gènes impliqués dans l'expression de caractères d'intérêt agronomique ou technologique.
- La sélection assistée par marqueurs (SAM).

### **3.2.1.2 Identification et développement de microsatellites chez les *Prunus***

Les amorces PCR sont construites par reproduction des régions flanquantes des microsatellites, qui sont spécifiques d'un locus donné.

Le premier isolement de marqueurs microsatellites chez les *Prunus* a été réalisé par Cipriani *et al.* (1999) à partir de deux banques d'ADN génomique de pêcher. Depuis, de nombreuses études ont permis le développement de ce type de marqueurs chez l'ensemble des espèces fruitières importantes du genre *Prunus* :

- le pêcher (Testolin *et al.*, 2000 ; Sosinski *et al.*, 2000 ; Aranzana *et al.*, 2002 ; Dirlewanger *et al.*, 2002 ; Yamamoto *et al.*, 2002)
- l'amandier (Testolin *et al.*, 2004 ; Mnejja *et al.*, 2005 ; Joobeur *et al.*, 2000)
- le cerisier (Downey et Iezzoni, 2000 ; Cantini *et al.*, 2001 ; Clarke et Tobutt, 2003 ; Struss *et al.*, 2003 ; Vaughan et Russel, 2004) ;
- le prunier (Mnejja *et al.*, 2004) ;
- l'abricotier (Lopes *et al.*, 2002 ; Decroocq *et al.*, 2003 ; Messina *et al.*, 2004 ; Hagen *et al.*, 2004 ; Vilanova *et al.*, 2006).

### **3.2.1.3 Synténie et transférabilité des microsatellites au sein du genre *Prunus***

La synténie se définit comme la conservation de l'ordre des gènes et des marqueurs entre les cartographies ou séquences de génomes de plantes différentes (Morot-Gaudry et Briat, 2007).

Chez deux espèces présentant un niveau de synténie suffisamment élevé, il peut y avoir transférabilité des marqueurs microsatellites : les microsatellites isolés à partir du génome d'une première espèce conservent leur pouvoir d'amplification et présentent le même polymorphisme chez l'autre espèce. De manière générale, plus les espèces sont proches, plus les microsatellites ont un niveau de transférabilité élevé.

Au sein du genre *Prunus*, plusieurs études ont montré un niveau de transférabilité élevé de la plupart des microsatellites isolés (Dirlewanger *et al.*, 2002 ; Messina *et al.*, 2004 ; Wunsch, 2008). Ainsi de nombreuses études ont montré l'efficacité de marqueurs microsatellites provenant du génome du pêcher pour étudier l'abricotier (Hormaza, 2002 ; Romero *et al.*, 2003 ; Zhebentyayeva *et al.*, 2003 ; Maghuly *et al.*, 2005 ; Sanchez-Perez *et al.*, 2005 ; Dondini *et al.*, 2007 ; Ruthner *et al.*, 2006 ; He *et al.*, 2007 ; Bourguiba *et al.*, 2010).

## **3.2.2 Inférence de la structure génétique**

### **3.2.2.1 Définition**

L'établissement de la structure génétique d'une population consiste à déterminer le nombre de groupes génétiques différents au sein de cette population, les individus composant chacun de ces groupes, et éventuellement les individus ayant des origines dans plusieurs groupes. Plusieurs méthodes statistiques permettent d'effectuer les regroupements à partir de données génétiques, comme l'Analyse Factorielle des Correspondances (AFC) ou l'établissement d'arbres phylogénétiques par l'application d'algorithmes de type Neighbor-Joining (Saitou et Nei, 1987).

Une autre approche plus récente consiste à supposer que les données génétiques peuvent être expliquées par un modèle probabiliste de type Bayésien dont les paramètres sont inconnus. De nombreuses méthodes actuelles sont basées sur le modèle bayésien de regroupement, proposé pour la première fois par Pritchard *et al.* (2000) et implémenté dans le programme STRUCTURE. Ce programme permet d'estimer conjointement les fréquences alléliques au sein des différentes populations et les coefficients d'appartenance des individus à chaque groupe (Jay, 2011).

### **3.2.2.2 Influence du mode de reproduction**

L'abricotier présente à la fois des variétés auto-compatibles et auto-incompatibles, suivant le groupe éco-géographique auxquelles elles appartiennent. Un mode de reproduction autogame a tendance à fixer les caractères d'une génération à l'autre. Ainsi il y a une homogénéisation des génomes au sein d'un groupe, et une augmentation de la différenciation génétique entre les groupes. A l'inverse, un mode de reproduction allogame favorise les recombinaisons et donc a tendance à diminuer les différences génétiques entre les groupes.

Chez les espèces présentant ces deux modes de reproduction, comme l'abricotier, les proportions de variétés autogames et allogames devraient être un des facteurs déterminant la structure génétique (Loveless et Hamrick, 1984).

### **3.2.2.3 Applications de la structure génétique**

La génétique d'association est une méthode permettant d'identifier des gènes qui contribuent à la variation de caractères. Elle nous intéresse tout particulièrement dans le cas de l'abricotier du fait de la diversité phénotypique observée. Elle consiste à chercher des associations statistiques entre les variations du génotype et du phénotype, au sein de populations présentant une forte diversité. La génétique d'association s'appuie sur le déséquilibre de liaison (DL), qui rend compte de l'association non aléatoire entre des allèles situés à des locus différents.

Le DL peut être dû à une liaison physique entre les allèles concernés, mais également à des forces évolutives liées à l'histoire de l'espèce et qui influent directement sa structure : mutation, migration et sélection (Rafalski et Morgante, 2004 ; Flint-Garcia *et al.*, 2003). Pour pouvoir être utilisé en cartographie, le DL doit être dû uniquement à une liaison physique, et ne doit pas dépendre de la structure de population. C'est pourquoi les analyses sur la structure génétique d'une population sont importantes en amont des études de DL.

L'établissement de la structure génétique d'une population se fait souvent à partir d'effectifs importants, afin que toute la diversité de la population soit représentée. Or, l'augmentation des effectifs des collections entraînent des problèmes de gestion de l'espace. L'idée de réduire les effectifs tout en conservant ou maximisant la diversité d'une collection a été émise par Brown (1989). Ces « core collections » doivent représenter la diversité génétique des collections d'origine, en renfermant au moins 80% de la variabilité.

De nombreuses études ont eu pour objectif de développer des « core collections » chez différentes espèces, par exemple le pois (Logozzo *et al.*, 2007), le soja (Wang *et al.*, 2006), la canne à sucre (Nair *et al.*, 1998), et récemment l'abricotier (Krichen *et al.*, 2012). La caractérisation de ces core-collections vise à s'assurer de la maîtrise et du maintien de la diversité génétique, en même temps qu'elle va permettre d'étudier de manière plus fine les relations entre génotype et environnement, notamment lorsqu'elles sont partagées dans des conditions pédoclimatiques contrastées.

## 4. Matériels et méthodes

### 4.1 Matériel végétal

Dans le cadre de cette étude, 1253 accessions d'abricotiers de différentes origines géographiques ont été analysées (Figure 2a et 2b). Ces accessions sont issues de différentes collections nationales et de campagnes de prospection. On différencie, au sein des collections, les accessions dites patrimoniales ou dérivées de prospections, pour lesquelles le lien avec la région d'origine est ancien, et les accessions hybrides, issues de programmes d'amélioration.

En vue des analyses de diversité génétique, un regroupement des pays en zones géographiques a été effectué. Cela permet de simplifier les analyses et de repérer des grandes tendances, en passant d'une quarantaine de pays à six grandes zones géographiques : l'Amérique (principalement États-Unis et Canada); l'Europe méditerranéenne (France, Italie, Espagne et Grèce); le Maghreb (Algérie, Maroc et Tunisie); l'Europe centrale (Roumanie, République Tchèque, Slovaquie pour les pays les plus représentés) ; l'Asie centrale (Turquie, Iran, Afghanistan, Kazakhstan, Pakistan, Turkménistan, Russie); l'Asie de l'Est (Chine et Japon). D'autres accessions en nombre beaucoup moins important proviennent de Nouvelle-Zélande, d'Australie et d'Afrique du Sud. La moitié des accessions est originaire d'Europe méditerranéenne (444) et du Maghreb (218). L'autre moitié est principalement constituée d'accessions originaires d'Asie centrale (178), d'Asie de l'Est (57), d'Amérique du Nord (180) et d'Europe continentale (134). Les pays les plus représentés sont la France (248 accessions), les États-Unis (145), l'Italie (120), la Tunisie (90), le Maroc (81), l'Espagne (65), la Turquie (57) et l'Algérie (46). Les variétés hybrides, au nombre de 419, sont principalement originaires d'Europe méditerranéenne (193) et d'Amérique du Nord (147).

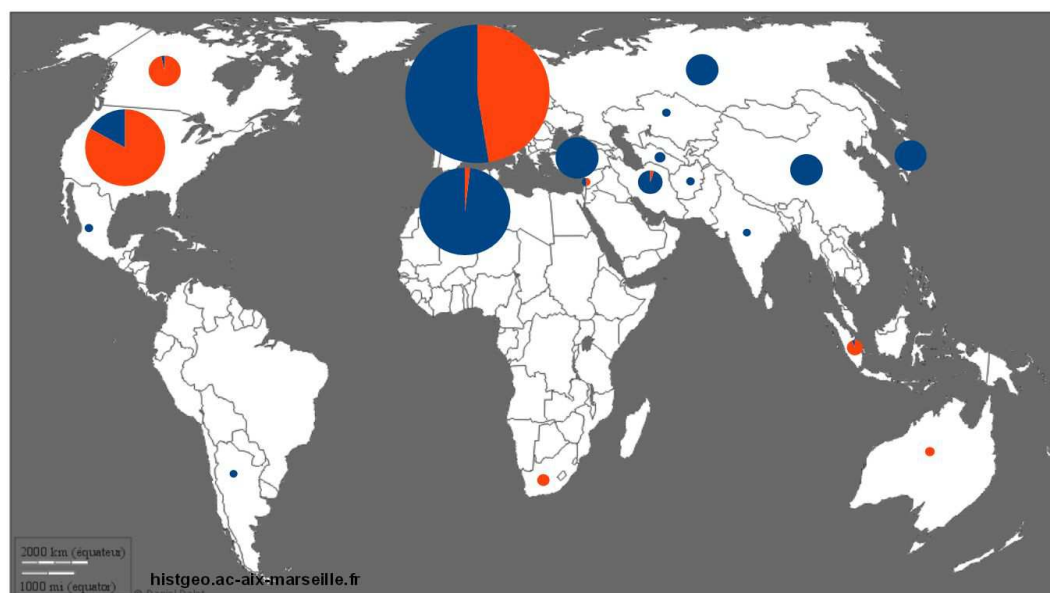


Figure 2a : Origine géographique des accessions. En bleu, les accessions patrimoniales et issues de prospections ; en orange, les accessions hybrides.

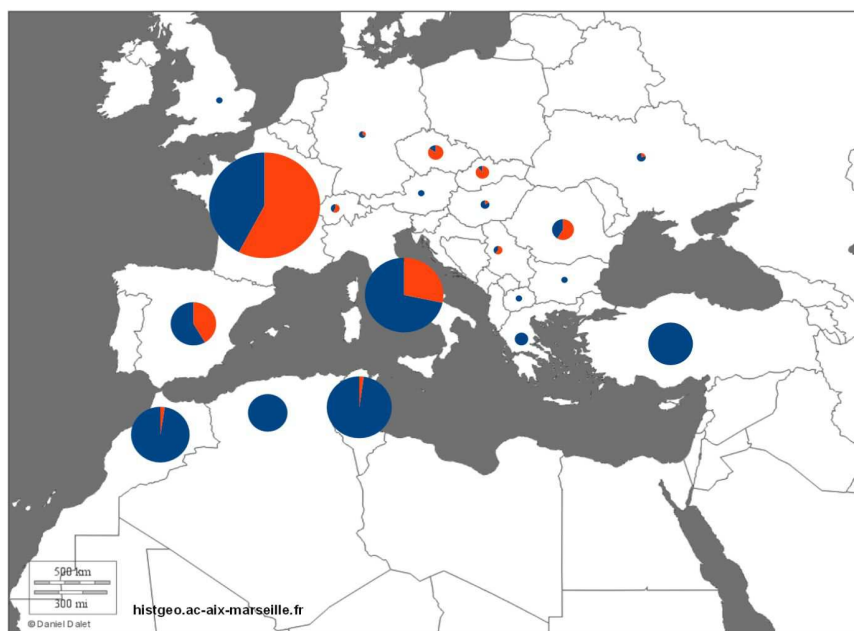


Figure 2b : Origine géographique des accessions issues d'Europe et du Maghreb. En bleu, les accessions patrimoniales et issues de prospections ; en orange, les accessions hybrides.

Il est important de noter que le génotypage des 1253 accessions a été fait en plusieurs temps, et sur plusieurs plateformes différentes (Tableau 2). Les accessions que j'ai génotypées, au nombre de 327, sont issues de la collection de l'INRA et sont originaires pour la plupart de France, d'Italie, des Etats-Unis et d'Europe centrale. Par la suite, 253 autres accessions ont été également génotypées au laboratoire BioGEVES puis ajoutées au matériel total. Ces dernières accessions proviennent du Japon, des Etats-Unis, de Suisse, d'Espagne et d'Asie centrale (principalement Russie/ex-URSS, Pakistan, Turkménistan). Les conditions d'extraction / d'amplification de l'ADN et de traitement des produits de PCR, décrites dans les paragraphes suivants, concernent les 580 accessions génotypées au laboratoire BioGEVES. Un ensemble commun de 25 marqueurs microsatellites (Figure 3) a été utilisé pour l'ensemble des 1253 accessions, ce qui rend possible une analyse commune de la totalité du matériel caractérisé.

Tableau 2 : Répartition des accessions génotypées en fonction de la plateforme de génotypage

Lieu de génotypage	Nombre d'accessions	Origine géographique
Plate-forme INRA de Montpellier	584	Algérie, Maroc, Tunisie, France, Espagne, Italie, Turquie...
Plate-forme INRA GENTIANE de Clermont-Ferrand	89	Europe centrale, Asie centrale, Chine...
Laboratoire BioGEVES de Surgères	327 + 253	France, Italie, Espagne, USA, Japon, Europe centrale, Asie centrale

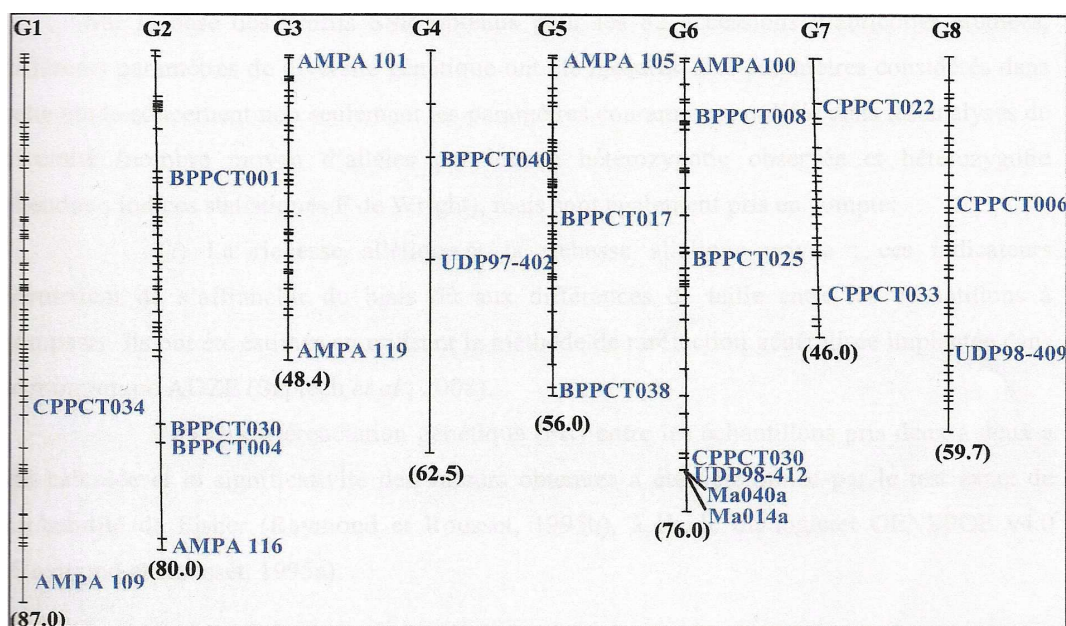


Figure 3 : Positionnement de 24 des 25 marqueurs microsatellites utilisés sur la carte de référence du génome des espèces du genre *Prunus* (Bourguiba *et al.*, 2012b). Le marqueur UDP96-018, non représenté ici, est situé à l'extrémité du groupe de liaison G1.

