



HAL
open science

Evolutionary history of apricot (*Prunus armeniaca* L.) and impact of different processes of evolution on genetic diversity

Shuo Liu

► **To cite this version:**

Shuo Liu. Evolutionary history of apricot (*Prunus armeniaca* L.) and impact of different processes of evolution on genetic diversity. Agricultural sciences. Université de Bordeaux, 2019. English. NNT : 2019BORD0190 . tel-02421879v1

HAL Id: tel-02421879

<https://theses.hal.science/tel-02421879v1>

Submitted on 20 Dec 2019 (v1), last revised 2 Jan 2020 (v2)

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

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

THÈSE PRÉSENTÉE
POUR OBTENIR LE GRADE DE
DOCTEUR DE
L'UNIVERSITÉ DE BORDEAUX

ÉCOLE DOCTORALE des Sciences de la Vie et de la Santé
SPÉCIALITÉ Biologie végétale

Par

Shuo LIU

**Histoire évolutive et impact des différents processus évolutifs
sur la diversité génétique de l'abricotier (*Prunus armeniaca* L.)**

Sous la direction de : **Véronique DECROOCQ**

Soutenue le 24 Octobre 2019

Membres du jury :

M. HERNOULD Michel	Professeur, Université de Bordeaux	Président
Mme RONFORT Joelle	Directrice de Recherche, INRA	Rapporteuse
M. GLEMIN Sylvain	Directeur de Recherche, CNRS	Rapporteur
Mme GRAPIN Agnès	Maître de Conférences, AgroCampus Ouest	Examinatrice
M. ARUS Père	Docteur, IRTA	Examineur

“Any man who is really a man must learn to be alone in the midst of others, to think alone for others, and, if necessary, against others”

Romain Rolland

致谢 (THANKS)

人的一生中，最光辉的一天并非是功成名就那天，而是从悲叹与绝望中产生对人生的挑战，以勇敢迈向意志那天。

在法国的学习生活经历注定在我人生中留下一抹明亮的痕迹。四年的博士学习带给我许多财富，这些财富包含丰富的科学知识、缜密的思维逻辑、可贵的分享意识及难得的科学精神等。此期间无数次的过程，从迷惑不解到豁然开朗，从妄自菲薄到满怀信心，从不切实际到脚踏实地，从失望无助到独立自强，从固执己见到从善如流，都离不开身边的同事、朋友和家人的帮助和理解。

我的博士导师，Veronique Decroocq，一位法国女科学家，是我在法国最为尊重、喜爱和崇拜的良师益友。「古之学者必有师。师者，所以传道受业解惑也」，及「授人以鱼不如授人以渔」这两句中国古谚语用于概括她在我博士学习的贡献，最为恰当。每遇到学习困难，她总会给我鼓励打气，倾听我的想法并提出建议，协调团队人员积极协助。从她那里获得的科学逻辑的训练和独立判断的培养，对我而言受益一生。通过多年的师生相处，我们建立了很好的彼此信任，十分珍贵。我非常感激她。

工作在波尔多和巴黎 INRA 的法国同事对我博士的学习帮助巨大。感谢在巴黎 INRA 工作的 Amandine Cornille 和巴黎第十一大学的 Tatiana Giraud 两位科学家，同时也是我博士学术委员会成员在群体遗传学、统计学和生物信息学方面的细心指导帮助，让我更加全面地了解我所从事的课题研究并拓宽了我的研究领域，接触到更多有趣的科学知识。感谢波尔多 Genetics 团队 Stephane Decroocq 在我博士学习中的数据分析和帮助，让我学习到了很多高效方便的方法，获得了准确可靠的结果。感谢团队的 Aurelie Chague，David Tricon 和 Jean-Philippe Eyquard 在我博士实验过程中提供的数据来源和样品材料。感谢法国波尔多大学的 Alexis Groppi 和 Barre Aurelien 对我数据处理方面极大的支持和帮助。此外，我要感谢 INRA-BFP 的 Thierry Candresse 所长接受我的博士学习申请，Luc Sofer 和 Chantal Faure 等同事在我博士学习期间对我学习、生活的关心。感谢我博士学术委员会成员 Rémy Petit，博士毕业答辩过程中论文审阅人 Joelle Ronfort 和 Sylvain Glémin，及答辩评审成员 Père Arus, Agnès Grapin 和 Michel Hernould 的付出和努力。

友情无处不在，它伴随你左右，萦绕在你身边，和你共渡一生。留学生活不仅是一段学习经历，同时也是一种人生历练。感谢一同留法朋友马宇欣、蒋建桥、姜南、樊磊、李小明等在我博士学习过程中的帮助和建议，及在法生活中分享的忧愁、烦恼和快乐。这份友情

的光辉也将在我的人生痕迹中闪耀。虽然身处异乡，但来自家人的关心、关爱始终没有被距离阻断过。感谢我的妻子刘有春，在我攻读博士期间对家庭、孩子和父母的悉心照顾和辛苦的付出，及对我的关爱。每逢最孤独、寂寞的低谷时，总有来自妻子的问候让我倍感欣慰。感谢父母和长辈对我的鼓励和支持，父母的关心是我不断前行的动力，刘威生