## 4.2 Extraction et amplification de l'ADN

Les échantillons de jeunes feuilles ont été prélevés en mai 2012, puis stockés à -80°C. Les feuilles ont été ensuite lyophilisées avant broyage (broyeur Retsch MM301).

Les ADN génomiques ont été extraits grâce au kit NucleoSpin 96 Plant II (Macherey-Nagel). Environ 20 mg de poudre sèche ont été mélangés dans chaque tube de 2 mL avec 400 µL de tampon PL2 (NaCl) et 10 µL de RNase. Les échantillons ont été incubés à 65°C pendant 30 min, puis mélangés avec 100 µL de tampon PL3 (solution de tampon acétique). Ils ont ensuite été incubés 5 min dans la glace, puis centrifugés à 6000 g pendant 5 min. Le surnageant de chaque échantillon a été ajouté à 450 µL de Buffer PC (Guanidine chlorhydrate), préalablement distribué dans une plaque 96 2.2 mL. Les échantillons ont été transférés dans une plaque de mini colonnes Nucleospin Plant II, puis centrifugés 5 min à 6000 g pour séparer l'ADN (qui se fixe à la membrane de silice des colonnes) du lysat. Les membranes ont ensuite été lavées en trois fois par l'ajout de tampon PW1, PW2 et PW3. Entre chaque lavage, la plaque d'ADN a été centrifugée quelques minutes à 6000g. La plaque de mini colonnes a été placée sur une plaque d'élution semi deepwell, puis 100 µL de tampon d'élution préchauffé à 65°C ont été distribués dans chaque puits. La plaque a ensuite été incubée à température ambiante et centrifugée à 6000g pendant 2 minutes. Cette dernière étape d'élution a été répétée une fois, puis la plaque a été scellée et stockée à 4°C, en vue de l'amplification. Préalablement à l'amplification, l'ADN extrait a été dilué au 1/5.



L'amplification a été faite par Polymerase Chain Reaction (PCR), en utilisant 25 marqueurs microsatellites issus des génomes du pêcher et de l'abricotier. Les 25 amorces microsatellites ont été isolées et testées chez diverses espèces du genre *Prunus*. Ces amorces ont été choisies de façon à couvrir l'ensemble du génome de manière homogène, en s'appuyant sur la carte de référence *Prunus* (Aranzana *et al.*, 2002 ; Bourguiba *et al.*, 2010 ; 2012 ; Cipriani *et al.*, 1999 ; Dirlewanger *et al.*, 2002 ; Hagen *et al.*, 2004 ; Testolin *et al.*, 2000 ; Yamamoto *et al.*, 2002). Quatre fluorochromes ont été utilisés pour marquer les amorces: VIC (vert), PET (rouge), FAM (bleu) et NED (jaune) ; un fluorochrome a été attribué à chacun des 25 marqueurs SSR. Le kit PCR utilisé est le Twelve Paq 10X PCR Buffer II, contenant l'ADN polymérase AmpliTaq Gold® (Applied Biosystem, Foster City, Californie).

Deux programmes d'amplification ont été utilisés.

Le programme « pêcher » (Figure 3) a été utilisé pour 17 marqueurs. Chaque mélange PCR destiné à ce programme, d'un volume final de 6 µL, contient :

- 10 ng d'ADN
- 1 µL de tampon 10X
- 0.2 µL de chaque dNTP (10mM)
- 0.6 µL de MgCl<sub>2</sub> (25mM)
- 0.10 µL d'amorce sens (1 µM)
- 0.10 µL d'amorce anti-sens (10 µM)
- 0.10 µL de fluorochrome (10 µM)
- 0.10 µL d'AmpliTaq Gold

Le programme PCR « pêcher » consiste en une dénaturation pendant 10 min à 94 °C suivie par 14 cycles de : (1) 30 s à 94°C, (2) 45 s à la température d'hybridation 62°C (cette température diminue de 0.5°C par cycle) et (3) 1 min d'élongation à 72°C. Suivent ensuite 25 cycles de : (1) 30 s à 94°C, (2) 45 s à 50°C et (3) 1 min à 72°C, puis une étape finale d'élongation de 30 min à 72°C. Après l'amplification les échantillons sont stockés à 4°C.

Le programme « abricotier » (Figure 4) a été créé à partir du programme « pêcher » afin d'amplifier certains marqueurs (8) dont l'amplification par le programme « pêcher » n'était pas optimale. Chaque mélange PCR destiné à ce programme, d'un volume final de 6  $\mu\text{L}$ , contient :

- 10 ng d'ADN
- 1  $\mu\text{L}$  de tampon 10X
- 0.125  $\mu\text{L}$  de chaque dNTP (10mM)
- 1.2  $\mu\text{L}$  de  $\text{MgCl}_2$  (25mM)
- 0.25  $\mu\text{L}$  d'amorce sens (1  $\mu\text{M}$ )
- 0.25  $\mu\text{L}$  d'amorce anti-sens (10  $\mu\text{M}$ )
- 0.25  $\mu\text{L}$  de fluorochrome (10  $\mu\text{M}$ )
- 0.10  $\mu\text{L}$  d'AmpliTaq Gold

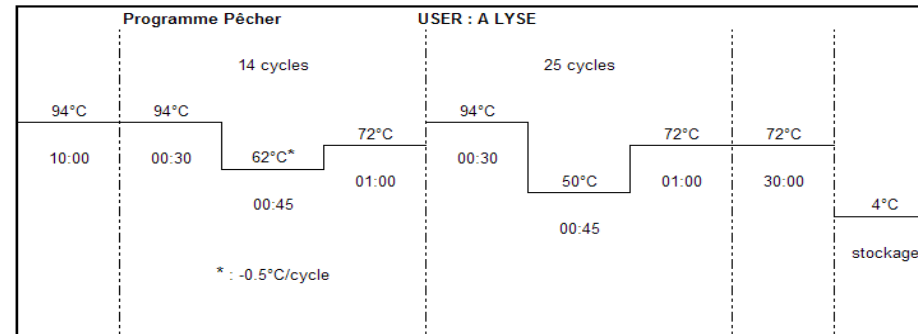
Le profil d'amplification consiste en une dénaturation pendant 10 min à 94 °C, suivie par 10 cycles de : (1) 30 s à 94°C, (2) 1 min à la température d'hybridation 61°C (cette température diminue de 1°C par cycle) et (3) 1 min d'élongation à 72°C. Suivent ensuite 32 cycles de : (1) 30 s à 94°C, (2) 1 min à 52°C, et (3) 1 min à 72°C, puis une étape finale d'élongation de 10 min à 72°C. Après l'amplification les échantillons sont stockés à 10°C.

Les réactions de PCR ont été effectuées avec les thermocycleurs GeneAmp PCR System 9700 et 9800 Fast Thermal Cycler (Applied Biosystems).

A

Simplex	Taq Gold+TP AB	
H2O up	3.80	399.00
TP 10X	1.00	105.00
dNTP (10mM)	0.20	21.00
MgCl2 (25 mM)	0.60	63.00
Oligo R (10 µM)	0.10	10.50
Oligo FZ (1 µM)	0.10	10.50
Oligo 35S* IRD (10 µM)	0.10	10.50
AmpliTaq Gold (5U/µl)	0.10	10.50
Vol mix 1X (µl)	6.00	
Vol mix n+10% (µl)		630.00
ADN (2,5ng/µl)		4 µl
Vol final PCR 1x	10 µl	
Nb d'échant n+10%		105

B



C

MARQUEUR	Couleur 35S
UDP97.402	FAM
UDP98.409	FAM
AMPA 101	NED
BPPCT030	NED
BPPCT040	NED
BPPCT001	NED
BPPCT038	NED
CPPCT033	PET
AMPA 116	PET
UDP98.412	PET
BPPCT004	VIC
AMPA 109	VIC
AMPA 119	VIC
CPPCT022	VIC
BPPCT017	VIC
BPPCT025	VIC
CPPCT006	VIC

Figure 4 :

A. Composition des mix PCR pour le programme « pêcher ». La première colonne indique les quantités de réactifs pour un échantillon ; la deuxième colonne indique les quantités de réactifs pour 105 échantillons (ce qui correspond à une plaque 96, plus 10% de marge).

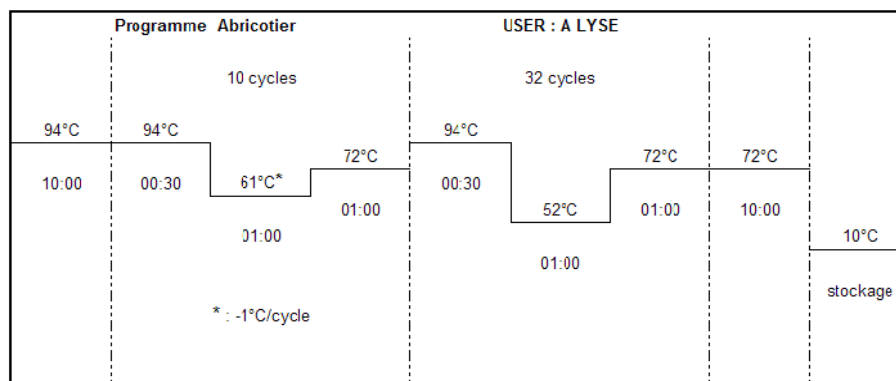
B. Description du programme PCR « pêcher ».

C. Liste des marqueurs SSR destinés au programme « pêcher », et fluorochromes correspondant.

A

Simplex	Taq Gold+TP AB	
H2O up	2.825	296.625
TP 10X	1.00	105
dNTP (10mM)	0.125	13.125
MgCl2 (25 mM)	1.20	126
Oligo R (10 $\mu$ M)	0.25	26.25
Oligo FZ (1 $\mu$ M)	0.25	26.25
Oligo 355* IRD (10 $\mu$ M)	0.25	26.25
Ampli Taq Gold (EU/ $\mu$ l)	0.10	10.5
<b>Vol mix 1X (<math>\mu</math>l)</b>	<b>6.00</b>	
<b>Vol mix n+10% (<math>\mu</math>l)</b>		<b>630.0</b>
ADN (1/10)		4 $\mu$ l
<b>Vol final PCR 1x</b>	<b>10 <math>\mu</math>l</b>	
Nb c'échant n+10%		<b>105</b>

B



C

MARQUEUR	Couleur 35S
MA040a	FAM
AMPA 105	FAM
CPPCT030	FAM
BPPCT008	FAM
CPPCT034	PET
UDP96-018	PET
AMPA 100	PET

Figure 5 :

A. Composition des mix PCR pour le programme « abricotier ». La première colonne indique les quantités de réactifs pour un échantillon ; la deuxième colonne indique les quantités de réactifs pour 105 échantillons (ce qui correspond à une plaque 96, plus 10% de marge).

B. Description du programme PCR « abricotier ».

C. Liste des marqueurs SSR destinés au programme « abricotier », et fluorochromes correspondants.

### ***4.3 Analyse des produits de PCR***

Avant analyse par le séquenceur, 4 µL des produits PCR de chaque échantillon sont mélangés avec 5.8 µL de formamide (qui permet la rupture des liaisons hydrogènes entre les bases d'ADN) et 0.2 µL de marqueur de taille GeneScan 500 LIZ (Applied Biosystems). Quand cela est possible, les dépôts se font par multiplexage. Les échantillons obtenus sont ensuite soumis à migration par le séquenceur de gène capillaire 3730 DNA Analyzer (Applied Biosystems).

Le logiciel Gene Mapper (Applied Biosystems) permet l'analyse des données brutes obtenues et les assigne à des données biologiques : les allèles sont visualisés sous forme de pics.

### ***4.4 Analyses statistiques des données***

#### **4.4.1 Harmonisation des données**

Une fois les allèles présents pour chaque marqueur identifiés grâce à Gene Mapper, on établit une matrice de données des allèles pour l'ensemble des accessions. Le génotypage de la totalité des accessions a été réalisé sur des plate-formes différentes. De ce fait, il existe des décalages entre les tailles alléliques déterminées au BioGEVES et celles déterminées sur les autres plateformes. Ce décalage d'environ 20 paires de bases varie selon les allèles (entre 17 et 23 paires de bases). Il était essentiel de déterminer la correspondance exacte entre les tailles alléliques du BioGEVES et les autres, afin d'obtenir une matrice de données cohérente. Pour ce faire, un certain nombre d'accessions a été génotypé sur les trois plate-formes. Une fois la correspondance entre les différentes tailles alléliques déterminée, un codage allélique commun aux trois plate-formes a été créé.

Les profils génétiques présentant un degré de ploïdie supérieur à deux n'ont pas été conservés dans la matrice de données finale, car les analyses statistiques ne peuvent s'appliquer qu'aux individus haploïdes ou diploïdes. Ces accessions, au nombre de cinq, correspondent à des porte-greffes ou des variétés résultant d'hybridations inter-spécifiques, ce qui explique leur degré élevé de ploïdie.

Les homonymes (accessions présentant un nom identique) ont été supprimés, sauf s'ils présentent des profils génétiques différents. Des noms identiques attribués à des accessions présentant des profils génétiques différents peut s'expliquer soit par une mauvaise identification de certaines accessions, soit par une variabilité au sein d'une variété.

Les accessions dont la mauvaise attribution du nom a pu être repérée grâce au génotypage sont conservées pour la suite de l'analyse mais sont signalées. Par exemple, quatre accessions de la variété « Stark Early Orange » ont été génotypées, dont une présente un profil génétique très différent des autres. Cette accession a été très probablement mal nommée, et son nom est changé en « Stark Early Orange faux » pour la suite de l'analyse. Cinq accessions sont dans ce cas, sur les 1253 au total. Des accessions « de référence », pour lesquelles on est sûr de la bonne attribution du nom, sont très utiles. Il peut arriver que seules deux accessions aux profils très différents et portant un même nom ont été génotypées, ou qu'il n'y ait pas d'accession « de référence » permettant une comparaison de profils. Dans ce cas, on ne peut déterminer à quel profil génétique correspond ce nom. Moins d'une dizaine d'accessions sont dans ce cas, elles correspondent à des variétés issues de collections différentes pour lesquelles un effort de validation devra être conduit.

Si les différences de profils constatées entre des homonymes sont beaucoup plus légères, cela peut être dû à une variabilité au sein d'une même variété. Ainsi, la variété Bergeron peut être considérée comme une «variété population», et de légères différences aussi bien phénotypiques que génotypiques peuvent exister entre les individus de cette variété, ce qui avait conduit l'organisme en charge de la certification à identifier deux clones pour la variété.

De légères différences entre des homonymes peuvent également être dues à des erreurs de lecture des allèles sur le logiciel Gene Mapper. Cette lecture a été faite à différents moments par des personnes différentes, le risque de décalage de lecture n'est donc pas négligeable. Cela est visible sur les profils lorsque, chez des homonymes, la différence pour un marqueur donné n'est que de une ou deux paires de base. Quand ce cas s'est produit, une relecture des profils avec Gene Mapper a été nécessaire pour décider de l'affectation de l'allèle à attribuer.

Les synonymes sont les accessions aux profils génétiques identiques ayant des noms différents. Il peut s'agir d'accessions extrêmement proches génétiquement, que les 25 microsatellites ne peuvent différencier. La synonymie de certaines accessions peut également être due à une mauvaise attribution de nom ou à une erreur de prélèvement précédant le génotypage (on a ainsi pu observer des cas d'homonymie qui n'en étaient pas par confusion entre collections des noms de marques et des dénomination variétales). C'est le cas lorsque des accessions aux profils identiques ou très proches portent des noms de variétés que l'on sait distinctes. L'observation phénotypique des abricotiers sur lesquels ont été effectués les prélèvements peut aider à vérifier que des accessions sont différentes.

#### 4.4.2 Paramètres de diversité génétique

La matrice de données représentant les allèles (par leurs codes et non leurs tailles) servira à calculer certains paramètres d'estimation du polymorphisme des microsatellites, de diversité génétique et de différenciation génétique :

- le nombre moyen d'allèles et de génotypes par locus
- le taux de polymorphisme, ou Polymorphic Information Content (PIC) : il met en évidence la capacité d'un marqueur à détecter un polymorphisme dans une population (Botstein *et al.*, 1980).
- l'hétérozygotie observée ( $H_o$ ) : c'est la moyenne des fréquences des hétérozygotes observées à chacun des locus étudié.
- l'hétérozygotie attendue ( $H_e$ ) : c'est la probabilité qu'à un locus donné, deux allèles choisis aléatoirement soient différents.
- Fis: Déficit en hétérozygote à l'intérieur des populations.
- Fst : Déficit en hétérozygote dû à la subdivision de la population totale en sous-populations ; correspond à la différenciation des populations par rapport au total.
- Fit : Déficit global en hétérozygote sur l'ensemble des populations, c'est-à-dire en considérant une seule et même population.

Les indices de fixation de Wright (F), estimés par la formule de Weir et Cockerham (1984), sont liés par la formule :  $(1-Fit) = (1-Fis)(1-Fst)$ .

- l'indice de diversité de Shannon (I) : il mesure le degré de diversité d'une population.
- le flux de gènes (Nm) : c'est l'échange de plusieurs gènes ou allèles entre différentes populations. Il se calcule par la formule de Nei :  $Nm = 0,25 (1/F_{st}-1)$ .

L'ensemble de ces paramètres de diversité génétiques a été calculé à l'aide des logiciels GENETIX 4.0.5 (Belkhir *et al.*, 1996), GENEPOP 4.2 (Raymond et Rousset, 1995), et FSTAT 2.9.3.2 (Goudet, 2002).

#### **4.4.3 Analyse factorielle des correspondances**

L'analyse factorielle des correspondances (AFC) vise à rassembler en un nombre réduit de dimensions la plus grande partie de l'information initiale en s'attachant aux valeurs relatives, soit les correspondances entre les variables. L'AFC fournit un espace de représentation commun aux variables et aux individus. Les individus analysés sont vus comme un nuage de points dans un hyperespace qui a autant de dimensions qu'il y a de modalités (ici les allèles) sur toutes les variables (ici les allèles aux différents loci). Cette analyse a été effectuée à l'aide du logiciel GENETIX 4.0.5 (Belkhir *et al.*, 1996-2004).

#### **4.4.4 Analyse Bayésienne de regroupement**

L'analyse Bayésienne de regroupement vise à estimer la structure génétique d'un ensemble d'individus en rassemblant ces derniers dans des groupes, sur la base de leurs profils génétiques, elle permet l'identification de groupes génétiques différents. On obtient après analyse des groupes distincts, génétiquement différents. Cette approche est implémentée dans le programme STRUCTURE (Pritchard *et al.*, 2000). Ce programme travaille sous deux hypothèses : les populations sont à l'équilibre de Hardy-Weinberg (panmixie et pas de pression évolutive), et il y a existence d'un équilibre de liaison entre les différents locus.

Dans cette étude réalisée avec le logiciel STRUCTURE, le modèle avec admixture (le génotype d'un individu peut provenir de plusieurs populations) et l'option des fréquences alléliques indépendantes ont été choisis. On a fait varier le nombre K de populations de 1 à 10, en effectuant 10 simulations par valeur de K. Pour déterminer le nombre optimal K de populations, la procédure définie par Evanno *et al.* (2005) a été utilisée. Au final, le logiciel STRUCTURE attribue à chaque individu sa probabilité d'appartenance à chaque population. On utilise ces probabilités pour déterminer la composition de chaque population.

Les analyses par STRUCTURE ont été réalisées sur le serveur Titan de l'université d'Oslo (Kumar *et al.*, 2009). Ce serveur permet de diminuer considérablement la durée des analyses.

## **5. Résultats et discussion**

### ***5.1 Analyse de la diversité observée***

#### **5.1.1 Diversité allélique des accessions patrimoniales et issues de prospections**

Concernant les 777 accessions patrimoniales et issues de prospections, un total de 448 allèles a été détecté pour les 25 marqueurs SSR utilisés, de 12 (BPPCT001, UDP96018) à 28 par locus

(UDP98409) (Tableau 3).

L'hétérozygotie attendue ( $H_e$ ) pour chaque locus va de 0.157 pour UDP-96018 à 0.874 pour UDP98-409, avec une moyenne de 0.682. MA014a et UDP-96018 sont les locus les moins polymorphes, tandis que UDP98-409 est le plus polymorphe. La valeur moyenne de l'hétérozygotie observée est de 0.566.  $H_o$  varie de 0.086 (UDP-96018) à 0.759 (CPPCT006). L'ensemble des marqueurs présente un déficit d'hétérozygotie, excepté AMPA119.

**Tableau 3 : Paramètres de diversité génétiques des locus microsatellites utilisés dans l'étude de la diversité génétique patrimoniale**

Locus	Position du groupe de liaison	nombre d'allèles	nombre d'allèle fréquents ( $f > 0,05$ )	$H_e$	$H_o$	Fis
AMPA100	6	18	5	0,802	0,73	0,091
AMPA101	3	17	5	0,695	0,609	0,126
AMPA105	5	18	5	0,829	0,632	0,237
AMPA109	1	22	3	0,477	0,316	0,339
AMPA116	2	17	7	0,722	0,611	0,158
AMPA119	3	13	3	0,465	0,478	0,19
BPPCT001	2	12	2	0,46	0,437	0,048
BPPCT004	2	23	7	0,845	0,7	0,171
BPPCT008	6	17	4	0,8	0,726	0,092
BPPCT017	5	22	3	0,642	0,589	0,083
BPPCT025	6	12	4	0,659	0,549	0,17
BPPCT030	2	17	5	0,798	0,707	0,116
BPPCT038	5	16	5	0,793	0,719	0,094
BPPCT040	4	13	3	0,592	0,507	0,144
CPPCT006	8	20	6	0,827	0,759	0,084
CPPCT022	7	25	6	0,86	0,695	0,194
CPPCT030	6	24	6	0,76	0,64	0,159
CPPCT033	7	16	4	0,61	0,449	0,264
CPPCT034	1	16	2	0,579	0,486	0,159
MA014a	6	11	2	0,556	0,438	0,21
MA040a	6	24	3	0,725	0,506	0,302
UDP96-018	1	12	1	0,157	0,086	0,452
UDP-97402	4	17	5	0,722	0,401	0,443
UDP-98409	8	28	7	0,874	0,738	0,157
UDP-98412	6	18	5	0,799	0,648	0,189
Moyenne		17,92	4,32	0,682	0,566	0,187



### **5.1.2 Diversité allélique des accessions issues d'amélioration génétique**

Concernant les 460 accessions issues de programmes d'amélioration, un total de 288 allèles a été détecté pour les 25 marqueurs SSR utilisés, de 5 (MA014a) à 20 par locus (CPPCT022, UDP-98409) (Tableau 4). Le nombre moyen d'allèles par locus est de 12.96, mais il chute à 4.17 lorsque l'on ne considère pas les allèles rares (ceux qui ont une fréquence inférieure à 0,05). Le nombre d'allèles par locus avec une fréquence supérieure à 0,05 va de 1 (UDP-96018) à 8 (CPPCT022).

L'hétérozygotie attendue ( $H_e$ ) pour chaque locus va de 0.16 pour UDP-96018 à 0.87 pour CPPCT022, avec une moyenne de 0.68. MA014a et UDP-96018 sont les marqueurs les moins polymorphes, tandis que UDP98-409 est le plus polymorphe. La valeur moyenne de l'hétérozygotie observée est de 0.62.  $H_o$  varie de 0.13 (UDP-96018) à 0.81 (BPPCT028, CPPCT022). La plupart des marqueurs ne présente ni de déficit, ni d'excès significatif d'hétérozygotie. Les marqueurs présentant un déficit d'hétérozygotie sont AMPA105, BPPCT008, CPPCT030, MA040a et UDP-97402. BPPCT040 présente un léger excès d'hétérozygotie.

**Tableau 4 : Paramètres de diversité génétiques des locus microsatellites utilisés dans l'étude de la diversité génétique hybride**

Locus	Position du groupe de liaison	nombre d'allèles	nombre d'allèle fréquents ( $f > 0,05$ )	He	Ho	Fis
AMPA100	6	14	4	0,74	0,70	0,055
AMPA101	3	14	4	0,76	0,71	0,064
AMPA105	5	14	4	0,79	0,64	0,184
AMPA109	1	14	4	0,55	0,51	0,069
AMPA116	2	13	5	0,79	0,71	0,097
AMPA119	3	11	3	0,56	0,54	0,038
BPPCT001	2	8	2	0,36	0,35	0,009
BPPCT004	2	17	5	0,77	0,77	0,010
BPPCT008	6	13	5	0,76	0,67	0,118
BPPCT017	5	17	3	0,70	0,70	-0,002
BPPCT025	6	11	5	0,61	0,61	0,001
BPPCT030	2	9	5	0,75	0,74	0,014
BPPCT038	5	17	5	0,82	0,81	0,014
BPPCT040	4	11	4	0,72	0,77	-0,062
CPPCT006	8	11	6	0,69	0,70	-0,012
CPPCT022	7	20	8	0,87	0,81	0,065
CPPCT030	6	16	5	0,70	0,54	0,23
CPPCT033	7	12	4	0,60	0,51	0,151
CPPCT034	1	11	3	0,57	0,48	0,16
MA014a	6	5	2	0,55	0,57	-0,025
MA040a	6	14	4	0,74	0,42	0,442
UDP96-018	1	6	1	0,16	0,13	0,204
UDP-97402	4	12	6	0,80	0,61	0,228
UDP-98409	8	20	5	0,79	0,79	0,003
UDP-98412	6	115	5	0,78	0,76	0,034
<b>Moyenne</b>		<b>12,96</b>	<b>4,17</b>	<b>0,68</b>	<b>0,62</b>	<b>0,008</b>

### 5.1.3 Comparaison entre diversité patrimoniale et diversité hybride, et comparaison avec une étude antérieure

La comparaison entre les accessions patrimoniales et les accessions hybrides (Tableau 5) montre que les premières présentent un nombre d'allèles moyen total nettement supérieur (17,9) par rapport aux deuxièmes (12,9), ce qui rend compte que les fonds génétiques améliorés sont beaucoup plus réduits que la variabilité génétique disponible au sein de l'espèce.

Nous noterons aussi l'apport significatif des accessions issues de régions complémentaires par rapport à l'étude de Bourguiba et *al.* (2012b) qui contribue à augmenter significativement le nombre d'allèles détectés chez les accessions patrimoniales, la plupart de ces nouveaux allèles étant peu fréquents (fréquence inférieure à 0,05).

Le nombre moyen de 17.9 allèles par locus est ainsi beaucoup plus élevé dans la nouvelle collection patrimoniale que celui observé lors de l'étude de Bourguiba et *al.* (2012b) qui était de 10,3 (Tableau 5). Cependant, un certain nombre de ces allèles présentent une fréquence faible ( $<0.05$ ). Si l'on prend en compte uniquement les allèles les plus fréquents ( $f > 0.05$ ), on trouve un nombre d'allèles moyen par locus de 4.32, contre 3,96 pour Bourguiba et *al.* (2012). Le nombre d'allèles par locus avec une fréquence supérieure à 0,05 va de 1 (UDP96018) à 7 (AMPA 116, BPPCT004, UDP98409).

L'analyse des paramètres de diversité génétique par groupe géographique va permettre d'examiner l'origine des nouveaux allèles détectés.

**Tableau 5 : Comparaison du nombre d'allèles détectés avec l'étude de Bourguiba et al. (2012)**

	Bourguiba <i>et al.</i> (2012)	Accessions patrimoniales et prospections	Accessions hybrides
Nombre d'accessions	207	777	460
Allèles détectés	257	448	288
Nombre minimal d'allèles détectés	5 (BPPCT001)	12 (BPPCT001, UDP96018)	5 (MA014a)
Nombre maximal d'allèles détectés	18 (UDP98409)	28 (UDP98409)	20 (CPPCT022, UDP98409)
Nombre moyen d'allèles détectés	10,28	17,92	12,96
Nombre minimal d'allèles fréquents détectés ( $f > 0,05$ )	1 (UDP96018)	1 (UDP96018)	1 (UDP96018)
Nombre maximal d'allèles fréquents détectés ( $f > 0,05$ )	8 (UDP98409)	7 (AMPA116, UDP98409, BPPCT004)	8 (CPPCT022)
Nombre moyen d'allèles fréquents détectés ( $f > 0,05$ )	3,96	4,32	4,17

### 5.1.4 Diversité allélique patrimoniale intra et inter groupes géographiques

La diversité génétique allélique observée diffère aussi en fonction des groupes géographiques. Le nombre total d'allèles détecté va de 164 (Amérique) à 342 (Asie centrale) (Tableau 6). Cependant, ce nombre d'allèles est fortement dépendant de la taille des échantillons, qui varie de 33 (Amérique) à 238 (Maghreb). La notion de richesse allélique, développée par El Mousadik et Petit (1996), permet de corriger ce biais par l'estimation du nombre d'allèles attendus dans un échantillon de taille donnée. Ici, pour  $n=23$ , la richesse allélique la plus importante est obtenue en Asie de l'Est (10,02), et la plus faible est obtenue au sein du Maghreb (5). L'hétérozygotie attendue ( $H_e$ ) la plus importante est obtenue par l'Asie de l'Est (0,78), et la plus faible est obtenue par l'Europe centrale (0,525). Les valeurs des  $F_{is}$  montrent un déficit d'hétérozygotie pour l'Amérique, l'Europe méditerranéenne, l'Asie centrale et l'Asie de l'Est. La valeur de la différenciation génétique globale ( $F_{st}$  global) est 0,086. Les valeurs de  $F_{st}$  entre paires de groupes géographiques varient de 0,036 (entre Amérique et Asie centrale) à 0,211 (entre Europe centrale et Asie de l'Est) (Tableau 7).

**Tableau 6 : Paramètres de diversité génétique au sein des groupes géographiques**

Groupes géographiques	Nombre d'accessions	$H_o$	$H_e$	$F_{is}$	Nombre total d'allèles	Richesse allélique
<b>Amérique</b>	33	0,580	0,655	0,130	164	6,13
<b>Asie centrale</b>	162	0,581	0,734	0,211	342	8,50
<b>Asie de l'Est</b>	56	0,604	0,780	0,235	310	10,02
<b>Europe centrale</b>	74	0,592	0,525	-0,122	204	5,53
<b>Europe méditerranéenne</b>	214	0,495	0,571	0,135	195	5,00
<b>Maghreb</b>	238	0,593	0,625	0,053	242	5,46

**Tableau 7 : Différenciation génétique au sein de chaque paire de groupes géographiques ( $F_{st}$ )**

	Amérique	Asie centrale	Asie de l'Est	Europe centrale	Europe méditerranéenne
Amérique	0				
Asie centrale	0,036	0			
Asie de l'Est	0,104	0,098	0		
Europe centrale	0,114	0,117	0,211	0	
Europe méditerranéenne	0,076	0,071	0,195	0,132	0
Maghreb	0,039	0,058	0,151	0,107	0,041

L'étude de l'allèle le plus fréquent pour chaque marqueur permet de repérer les allèles particulièrement présents dans chaque groupe géographique, de manière spécifique ou non (Tableau 8).

Chez certains marqueurs, l'allèle le plus fréquent est toujours le même, quel que soit le groupe géographique considéré (AMPA 119, BPPCT 001, BPPCT 017, CPPCT 034 et UDP96-018). Seule la fréquence de cet allèle varie selon le groupe géographique. Le marqueur CPPCT 022 présente la particularité de changer d'allèle le plus fréquent selon le groupe géographique considéré. Certains allèles semblent spécifiques de certains groupes géographiques. Ainsi, l'allèle 40 du marqueur CPPCT 030 est majoritaire au Maghreb et en Europe, mais est peu présent dans les autres régions. L'allèle 100 du marqueur BPPCT 040 et l'allèle 30 du marqueur BPPCT 038 sont particulièrement présents en Europe méditerranéenne, par rapport aux autres groupes géographiques. L'allèle 70 du marqueur MA 040a semble spécifique des accessions d'Europe centrale. L'allèle 40 du marqueur MA 014a est davantage présent en Asie centrale. L'Asie de l'Est est le groupe géographique qui présente le plus d'allèles qui lui sont spécifiques. Le cas le plus marqué est celui du marqueur BPPCT 025 : l'allèle 20 est majoritaire chez les accessions d'Asie de l'Est (68%), alors que chez tous les autres groupes géographiques, c'est l'allèle 50 qui est prépondérant. Cet allèle est donc une spécificité de l'Asie de l'Est.

**Tableau 8 : Allèle le plus fréquent par marqueur pour chaque groupe géographique**

Marqueur	Allèle le plus souvent détecté – Fréquence de l'allèle le plus souvent détecté					
	Maghreb	Europe méd	Europe centr	Asie centrale	Asie de l'Est	Amérique
AMPA100	60 – 0.51	60 – 0.37	10 – 0.32	60 – 0.24	50 – 0.29	60 – 0.35
AMPA101	30 – 0.60	30 – 0.52	30 – 0.51	30 – 0.44	90 – 0.38	30 – 0.47
AMPA105	110 – 0.45	50 – 0.29	30 – 0.38	110 – 0.21	50 – 0.43	80 – 0.27
AMPA109	130 – 0.90	130 – 0.70	130 – 0.85	130 – 0.60	100 – 0.25	130 – 0.60
AMPA116	30 – 0.36	30 – 0.41	30 – 0.69	30 – 0.48	70 – 0.24	30 – 0.38
AMPA119	20 – 0.89	20 – 0.70	20 – 0.57	20 – 0.58	20 – 0.30	20 – 0.82
BPPCT001	10 – 0.69	10 – 0.68	10 – 0.87	10 – 0.53	10 – 0.65	10 – 0.75
BPPCT004	160 – 0.39	130 – 0.30	60 – 0.32	150 – 0.20	60 – 0.22	130 – 0.33
BPPCT008	50 – 0.31	50 – 0.41	50 – 0.41	50 – 0.25	70 – 0.22	50 – 0.70
BPPCT017	90 – 0.52	90 – 0.53	90 – 0.53	90 – 0.57	90 – 0.29	90 – 0.47
BPPCT025	50 – 0.59	50 – 0.56	50 – 0.50	50 – 0.49	20 – 0.68	50 – 0.51
BPPCT030	50 – 0.39	60 – 0.34	10 – 0.30	10 – 0.27	10 – 0.27	40 – 0.38
BPPCT038	90 – 0.39	30 – 0.44	90 – 0.32	90 – 0.27	110 – 0.23	90/130 – 0.25
BPPCT040	50 – 0.72	100 – 0.46	50 – 0.71	50 – 0.58	50/90 – 0.25	50 – 0.49
CPPCT006	110 – 0.37	110 – 0.47	50 – 0.35	100 – 0.20	110 – 0.24	110 – 0.36
CPPCT022	110 – 0.34	250 – 0.34	150 – 0.32	160 – 0.16	60 – 0.37	90 – 0.26
CPPCT030	40 – 0.50	40 – 0.63	40 – 0.65	50 – 0.27	70 – 0.24	70 – 0.31
CPPCT033	60 – 0.80	60 – 0.75	60 – 0.45	60 – 0.33	50 – 0.30	60 – 0.51
CPPCT034	100 – 0.63	100 – 0.66	100 – 0.68	100 – 0.44	100 – 0.65	100 – 0.69
MA014a	30 – 0.67	30 – 0.63	30 – 0.59	40 – 0.41	30 – 0.56	40 – 0.56
MA040a	30 – 0.44	30 – 0.43	70 – 0.55	30 – 0.48	30 – 0.32	30 – 0.37
UDP96-018	10 – 1.00	10 – 0.93	10 – 1.00	10 – 0.87	10 – 0.54	10 – 0.97
UDP-97402	90 – 0.65	90 – 0.56	90 – 0.40	90 – 0.45	70 – 0.30	40 – 0.30
UDP-98409	80 – 0.19	60 – 0.30	170 – 0.29	20 – 0.19	100 – 0.20	60 – 0.37
UDP-98412	50 – 0.46	50 – 0.29	100 – 0.34	50 – 0.41	60 – 0.32	50 – 0.37

Lorsque l'on compare la diversité du groupe géographique Asie de l'Est avec la diversité de chacun des autres groupes (le groupe Amérique mis à part), on constate une réduction de la richesse allélique (Tableau 9). Cette réduction est d'autant plus importante que le groupe géographique comparé est situé plus à l'Ouest. Cela suggère qu'il existe un gradient décroissant de diversité génétique de l'Est vers l'Ouest, résultat qui va dans le même sens que l'étude conduite par H. Bourguiba et al. (2012b) à l'échelle du bassin Méditerranéen (au sens large).

**Tableau 9 : Réduction relative de la diversité entre l'Asie de l'Est et les autres groupes géographiques**

Réduction relative de diversité	Ho	He	Richesse allélique
Asie de l'Est vs Asie centrale	0,040	0,059	0,152
Asie de l'Est vs Europe centrale	0,020	0,327	0,448
Asie de l'Est vs Europe méditerranéenne	0,019	0,200	0,455
Asie de l'Est vs Maghreb	0,180	0,268	0,501

**Pour chaque estimateur X, la réduction relative de la diversité a été calculée par  $1 - (X1/Xn)$ , où X1 est l'estimateur du groupe Asie de l'Est, et Xn l'estimateur de chacun des autres groupes.**

### 5.1.5 Pertinence des marqueurs utilisés et positionnement dans le génome

L'analyse de la diversité allélique mise en évidence par les 25 marqueurs microsatellites utilisé peut amener à se poser la question de leur pertinence.

L'étude du nombre d'allèles détectés pour chaque marqueur et chaque groupe géographique permet d'évaluer ce niveau de pertinence (Tableau 10). 4 marqueurs sur 25 (AMPA 119, BPPCT001, MA014a et UDP96-018, en jaune) détectent en moyenne, pour un groupe donné, moins de 3 allèles présentant une fréquence supérieure à 0,05. Ces marqueurs présentent une diversité allélique peu importante, et sont donc moins efficaces que les autres pour différencier des accessions distinctes. Parmi ces marqueurs, UDP96-018 détecte le moins d'allèles pour l'ensemble des groupes géographiques. Il détecte un seul allèle (codé 10) chez les accessions du Maghreb et d'Europe Centrale, ce qui signifie que toutes ces accessions sont homozygotes 10 pour ce marqueur. Chez les autres groupes géographiques, d'autres allèles sont détectés, mais l'allèle 10 est le plus fréquent (sauf pour les accessions d'Asie de l'Est, où 3 allèles avec une fréquence supérieure à 0,05 sont détectés). Le marqueur UDP96-018 est donc extrêmement peu polymorphe, et ne permet pas de différencier la plupart des accessions. Il peut être considéré comme le marqueur le moins pertinent de l'ensemble des 25 microsatellites, son utilisation n'apportant que très peu d'informations dans cette étude de diversité génétique. Il peut être recommandé de ne plus utiliser ce marqueur dans les études à venir, et de le remplacer si possible par un marqueur plus informatif.

**Tableau 10 : Nombre d'allèles détectés au sein de chaque groupe géographique**

Marqueur	Nombre total d'allèles détectés – Nombre d'allèles détectés de fréquence > 0,05						
	Maghreb	Europe méd	Europe centr	Asie centrale	Asie de l'Est	Amérique	Moyenne
AMPA100	7 - 4	10 - 5	11 - 6	15 - 6	14 - 4	16 - 5	10,17 – 5,17
AMPA101	6 - 4	9 - 5	8 - 5	13 - 4	14 - 5	7 - 4	9,33 – 4,33
AMPA105	10 - 5	11 - 5	8 - 6	13 - 6	17 - 3	8 - 6	10,33 – 5,50
AMPA109	7 - 2	9 - 3	6 - 2	15 - 3	11 - 4	5 - 4	7,83 – 3,50
AMPA116	10 - 4	10 - 5	7 - 4	11 - 5	15 - 5	7 - 5	9,00 – 4,83
AMPA119	5 - 2	7 - 3	6 - 3	11 - 2	11 - 4	3 - 2	6,17 – 2,50
BPPCT001	2 - 2	4 - 2	3 - 2	8 - 2	4 - 3	3 - 2	4,00 – 2,17
BPPCT004	9 - 4	12 - 5	8 - 6	19 - 5	15 - 8	7 - 5	10,5 – 5,67
BPPCT008	10 - 6	11 - 5	8 - 3	13 - 6	11 - 6	6 - 3	8,33 – 4,00
BPPCT017	6 - 3	10 - 3	5 - 2	15 - 4	14 - 7	7 - 4	9,17 – 4,17
BPPCT025	6 - 3	7 - 4	6 - 4	7 - 4	9 - 4	6 - 4	6,67 – 4,00
BPPCT030	6 - 4	7 - 4	6 - 5	13 - 4	10 - 6	5 - 5	7,33 – 5,00
BPPCT038	12 - 4	13 - 4	9 - 6	13 - 6	14 - 7	9 - 5	10,50 – 5,67
BPPCT040	5 - 2	6 - 3	6 - 2	10 - 3	11 - 4	5 - 3	7,00 – 3,00
CPPCT006	9 - 5	11 - 5	8 - 4	16 - 7	12 - 7	8 - 6	10,00 – 6,00
CPPCT022	13 - 5	12 - 5	12 - 4	21 - 6	17 - 4	11 - 5	13,83 – 4,83
CPPCT030	11 - 5	11 - 4	8 - 4	22 - 5	14 - 7	9 - 6	11,83 – 5,67
CPPCT033	8 - 2	10 - 3	8 - 4	12 - 5	14 - 5	8 - 6	9,67 – 5,33
CPPCT034	9 - 2	9 - 3	5 - 3	13 - 3	11 - 2	7 - 4	8,33 – 3,33
MA014a	3 - 2	7 - 2	5 - 2	8 - 4	6 - 3	5 - 2	5,67 – 2,50
MA040a	10 - 3	10 - 3	6 - 3	13 - 5	14 - 4	5 - 4	8,00 – 4,00
UDP96-018	1 - 1	5 - 1	1 - 1	8 - 1	9 - 3	2 - 1	4,00 – 1,33
UDP-97402	8 - 2	8 - 4	11 - 4	14 - 5	10 - 8	9 - 6	10,33 – 5,83
UDP-98409	17 - 6	15 - 5	12 - 5	24 - 7	17 - 6	10 - 7	13,83 – 6,50
UDP-98412	6 - 4	13 - 5	7 - 5	13 - 7	16 - 5	5 - 5	8,50 – 5,33
Moyenne	7,84 – 3,44	9,48 – 3,84	7,20 – 3,80	13,60 – 4,60	12,40 – 4,96	6,56 – 4,36	

Les cellules colorées correspondent aux allèles particulièrement présents dans un groupe géographique, par rapport à l'ensemble des accessions.

Le choix de marqueurs microsatellites couvrant l'ensemble du génome (Figure 3) semble avoir une importance dans la ségrégation des accessions en fonction de leurs différences de profils. En effet, les marqueurs correspondant aux allèles « spécifiques » mis en évidence dans le tableau 7 sont répartis sur les huit groupes de liaison de la carte de référence du génome des espèces du genre *Prunus*. Le groupe de liaison comportant le plus d'allèles spécifiques est le groupe 6, qui est aussi le groupe de liaison qui porte le plus de marqueurs utilisés dans cette étude (7). De plus, six des marqueurs identifiés comme étant les plus polymorphes (BPPCT004, BPPCT038, AMPA116, UDP98-409, UDP97-402, CPPCT030) (Tableau 11) sont situés sur les groupes de liaison G2, G4,

G5, G6, G8. Les marqueurs identifiés comme étant les moins polymorphes (UDP96-018, AMPA119, BPPCT001, MA014a) sont situés sur les groupes de liaison G1, G2, G3, G6.

Ces positionnements ne permettent pas de considérer qu'un groupe de liaison est plus informatif qu'un autre, le fait d'utiliser des marqueurs couvrant l'ensemble du génome permet donc une représentation plus fidèle de la diversité génétique des accessions étudiées.



### 5.1.6 Diversité variétale

Les 25 marqueurs microsatellites utilisés permettent, comme le montrent les résultats précédents, d'identifier la plupart des accessions analysées, si elles présentent des différences suffisamment importantes. Concernant les accessions très proches, on peut se demander si ces marqueurs sont suffisants, notamment si les accessions identifiées comme des synonymes ne présentent pas des différences qu'on ne peut détecter ici. L'ajout d'autres marqueurs permettrait peut-être de mettre en évidence d'éventuelles nouvelles différences. De même, on peut se demander si les accessions présentant des profils génétiques très proches sont réellement différentes. En d'autres termes, à partir de combien de différences alléliques peut-on dire que deux accessions sont distinctes ? Cette question du seuil à considérer est essentielle pour les aspects d'identification variétale. Ceci étant ce questionnement relatif aux accessions proches doit être rapproché et ne peut être objectivement analysé qu'avec les variétés ayant fait l'objet d'une épreuve DHS qui permet de mesurer l'existence d'une distance phénotypique suffisante pour considérer des accessions comme identiques ou différentes.

Il faut également insister sur le fait que les marqueurs génétiques n'ont pas, pour le moment, vocation à se substituer aux caractères phénotypiques fixés dans le cadre des descripteurs OCVV et UPOV pour l'identification variétale. Les critères phénotypiques (basés sur des caractères morphologiques, agronomiques et physiologiques) sont essentiels et les seuls réglementairement reconnus pour déterminer l'identité d'une accession et pour éventuellement la différencier d'une autre.

Environ 90 accessions sont évaluées comme ayant un profil génétique identique ou extrêmement proche d'une ou plusieurs autres accessions (en dehors des doublons, c'est-à-dire des accessions au profil identique portant le même nom).

- Parmi elles, 5 sont des mutants et leurs variétés de référence : Boutard est une variété mutante de Bergeron, qui devient mature avec un mois de décalage. Bergecot est un autre mutant de Bergeron, et ne peut être différencié de Boutard par les critères morphologiques et physiologiques. Priboto est un mutant de Goldrich à maturité tardive. Il n'est pas surprenant que ces mutants aient un profil très proche de leur variété de référence, leurs différences tous critères confondus étant très faibles.
- 6 accessions sont en fait des démarques et leurs variétés de référence, c'est-à-dire qu'elles appartiennent à la même variété sous des noms différents. Blarou et Chrisgold sont ainsi des démarques de Bhart, et Orangered en est un nom commercial. Il est logique là encore de retrouver des profils génétiques identiques chez ces accessions.
- En revanche, environ 80 accessions sont sensées être des variétés distinctes, et ont un profil génétique identique à une autre. Dans ce cas, soit ces accessions sont effectivement distinctes, et les marqueurs microsatellites utilisés ne sont pas suffisants pour détecter leurs différences ; soit elles sont identiques, et des noms différents désignent en fait la même variété. A titre d'exemple, les variétés Pazza et Farely présentent des profils génétiques strictement identiques. Toutefois, dans la plupart des cas, ces variétés sont issues de collections différentes. Il va donc être particulièrement important de les comparer à l'aide des descripteurs OCVV-UPOV afin de statuer sur leur distinction avant d'aller plus avant dans la décision.

Concernant la gestion des collections, l'identification des accessions est une étape essentielle. Elle permet d'avoir des informations pour décider ou non de ne pas conserver certaines accessions, notamment les doublons et les variétés très proches. Après analyse de l'ensemble des observations collectées, il pourra être envisagé de définir des « core collections » par pays et pour l'espèce.

## ***5.2 Structure phylogéographique des accessions patrimoniales***

### **5.2.1 Analyses factorielles des correspondances**

#### ***5.2.1.1 AFC concernant les accessions patrimoniales***

Cette AFC concerne uniquement les variétés patrimoniales et issues de prospections. Les accessions sont catégorisées par pays d'origine (Figure 6a et 6b) et par zone géographique d'origine (Figure 7). On observe que la diversité allélique s'exprime par un continuum qui suit un axe géographique Est – Ouest.

Les accessions en provenance d'Asie centrale et d'Asie de l'Est (Japon, Chine et ex-URSS) sont très différentes des populations Euro-méditerranéennes. Elles expriment clairement une diversité allélique très structurée où les accessions Japonaises sont très différentes des accessions Chinoises, ce qui pourrait suggérer des voies de diversification distinctes. Il pourrait en être de même pour les accessions issues d'Ex-URSS mais leur effectif devra être consolidé pour permettre une interprétation pertinente.

L'extrémité Ouest, très compacte, regroupe les accessions originaires d'Algérie, du Maroc, de Tunisie, d'Espagne, puis d'Italie et de France (Figure 6a et 6b). Toutefois un examen par zone géographique (Figure 7) traduit l'existence de 2 groupes distincts dans la zone Euro-méditerranéenne.

#### ***5.2.1.2 Comparaison avec les accessions issues d'amélioration***

Cette deuxième AFC regroupe toutes les accessions étudiées (Figure 8), ce qui consiste en un ajout des accessions hybrides par rapport à la première AFC. Les accessions sont catégorisées par nature : variétés patrimoniales, prospections et hybrides. La Figure 8 représente un focus sur la zone Euro-méditerranéenne qui concentre toutes les variétés améliorées. Nous pouvons constater que les accessions hybrides épousent l'extrémité « Ouest » du graphe, et restent profondément ancrées sur la variabilité disponible au sein de la zone Euro-méditerranéenne.

Quelques hybrides sont en marge : Cegledi Bibor (Hongrie) ; Lejara (Rép. Tchèque) ; Mamaia (Roumanie) ; Lednice 1 (Rép. Tchèque) ; Čačansko zlato (Serbie). Ces hybrides d'Europe de l'Est sont obtenus par incorporation de matériels génétiques issues d'Asie centrale.

De manière globale, on observe que la diversité génétique des hybrides (en jaune) est loin de recouvrir la diversité des accessions patrimoniales et issues de prospections (en rose et gris). Les fonds améliorés explorent une partie très réduite de la diversité de l'espèce.

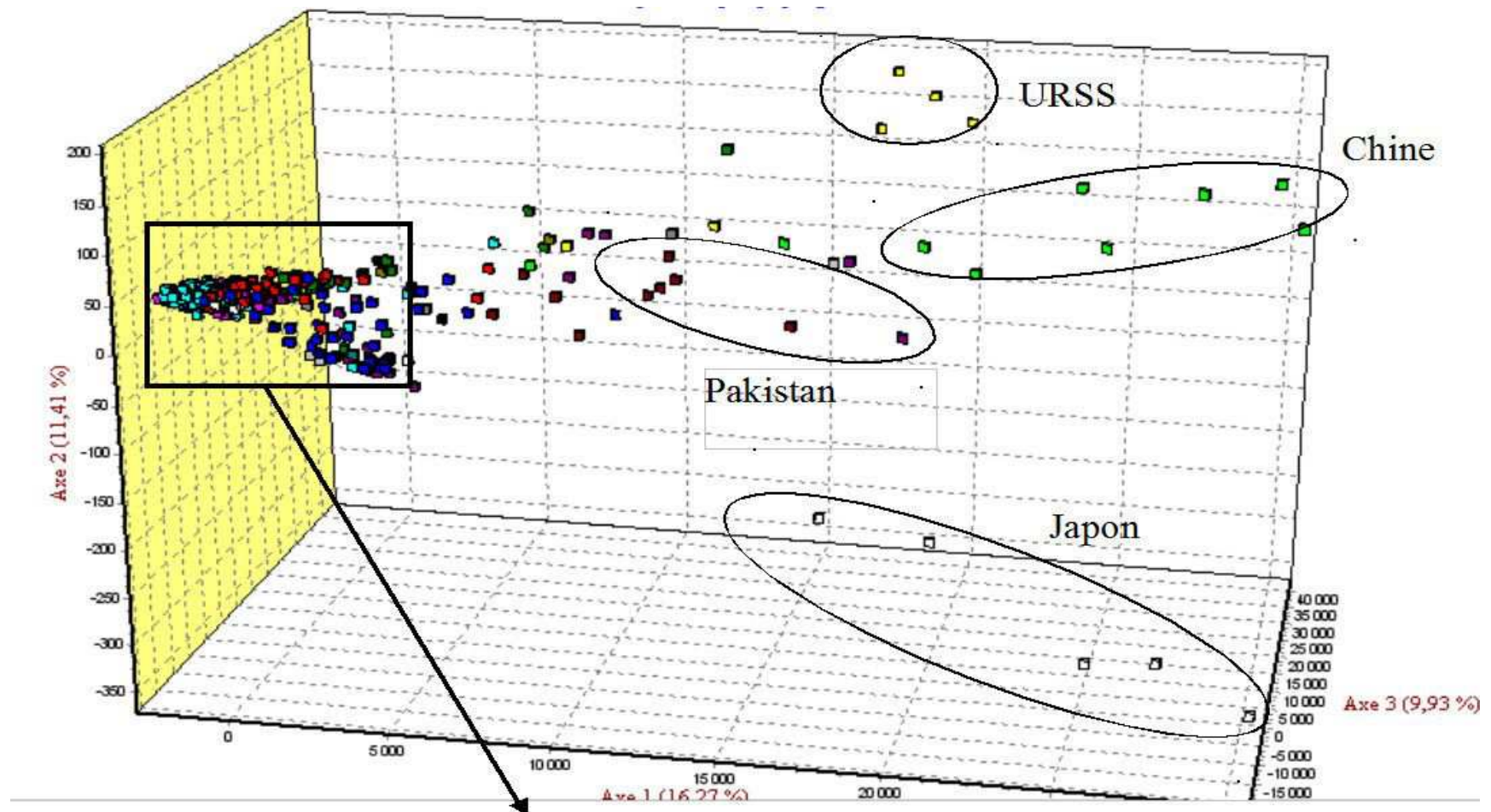


Figure 6a : AFC représentant les accessions patrimoniales et issues de prospections, et leurs pays d'origine (agrandissement de la zone encadrée en figure 6b)

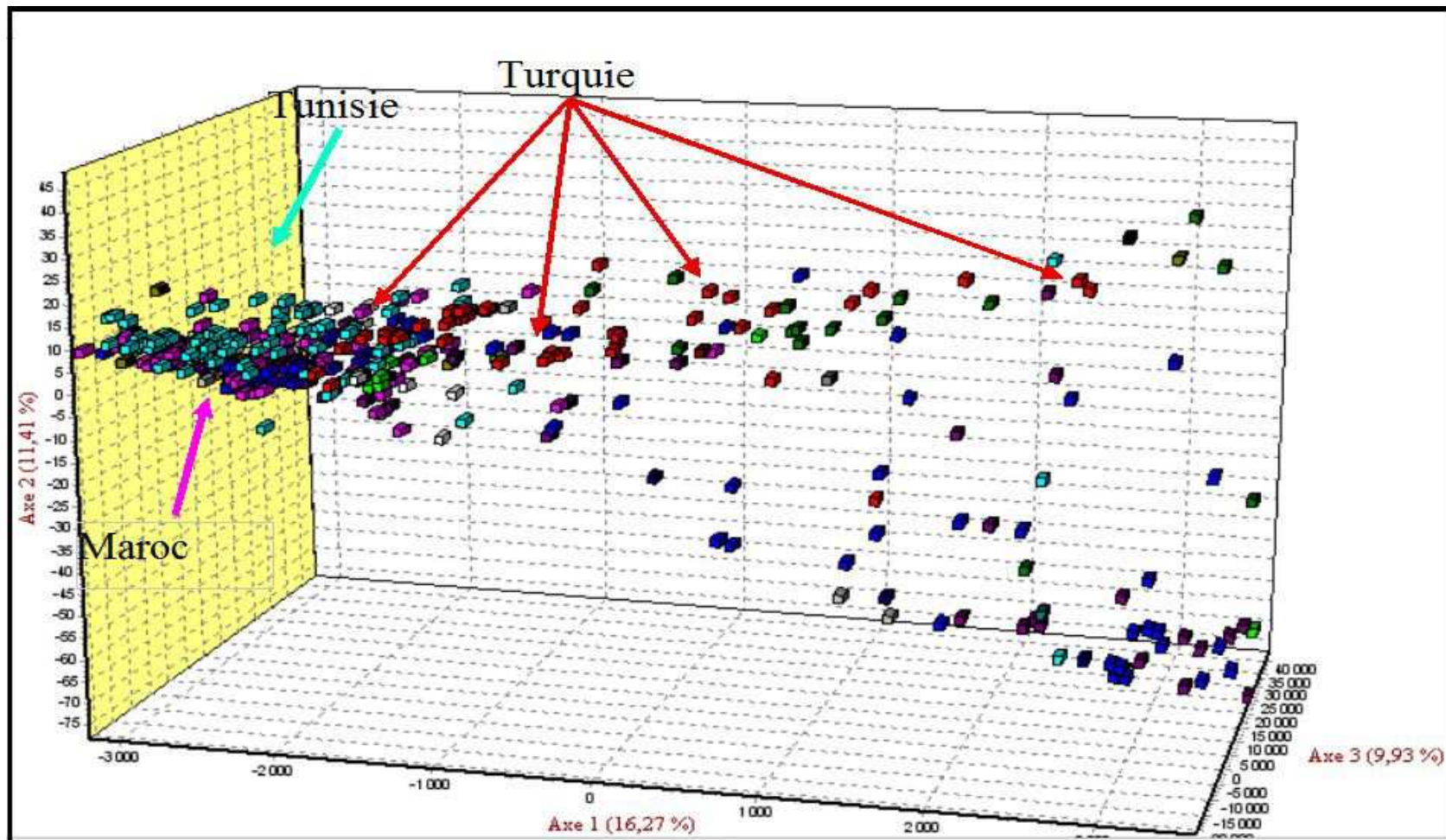


Figure 6b : Détail de l'AFC précédente, correspondant à la zone méditerranéenne



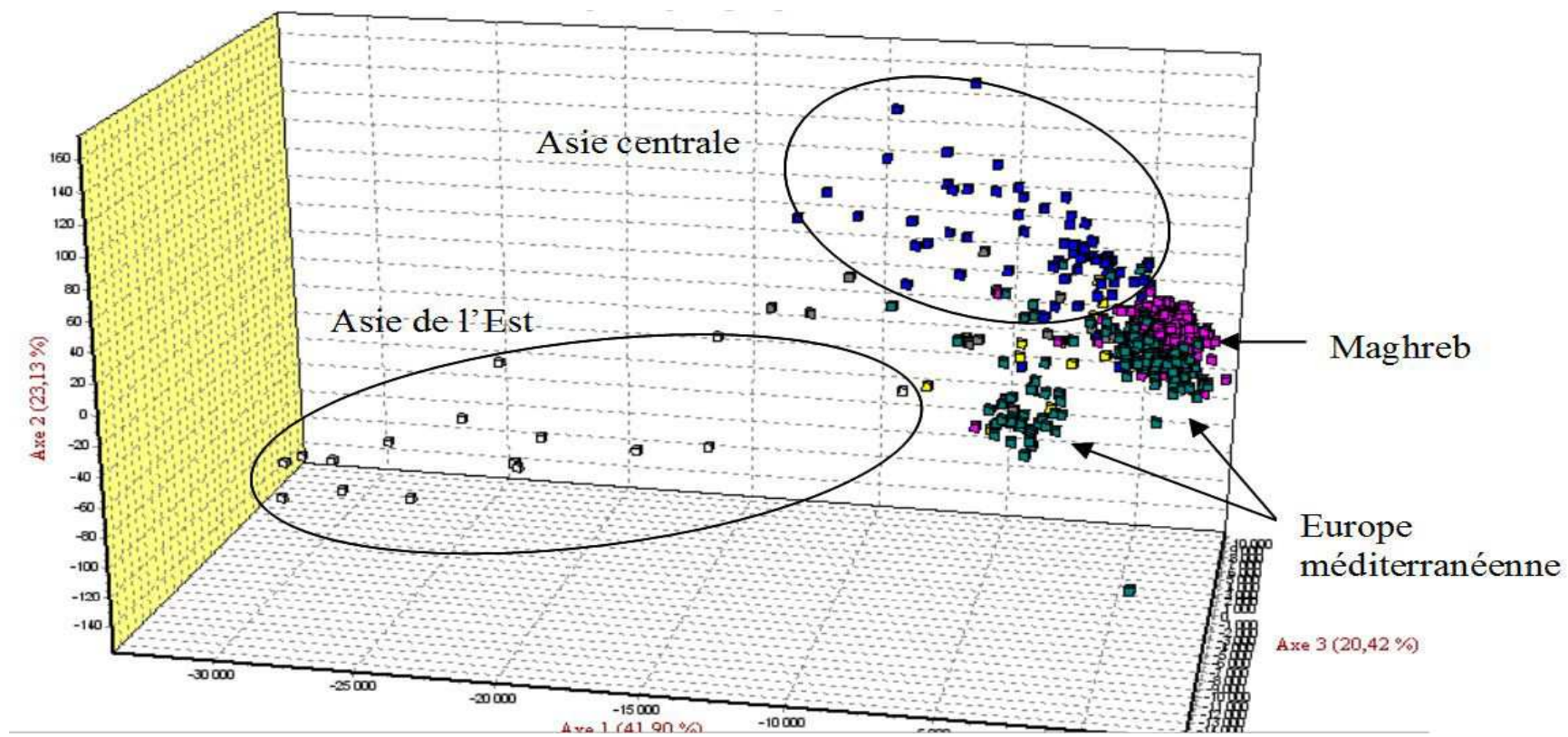


Figure 7 : AFC représentant les accessions patrimoniales et issues de prospections, et leurs zones géographiques d'origine

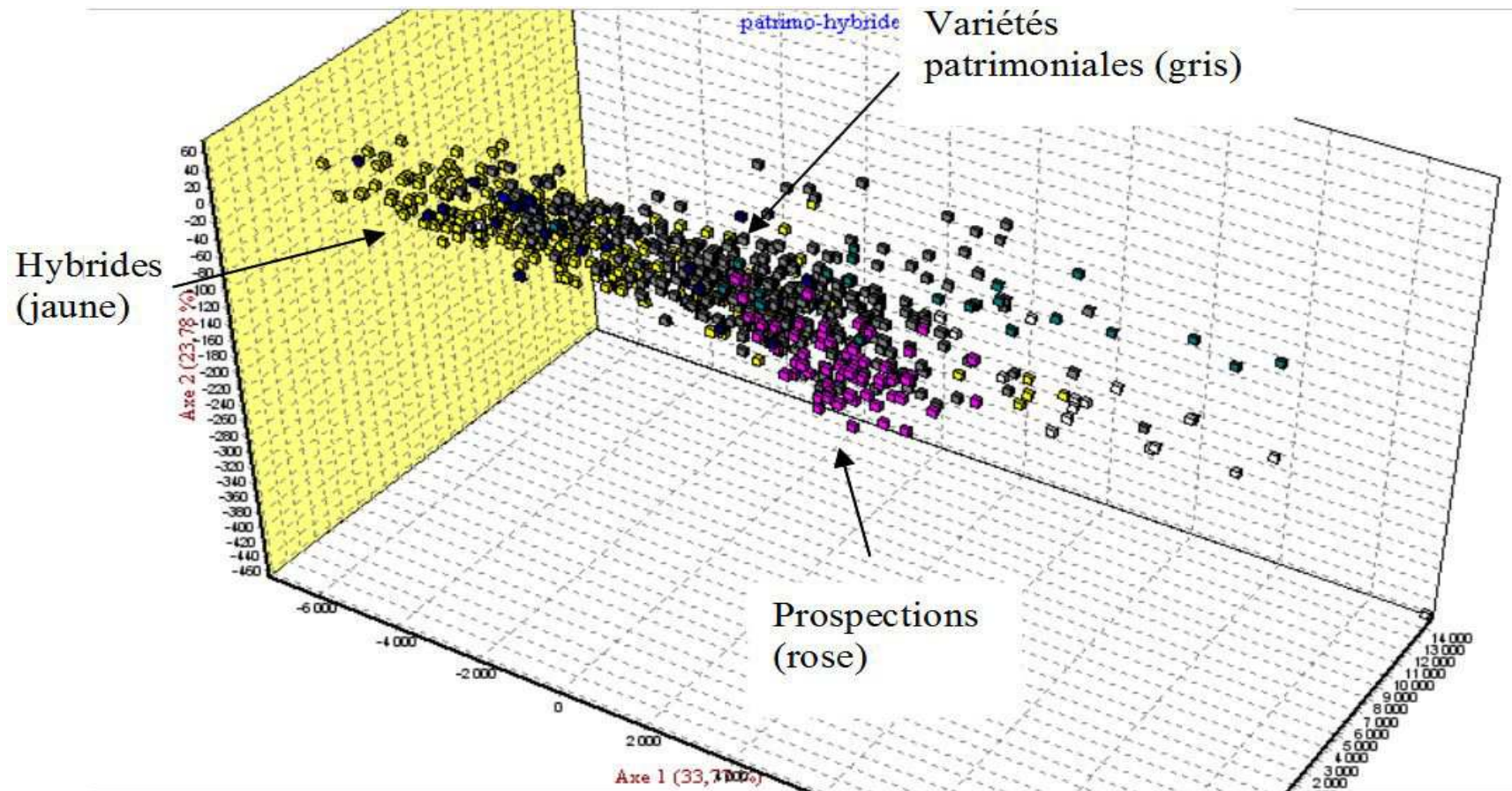


Figure 8 : AFC représentant les accessions patrimoniales et issues de prospection, ainsi que les accessions hybrides

## 5.2.2 Analyse Bayésienne de regroupement

L'analyse a été conduite sur la base de 673 accessions patrimoniales génétiquement distinctes. La procédure d'Evanno (2005), qui permet d'évaluer le nombre probable  $K$  de clusters en se basant sur le  $\Delta K$ , montre que la structure génétique la plus représentative est obtenue pour  $K=2$  ( $\Delta K=53685$ ) et  $K=4$  ( $\Delta K=13$ ). Ces résultats sont parfaitement cohérents avec ceux obtenus par Bourguiba *et al.* (2012b) lors de son étude conduite sur un ensemble d'accessions Euro-méditerranéennes. Pour  $K=2$ , les accessions d'Asie de l'Est et une partie des accessions d'Amérique et d'Asie centrale sont séparées des autres (autrement dit un pool constitué d'une variabilité génétique large, opposé à des sous-ensembles de variabilité réduite). Pour  $K=3$ , la plupart des accessions d'Europe centrale sont séparées des accessions méditerranéennes (Europe et Maghreb). A  $K=4$ , la moitié des accessions d'Asie de l'Est sont placées dans un nouveau groupe, traduisant l'existence de pools différents se différenciant à partir de la zone d'Asie centrale. Pour  $K=5$ , le nouveau groupe défini ne contient pas suffisamment d'accessions pour que l'on considère que la structure génétique de l'abricotier soit mieux définie. Pour  $K=6$ , tous les groupes comptent un nombre significatif d'accessions. Les groupes présents peuvent être définis comme : un groupe « Japon », un groupe « Bassin Méditerranéen Sud », un groupe « Bassin Méditerranéen Nord », un groupe « Europe Centrale », un groupe « Chine et Asie centrale », et un groupe « Asie Centrale ».

Le nombre de groupes  $K=4$  a été choisi comme définissant le mieux la structure génétique de l'abricotier, sur la base du paramètre  $\Delta K$  d'Evanno. Avec ce modèle, les groupes obtenus sont les suivants (Tableau 11 et Figure 9) :

- le cluster 1 « rouge » (114 accessions), composé principalement de la moitié des accessions d'Asie centrale et la moitié des accessions d'Asie de l'Est (partie chinoise) ;
- le cluster 2 « rose » (54 accessions), qui rassemble principalement l'autre moitié des accessions d'Asie de l'Est (correspondant notamment à la partie japonaise) ;
- le cluster 3 « bleu » (404 accessions), composé principalement d'accessions du bassin méditerranéen et d'Asie centrale ;
- le cluster 4 « jaune » (101 accessions), qui regroupe la plupart des accessions d'Europe centrale, et certaines accessions d'Europe méditerranéenne.

L'étude de Bourguiba *et al.* (2012b), portait sur 207 accessions originaires du Maghreb, de France, d'Espagne, d'Italie, d'Europe centrale, de Turquie et d'Iran. Le modèle  $K=4$  révélait dans cette étude les clusters suivants :

- un cluster « bassin méditerranéen sud », formé principalement des accessions du Maghreb et du sud de l'Espagne, correspondant à notre cluster « bassin méditerranéen sud », obtenu pour  $K=6$  ;
- un cluster « bassin méditerranéen nord », regroupant essentiellement des accessions de France et d'Italie, coïncidant avec notre cluster « bassin méditerranéen nord » obtenu pour  $K=6$  ;

Ces deux clusters sont inclus dans notre cluster 3 « bleu » obtenu pour  $K=4$ .

- un cluster « diversité adaptative », regroupant des accessions d'Europe continentale, du sud de la France et du sud de l'Italie, coïncidant avec notre cluster 4 « jaune », obtenu pour K=4 ;
- un cluster « irano-caucasien », formé d'accessions originaires d'Iran et de Turquie, qui se répartit entre nos clusters 3 « bleu » et 1 « rouge ».

L'élargissement de la zone d'étude par l'apport d'accessions d'origines complémentaire montre une concordance avec les résultats obtenus par Bourguiba *et al.* (2012b). Le modèle K=4 est estimé dans les deux études comme étant le plus probable. Un nouveau cluster apparaît dans notre étude, composé principalement d'accessions d'Asie de l'Est (cluster 2 « rose »).

Les accessions américaines ont la particularité d'être présentes dans les quatre groupes définis : la moitié se trouve dans le cluster 3 « méditerranéen » (qui regroupe l'ensemble des accessions maghrébines, la majorité des accessions d'Europe méditerranéenne, et la moitié des accessions d'Europe centrale). Une part importante de ces accessions américaines se trouvent dans le cluster 1, avec l'autre moitié des accessions d'Asie centrale, et les accessions chinoises d'Asie de l'Est. Enfin, dans une proportion moindre, on retrouve les accessions américaines restantes dans les clusters 2 (cluster « japonais ») et le cluster 4 (cluster « central européen »). Cette particularité des accessions américaines est sûrement due au fait que ces accessions ne sont pas réellement patrimoniales, mais dérivent de prospections d'origines diverses réalisées notamment en Asie centrale et présentant une diversité importante.

**Tableau 11 : Moyennes des probabilités d'affectation des accessions, triées par zone géographique d'origine, entre les quatre groupes phylogéographiques définis par le programme STRUCTURE.**

Zone géographique d'origine	Cluster assigné			
	1	2	3	4
Amérique	0.261	0.115	0.470	0.154
Asie centrale	0.498	0.020	0.418	0.064
Asie de l'Est	0.493	0.468	0.017	0.023
Europe centrale	0.085	0.020	0.135	0.760
Maghreb	0.008	0.013	0.918	0.061
Europe méditerranéenne	0.033	0.163	0.640	0.164



$K = 4$   
 $\Delta K = 13$

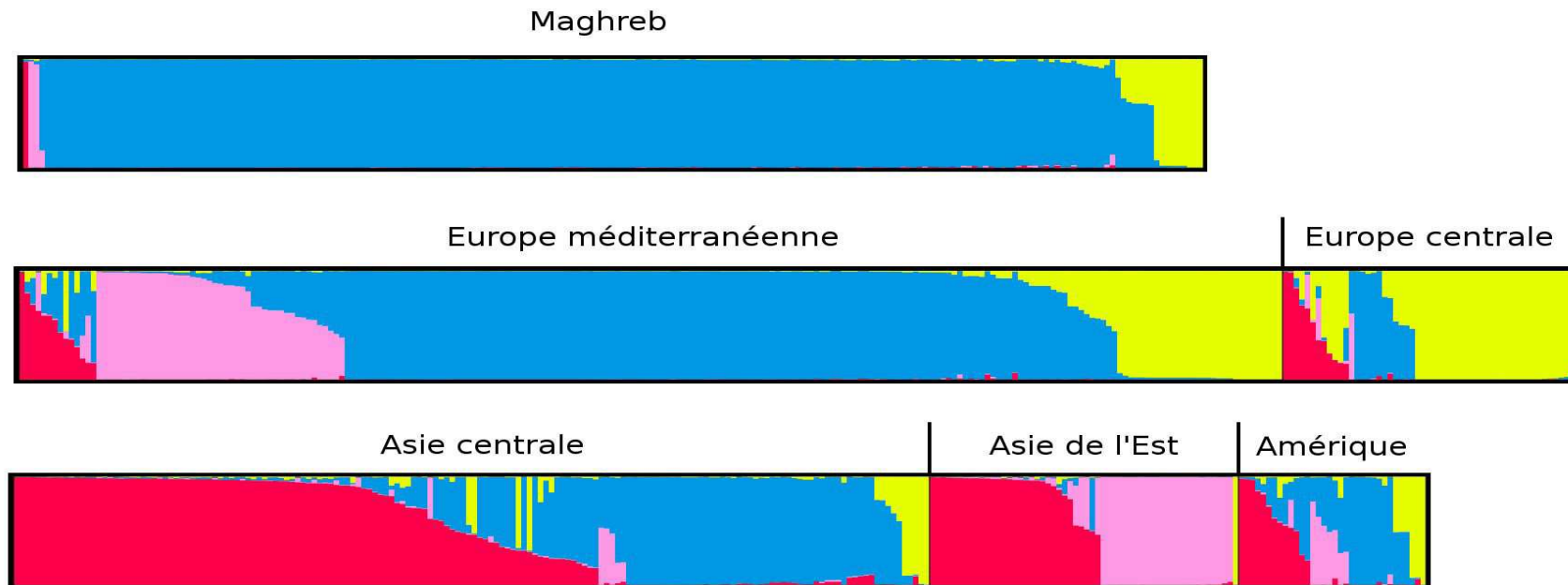


Figure 9 : Estimation de la structure de la population. Chaque individu est représenté par une ligne verticale segmentée en 4 parties colorées. Une partie correspond à un cluster d'appartenance déterminé par le programme STRUCTURE. Cette illustration a été obtenue par le logiciel DISTRUCT (Rosenberg, 2004), à partir des résultats du programme STRUCTURE.

On considère qu'une accession est assignée à un cluster quand sa probabilité d'appartenance à ce cluster est supérieure à 0.70. Lorsque cette probabilité d'affectation est plus faible, les accessions sont considérées en « admixture » : le génotype provient alors de recombinaisons entre plusieurs clusters. Au total, 601 accessions sur 777 (parmi les accessions patrimoniales et issues de prospections) sont clairement assignées, affectées à des clusters (probabilité d'assignation supérieure à 0.70). Les 176 accessions restantes sont en « admixture ». On trouve 52 % d'accessions en admixture pour l'Europe de l'Est (17/33), 44 % pour l'Amérique (15/34), 30 % pour l'Asie centrale (50/166), 24 % pour l'Europe méditerranéenne (56/229), 14 % pour le Maghreb (29/214), et 11 % pour l'Asie de l'Est (6/56).

Si l'on représente les 4 clusters existant pour chaque pays d'origine, on peut observer plus finement la structure génétique définie (Figures 10a et 10b). Les accessions en « admixture » ne sont ici pas représentées. Le cluster 3 (bleu) a clairement une origine méditerranéenne. Il est prépondérant au Maghreb et en Turquie, et très présent en Europe méditerranéenne ; le cluster 4 (jaune) est présent en ponctuellement dans tous les pays à l'Ouest, mais est sur-représenté en Europe continentale ; le cluster 1 (rouge) a une origine asiatique : il est largement majoritaire en Chine, Russie et Pakistan, et se retrouve jusqu'en Turquie. Il est en revanche pratiquement absent en Europe et au Maghreb ; le cluster 2 (rose) a une répartition particulière. Il regroupe la plupart des accessions japonaises et se retrouve également en France, Italie et Europe de l'Est, mais semble absent en Asie centrale. On le retrouve néanmoins dans cette zone géographique chez des accessions en admixture.

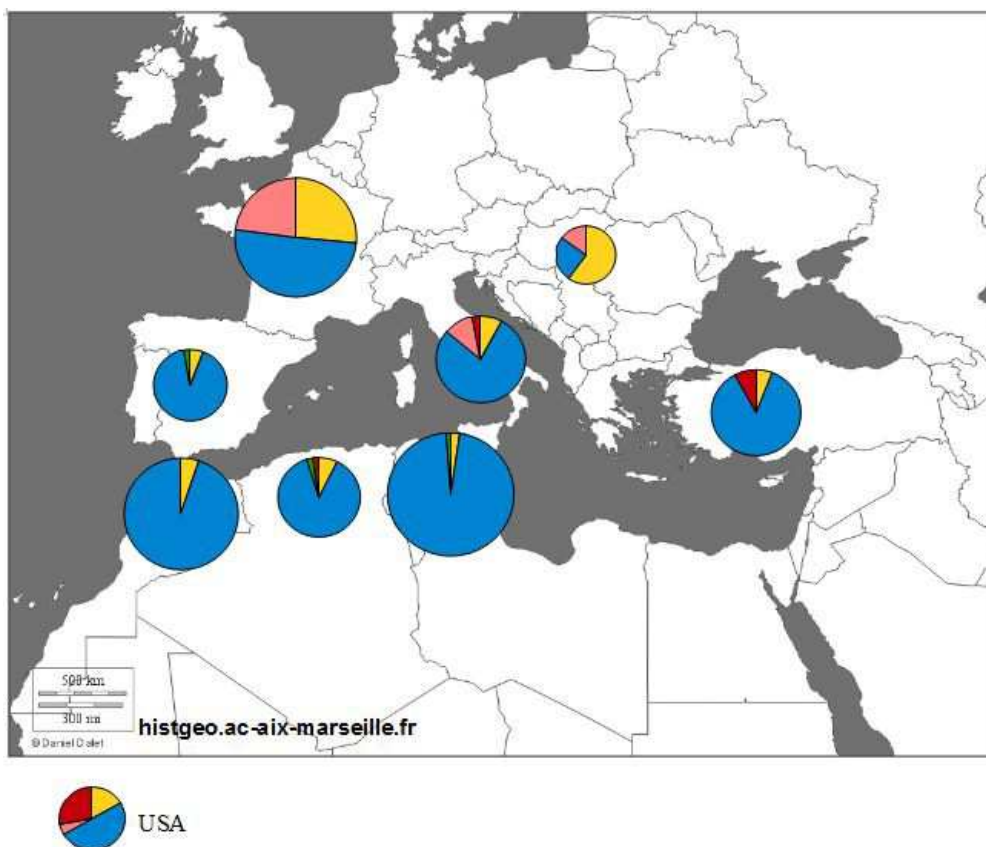


Figure 10a : Répartition des accessions européennes et maghrébines entre les quatre clusters (ou populations)

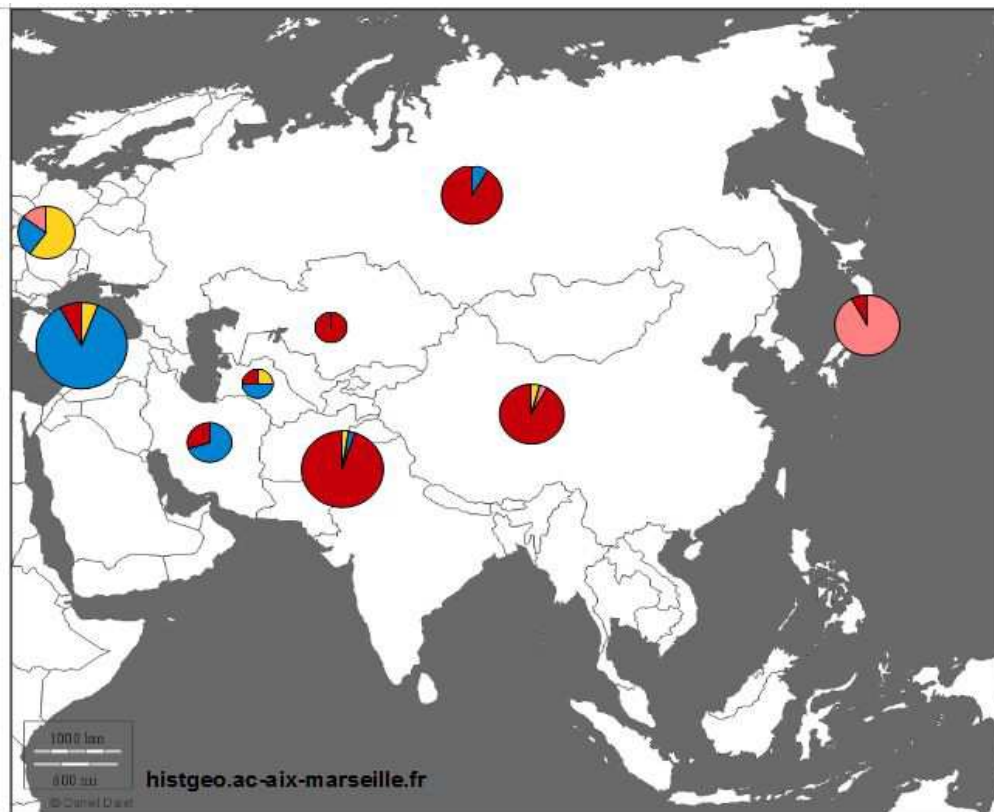


Figure 10b : Répartition des accessions asiatiques entre les quatre clusters (ou populations)

Cette structure phylogéographique semble soutenir l'hypothèse d'une origine de l'espèce en Asie Centrale où nous observons un maximum de diversité allélique. Puis, l'espèce aurait diffusé respectivement vers l'Est et vers l'Ouest, et de ce fait aurait suivi deux voies de différenciation et de spécialisation. L'arrivée de l'espèce dans le bassin méditerranéen se serait accompagnée d'une diminution de la diversité en raison de sa domestication avec 2 verrous de domestication l'un très net en zone méditerranéenne Sud caractérisé par une réduction drastique de diversité, l'autre important mais moins fort sur le continent Européen où vont co-exister 3 groupes l'un commun avec l'Afrique du nord, un second spécifique de l'Europe du Nord auquel vont pouvoir se raccrocher toutes les variétés traditionnellement adaptés en Europe centrale et un troisième groupe faiblement représenté sinon en admixture en Turquie mais retrouvé en Asie de l'Est.

Ces résultats sont partiellement compatibles avec les hypothèses émises par Bailey et Hough (1975) ou Faust et *al.* (1998) qui privilégiaient l'hypothèse Chinoise.

L'examen de cette structure phylogéographique semble particulièrement cohérent avec des éléments phénotypiques disponibles et notamment :

- le cluster 1 (Bleu), Méditerranéen, est caractérisé par des cultivars principalement à faible besoins en froid, bien adaptés en zone chaude mais non adaptés en zone froide,
- le cluster 4 (Jaune), Européen continental, est caractérisé par des variétés à forts besoins en froid, bien adaptées en zones froide mais inadaptées en région chaude,
- le cluster 2 (Rose), Asie de l'Est, est caractérisé par des variétés au spectre d'adaptation large, particulièrement sensibles aux pucerons et de piètre qualité gustatives,

Ils montrent notamment tout l'intérêt de l'exploitation d'une telle diversité dans le cadre de travaux de génétique d'association.

Au plan génétique cette structuration phylogéographique est liée à la présence d'allèles spécifiques à un cluster : ils sont présents chez une partie des accessions d'un cluster donné, mais absents ou très rares chez les accessions des autres clusters. Le tableau 12 rassemble les allèles à la fois présents chez plus de 10 % des accessions d'un cluster donné, et rares chez tous les autres clusters (présents à moins de 5 %).

**Tableau 12 : Allèles spécifiques pour chaque cluster**

marqueur	allèle	Présence de l'allèle au sein de chaque cluster (en %)			
		Cluster 1	Cluster 2	Cluster 3	Cluster 4
AMPA 100	50	<b>16,5</b>	0,9	0	0
	85	0	<b>15,7</b>	0	0
AMPA 101	50	4,5	0,9	2,4	<b>25</b>
	90	4,9	<b>55,6</b>	0,3	0
AMPA 105	20	0	<b>21,7</b>	0	0
AMPA 109	100	<b>45,5</b>	0	1	0
	250	0,5	<b>25</b>	0,1	0
AMPA 116	20	<b>18,3</b>	3,7	4,1	0
	60	1,4	2,8	<b>10,8</b>	2,5
	70	0,5	<b>47,2</b>	0,3	0
AMPA 119	50	0,5	<b>46,3</b>	1,1	0
BPPCT004	190	2,4	<b>30,8</b>	2,6	0
	200	1	0	<b>15,2</b>	0
BPPCT008	40	0	<b>32,4</b>	0,5	0
	90	4,1	<b>20,4</b>	0	0
BPPCT017	160	<b>11,3</b>	0	0	0
BPPCT038	50	<b>12,5</b>	2	0,5	0
	110	3,8	<b>35,3</b>	1,9	0
CPPCT022	90	4,9	2,9	<b>14,2</b>	0
CPPCT030	60	3,5	<b>29,3</b>	0,1	0
	90	0	0	<b>10,7</b>	0
	150	2	0	3,9	<b>32,5</b>
CPPCT033	50	<b>31,5</b>	1,4	2,3	0
CPPCT034	20	2,7	<b>20,4</b>	1	0
	90	4,9	0	<b>64,9</b>	0
MA040a	20	<b>12,8</b>	0	0,3	2,5
	50	<b>15,7</b>	1,9	2,1	0
UDP96-018	28	0	<b>10,2</b>	0	0
	45	1	<b>11,1</b>	1,1	0
UDP97-402	30	<b>25</b>	0,9	0,4	4,9
	70	2,9	<b>60,2</b>	0,5	0
UDP98-409	20	<b>17,9</b>	0	3	0
	25	3,1	<b>12</b>	0	0
	180	0,5	0	3	<b>34,7</b>
UDP98-412	100	4,9	0	3,4	<b>47,6</b>

## 6. Conclusion

L'analyse des profils génétiques de 1253 accessions d'abricotier fait apparaître la faculté des marqueurs microsatellites à caractériser des variétés distinctes, en complément des critères morphologiques, physiologiques et agronomiques plus anciennement utilisés.

L'ensemble des résultats obtenus par l'analyse des paramètres de diversité génétique et par l'analyse Bayésienne tendent à montrer l'existence d'une véritable structuration phylogéographique de l'abricotier à l'échelle mondiale. Cette structuration est cohérente avec celle précédemment établie à l'échelle du Bassin Méditerranéen (Hagen *et al.* 2002, Khadari *et al.*, 2006, Bourguiba *et al.*, 2010, 2012a, 2012b). Elle met en évidence deux voies de différenciation de l'espèce à partir de sa zone géographique d'origine, l'Asie centrale : une voie de différenciation vers l'Est, en direction de la Chine de l'Est et du Japon ; une deuxième voie de différenciation vers l'Ouest, à destination de l'Europe occidentale, en passant par l'Europe centrale et le Maghreb.

Une diminution de la diversité génétique de l'espèce entre l'Asie et le Bassin Méditerranéen a également été mise en évidence, caractéristique de l'effet « bottleneck » accompagnant la domestication de l'espèce.

Une étude davantage focalisée sur la partie asiatique des ressources génétiques permettrait de comprendre plus précisément les étapes de différenciation de l'abricotier du côté Est. Il faudrait pour cela de nouvelles accessions dont l'origine géographique précise est connue. L'incorporation d'accessions de la sous-espèce *Prunus mume* (abricotier du Japon), originaire de Chine et importée au Japon au VIII<sup>ème</sup> siècle, pourrait apporter un nouvel angle d'étude.

Parallèlement, nous devons constater que les fonds améliorés sont loin de recouvrir l'ensemble de la diversité observée. L'existence d'une diversité génétique importante en Asie mériterait être prise en compte dans les programmes d'innovation variétale, toujours à la recherche de caractères d'intérêt à intégrer dans les nouvelles variétés.

## 7. Bibliographie

### 7.1 Liste bibliographique

Aranzana, M. J., Garcia-Mas, J., Carbó, J., Arùs, P. (2002). "Development and variability analysis of microsatellite markers in peach." *Plant Breeding*, 121(1): 87-92.

Audergon, J. M. (1995). Variety and breeding. X International Symposium on Apricot Culture, Izmir, Turquie, *Acta Horticulturae*, 284 : 35-45.

Badenes, M., Asins, MJ, Carbonell, EA, Glacer, G (1996). "Genetic diversity in apricot, *Prunus armeniaca* L. aimed at improvement of resistance to plum pox virus." *Plant Breeding*, 115: 133-139.

Bailey, C. H., Hough, L.F. (1975). Apricots. *Advances in Fruit Breeding*. M. J. Janick J. Indiana, USA, Purdue University Press, 367-383.

Belkhir, K., Borsa, P., Chikhi, L., Raufaste, N., & Bonhomme, F. (1996). 2004. GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations.

Botstein, D., White, R.L., Skolnick, M., Davis, R.W. (1980). "Construction of a genetic linkage map in man using restriction fragment length polymorphisms." *The American Journal of Human Genetics*, 32 (3) : 314-331.

Bourguiba, H., Krichen, L., Audergon, J.M., Khadari, B., Trifi-Farah, N. (2010). "Impact of Mapped SSR Markers on the Genetic Diversity of Apricot (*Prunus armeniaca* L.) in Tunisia." *Plant molecular biology reporter*, 28(4): 578-587.

Bourguiba, H. (2012). Analyse de la structure génétique de l'abricotier au niveau du Bassin Méditerranéen : histoire et diffusion de l'espèce. Tunis, Université El Manar. Doctorat en Sciences biologiques: 83 p.

Bourguiba, H., Audergon, J.M., Krichen, L., Trifi-Farah, L., Mamouni, A., Trabelsi, S., Khadari, B. (2012a). "Genetic diversity and differentiation of grafted and seed propagated apricot (*Prunus armeniaca* L.) in the Maghreb region." *Scientia Horticulturae* 142: 7-13.

Bourguiba, H., Audergon, J.M., Krichen, L., Trifi-Farah, L., Mamouni, A., Trabelsi, S., D'Onofrio, C., Asma, B.M., Santoni, S., Khadari, B. (2012b). "Loss of genetic diversity as a signature of apricot domestication and diffusion into the Mediterranean Basin." *BMC Plant Biology*, 12 : 49.

Brown, A.H.D. (1989). "Core collections: a practical approach to genetic resources management." *Génome*, 31 (2) : 818-824.

Cantini, C., Iezzoni, A.F., Lamboy, W.F., Boritzki, M., Struss, D. (2001). " DNA fingerprinting of tetraploid cherry germplasm using simple sequence repeats " *Journal of the American Society for Horticultural Science*, 126(2): 205-209.

Cipriani, G., Lot, G., Huang, W-G., Marrazzo, M.T., Peterlunger, E., Testolin, R. (1999). "AC/GT and AG/CT microsatellite repeats in peach (*Prunus persica* (L.) Batsch): isolation, characterization and cross-species amplification in *Prunus*." *Theoretical and Applied Genetics*, 99: 65-72.

Clarke, J. B., Tobutt, K.R. (2003). "Development and characterization of polymorphic microsatellites from *Prunus avium* 'Napoleon'." *Molecular Ecology Notes*, 3(4): 578-580.

Decroocq, V., Fave, M.G., Hagen, L., Bordenave, L. Decroocq, S. (2003). " Development and transferability of apricot and grape EST microsatellite markers across taxa " *Theoretical and Applied Genetics*, 106(5): 912-922.

Dirlewanger, E., Cosson, P., Tavaud, M., Aranzana, M., Poizat, C., Zanetto, A., Arús, P., Laigret, F. (2002). "Development of microsatellite markers in peach [*Prunus persica* (L.) Batsch] and their use in genetic diversity analysis in peach and sweet cherry (*Prunus avium* L.)." *Theoretical and Applied Genetics*, 105(1): 127-136.

Dondini, L., Lain, O., Geuna, F., Banfi, R., Gaiotti, F., Tartarini, S., Bassi, D., Testolin, R. (2007). "Development of a new SSR-based linkage map in apricot and analysis of synteny with existing *Prunus* maps." *Tree Genetics and Genomes*, 3(3): 239-249.

Downey, S. L., Iezzoni, A.F. (2000). "Polymorphic DNA markers in black cherry (*Prunus serotina*) are identified using sequences from sweet cherry, peach, and sour cherry." *Journal of the American Society for Horticultural Science*, 125(1): 76-80.



- El Mousadik, A., Petit, R.J. (1996). "High level of genetic differentiation for allelic richness among populations of the argan tree [*Argania spinosa* (L.) Skeels] endemic to Morocco." *Theoretical and Applied Genetics*, 92 (7) : 832-839.
- Evanno, G., Regnaut, S., Goudet, J. (2005). "Detecting the number of clusters of individuals using the software structure: a simulation study." *Molecular Ecology*, 14 (8) : 2611-2620.
- Faust, M., Suranyl, D., Nyujto, F. (1998). "Origin and dissemination of apricot." *Horticultural Reviews*, 22: 225-265.
- Flint-Garcia, S.A., Thornsberry, J.M., Buckler, E.S.4 (2003). "Structure of linkage disequilibrium in plants." *Annual Review of Plant Biology*, 54 : 357-374.
- Goudet, J. (2002). Fstat. A program to estimate and test gene diversities and fixation indices. Release, 2(1).
- Hagen, L., Khadari, B, Lambert, P, Audergon, JM (2002). "Genetic diversity in apricot revealed by AFLP markers: species and cultivars comparisons." *Theoretical and Applied Genetics*, 105: 298-305.
- Hagen, L. S., Chaib, J., Fady, B., Decroocq, V., Bouchet, J.P., Lambert, P., Audergon, J.M. (2004). "Genomic and cDNA microsatellites from apricot (*Prunus armeniaca* L.) " *Molecular Ecology Notes*, 4(4): 742-745.
- He, T. M., Chen, X.S., Xu, Z., Gao, J.S., Lin, P.J., Liu, W., Liang, Q., Wu, Y. (2007). "Using SSR markers to determine the population genetic structure of wild apricot (*Prunus armeniaca* L.) in the Ily Valley of West China." *Genetic Resources and crop evolution*, 54(3): 563-572.
- Hormaza, J. I. (2002). "Molecular characterization and similarity relationships among apricot (*Prunus armeniaca* L.) genotypes using simple sequence repeats." *Theoretical and Applied Genetics*, 104(2-3): 321-328.
- Jay F, François O, Blum MGB (2011) Predictions of Native American Population Structure Using Linguistic Covariates in a Hidden Regression Framework. *PLoS ONE* 6(1): e16227.
- Joobeur, T., Periam, N., de Vicente, M.C., King, G.J., Arús, P. (2000). "Development of a second generation linkage map for almond using RAPD and SSR markers." *Genome*, 43(4): 649-655.

Khadari, B.; Krichen, L.; Lambert, P.; Marrakchi, M.; Audergon, J.M. (2006). "Genetic structure in Tunisian apricot, *Prunus armeniaca* L., populations propagated by grafting: a signature of bottleneck effects and ancient propagation by seedlings". *Genetic Resources and Crop Evolution*, 53 (4): 811-819.

Kostina, K. (1969). "The use of varietal resources of apricots for breeding." *Trud Nikit Bot Sada*, 40: 45-53.

Krichen, L., Audergon, J.M., Trifi-Farah, N. (2012). "Relative efficiency of morphological characters and molecular markers in the establishment of an apricot core collection." *Hereditas*, 149 (5) : 163-172.

Kumar, S., Skjæveland, A., Orr, R.J.S., Enger, P., Rude, T., Mevik, B.H., Burki, F., Botnen, A., Shalchian-Tabrizi, K. (2009). "AIR: A batch-oriented web program package for construction of supermatrices ready for phylogenomic analyses." *BMC Bioinformatics*, 10:357.

Lachkar, A., Mars, M. (2012). "Assessment of Tunisian apricot local and hybrids cultivars for new breeding program." *Annals of Biological Research*, 3(8): 4108-4118.

Layne, R.E.C., Bailey, C. H., Hough, L.F. (1996). Apricots. In Janick J. and Moore J.N. (eds). *Fruit Breeding : Tree and Tropical Fruits*, Vol. 1. John Wiley and Sons Inc., New York, 79-111.

Lichou, J., Jay, M. (2012). *Monographie Abricot*. Paris, CTIFL, 568 p.

Logozzo, G., Donnoli, R., Macaluso, L., Papa, R., Knüpfper, H., Spagnoletti Zeuli, P. (2007). "Analysis of the contribution of Mesoamerican and Andean gene pools to European common bean (*Phaseolus vulgaris* L.) germplasm and strategies to establish a core collection." *Genetic Resources and Crop Evolution*, 54 (8) : 1763-1779.

Lopes, M. S., Sefc, K.M., Laimer, M., Da Câmara Machado, A. (2002). "Identification of microsatellite loci in apricot." *Molecular Ecology Notes*, 2(1): 24-26.

Loveless, M. D., Hamrick, J. L. (1984). "Ecological determinants of genetic structure of plant populations." *Annual Review of Ecology and Systematics*, 15: 65-95.

- Maghuly, F., Borroto Fernandez, E., Ruthner, S., Pedryc, A., Laimer, M. (2005). "Microsatellite variability in apricots (*Prunus armeniaca* L.) reflects their geographic origin and breeding history." *Tree Genetics & Genomes*, 1(4): 151-165.
- Mehlenbacher, S., Cociu, V, Hough, LF (1991). "Apricots (*Prunus*). Genetic resources of temperate fruit and nut crops", *Acta Horticulturae*, 290: 65-110.
- Messina, R., Lain, O., Marrazzo, M.T., Cipriani, G., Testolin, R. (2004). "New set of microsatellite loci isolated in apricot." *Molecular Ecology Notes*, 4(3): 432-434.
- Mnejja, M., Garcia-Mas, M., Howad, W., Badenes, M.L., Arus, P. (2004). "Simple-sequence repeat (SSR) markers of Japanese plum (*Prunus salicina* Lindl.) are highly polymorphic and transferable to peach and almond." *Molecular Ecology Notes*, 4(2): 163-166.
- Mnejja, M., Garcia-Mas, J., Howad, W., Arus, P. (2005). " Development and transportability across *Prunus* species of 42 polymorphic almond microsatellites " *Molecular Ecology Notes*, 5(3): 531-535.
- Morot-Gaudry, J.F. et Briat, J.F. (2007). *Functional Plant Genomics*. Science Publishers, Enfield, USA, 708 p.
- Nair, N.V., Balakrishnan, R., Sreenivasan, T.V. (1998). "Variability for quantitative traits in exotic hybrid germplasm of sugarcane." *Genetic Resources and Crop Evolution* 45: 459–464.
- Powell, W., Machray, GC., Provan, J. (1996). "Polymorphism as revealed by simple sequence repeats." *Trends in Plant Science*, 1: 215-222.
- Pritchard, J.K., Stephens, M., Donnelly, P. (2000). "Inference of Population Structure Using Multilocus Genotype Data." *Genetics*, 155 (2) : 945-959.
- Rafalski A, Morgante M. (2004). "Corn and humans: recombination and linkage disequilibrium in two genomes of similar size." *Trends in Genetics*, 20 (2) : 103-111.
- Raymond, M., Rousset, F. (1995). GENEPOP (Version 1.2): "Population Genetics Software for Exact Tests and Ecumenicism." *The Journal of Heredity*, 86 (3) : 246-249.

Rehder, A. (1940). Manual of cultivated trees and shrubs hardy in North America exclusive of the subtropical and warmer temperate regions. New York, Macmillan, 996 p.

Rehder, A. (1949). Bibliography of cultivated trees and shrubs hardy in the cooler temperate regions of the Northern hemisphere. Boston, The Arnold Arboretum of Harvard University, 300 p.

Ricciardi, L., Giorgio, V., De Giovanni, C., Lotti, C., Gallotta, A., Fanizza, G. (2002). "The genetic diversity of Apulian apricot genotypes (*Prunus armeniaca* L.) assessed using AFLP markers." Cellular & molecular biology letters, 7: 431-436.

Romero, C., Pedryc, A., Muñoz, V., Llácer, G., Badenes, M.L. (2003). "Genetic diversity of different apricot geographical groups determined by SSR markers." Genome, 46(2): 244-252.

Rosenberg, N.A. (2004). "Distruct: a program for the graphical display of population structure." Molecular Ecology Notes, 4: 137-138.

Ruthner, S., Pedryc, A., Kriska, B., Romero, C., Badenes, M.L. (2006). "Molecular characterization of apricot (*Prunus armeniaca* L.) cultivars using cross species SSR amplification with peach primers." International Journal of Horticultural Science, 12(3): 53-57.

Saitou, N., Nei, M. (1987). "The neighbor-joining method: a new method for reconstructing phylogenetic trees." Molecular Biology and Evolution, 4 (4) : 406-425.

Sanchez-Perez, R., Ruiz, D., Dicenta, F., Egea, J., Martinez-Gomez, P. (2005). "Application of simple sequence repeat (SSR) markers in apricot breeding: molecular characterization, protection, and genetic relationships." Scientia Horticulturae, 103(3): 305-315.

Sosinski, B., Gannavarapu, M., Hager, L.D., Beck, L.E., King, G.J., Ryder, C.D., Rajapakse, S., Baird, W.V., Ballard, R.E., Abbott, A.G. (2000). " Characterization of microsatellite markers in peach [*Prunus persica* (L.) Batsch] " Theoretical and Applied Genetics, 101(3): 421-428.

Struss, D. A., R., Southwick, S.M., Boritzki, M. (2003). " Analysis of sweet cherry (*Prunus avium* L.) cultivars using SSR and AFLP markers " Journal of the American Society for Horticultural Science, 128(6): 904-909.

- Tanksley, S. D., Orton, T.J. (1983). Isozymes in plant genetics and breeding. Part A. New York, Elsevier, 526 p.
- Testolin, R., Marrazzo, T., Cipriani, G., Quarta, R., Verde, I., Dettori, M.T., Pancaldi, M., Sansavini, S. (2000). "Microsatellite DNA in peach (*Prunus persica* L. Batsch) and its use in fingerprinting and testing the genetic origin of cultivars." *Genome*, 43(3): 512-520.
- Testolin, R., Messina, R., Lain, O., Marrazzo, M.T., Huang, W.G., Cipriani, G. (2004). "Microsatellites isolated in almond from an AC-repeat enriched library " *Molecular Ecology Notes*, 4(3): 459-461.
- Vaughan, S. P., Russell, K. (2004). "Characterization of novel microsatellites and development of multiplex PCR for large-scale population studies in wild cherry, *Prunus avium*." *Molecular Ecology Notes*, 4(3): 429-431.
- Vavilov, N. I. (1951). *Phylogeographic basis of plant breeding. The origin, variation, immunity and breeding of cultivated plants.* Londres, Dawson and Sons: 13-54.
- Vavilov, N. I. (1992). *Origin and Geography of Cultivated Plants.* Cambridge, Cambridge University Press, 505 p.
- Vilanova, S., Romero, C., Abbott, A.G., Llácer, G., Badenes, M.L. (2003). "An apricot (*Prunus armeniaca* L.) F2 progeny linkage map based on SSR and AFLP markers, mapping plum pox virus resistance and self-incompatibility traits." *Theoretical and Applied Genetics*, 107(2): 239-247.
- Vilanova, S., Soriano, J.M., Lalli, D.A., Romero, C. Abbott, A.G., Llacer, G., Badenes, M.L. (2006). "Development of SSR markers located in the G1 linkage group of apricot (*Prunus armeniaca* L.) using a bacterial artificial chromosome library." *Molecular Ecology Notes*, 6(3): 789-791.
- Wang, L. *et al.* (2006). "Establishment of Chinese soybean *Glycine max* core collections with agronomic traits and SSR markers." *Euphytica*, 151 (2) : 215-223.
- Weir, B. S., Cockerham, C. C. (1984). "Estimating F-statistics for the analysis of population structure." *Evolution*, 1358-1370.

Wünsch, A. (2008). "Cross-transferable polymorphic SSR loci in *Prunus* species." *Scientia Horticulturae*, 120(3): 348-352.

Yamamoto, K., Mochida, T., Imai, Y.Z., Shi, I., Ogiwara, T., Hayashi, T. (2002). "Microsatellite markers in peach [*Prunus persica* (L.) Batsch] derived from an enriched genomic and cDNA libraries." *Molecular Ecology Notes*, 2(3): 298-301.

Zhebentyayeva, T., Reighard, G., Gorina, V., Abbott, A. (2003). "Simple sequence repeat (SSR) analysis for assessment of genetic variability in apricot germplasm." *Theoretical and Applied Genetics*, 106(3): 435-444.

## ***7.2 Liste sitographique***

FAOSTAT. Consulté en novembre 2012.

<http://faostat.fao.org/site/339/default.aspx>

GNIS. "Les marqueurs moléculaires." Consulté le 23/10/2012.

<http://www.gnis-pedagogie.org/biotechnologie-amelioration-marqueur-moleculaire.html>

Thompson, B., Structure and Distruct Tutorial. Consulté en mars 2013.

<http://bodegaphylo.wikispot.org/structure>

Université d'Oslo, Bioportal. Consulté en janvier et février 2013.

<http://www.bioportal.uio.no/about.php>



Diplôme : ingénieur en horticulture  
Spécialité :  
Spécialisation / option : GDV  
Enseignant référent : Claudie Lambert

Auteur(s) : Adrien Lasnier

Organisme d'accueil : INRA Avignon & BioGEVES  
Adresse : INRA GAFL Domaine St Maurice

Date de naissance\* : 16/04/1988

67 allée des chênes 84143 Montfavet

Nb pages : 53      Annexe(s) :

BioGEVES - Domaine du Magneraud  
17700 Surgères

Année de soutenance : 2013

Maîtres de stage : J-M. Audergon et A. Bernole

Titre français : Analyse de la diversité génétique de l'abricotier (*Prunus armeniaca* L.)

Titre anglais : Analysis of genetic diversity of apricot (*Prunus armeniaca* L.)

Résumé (1600 caractères maximum) :

Ce travail a consisté à étudier la diversité et la structure génétique de l'abricotier (*Prunus armeniaca* L.), en utilisant des marqueurs microsatellites. 25 marqueurs microsatellites, choisis pour leur répartition sur la carte de référence de *Prunus*, ont servi à génotyper 580 accessions d'abricotier. Ces accessions ont été ajoutées à 673 accessions génotypées précédemment, pour former un ensemble de 1253 profils génétiques d'abricotiers, originaires principalement d'Europe, du Maghreb, d'Amérique du Nord, d'Asie centrale et d'Asie orientale. Outre l'établissement d'un profil génétique par variété, l'analyse des données par les logiciels GENETIX et STRUCTURE a permis de confirmer l'existence d'une structure phylogénique très marquée depuis le centre d'origine vraisemblablement situé en Asie centrale. *A minima* 3 voies de diversification orientées respectivement vers l'Asie de l'Est, le Bassin Méditerranéen et l'Europe centrale ont été identifiées. De manière complémentaire, la comparaison entre les accessions dites patrimoniales, qui ont un lien ancien avec leur région de culture, et les accessions hybrides, issues de programmes d'amélioration, a permis d'identifier un déficit de diversité chez les variétés hybrides, qui pourrait être diminué par l'apport de variétés asiatiques dans les programmes d'hybridation.

Abstract (1600 caractères maximum) :

The present work consisted in studying the apricot (*Prunus armeniaca* L.) genetic diversity and phylogenetic structure through SSRs markers fingerprinting. 25 microsatellite markers, selected to cover the *Prunus* reference map, were used to genotype 580 apricot accessions. These accessions, and 673 previously genotyped accessions, were gathered to constitute a set of 1253 apricot genetic profiles, most of them coming from Europe, Maghreb, North America, Central Asia and East Asia. Out of the identification of the related accessions, the whole data set has been analyzed with GENETIX and STRUCTURE softwares. An evident phylogenetic structure has been shown from the origin center probably localized in Central Asia. At least 3 major diversification phylum have been identified dedicated to Eastern Asia, Mediterranean basin and Central Europe. In addition a comparative analysis of both local cultivars and improved cultivars by breeding has been initiated. This comparison emphasized a lack of genetic diversity among the hybrid varieties, which could be reduced by the use of Asiatic varieties in the hybridization programs.

Mots-clés : diversité génétique, microsatellites, *Prunus armeniaca* L., structure génétique.

Key Words: genetic diversity, genetic structure, microsatellites, *Prunus armeniaca* L.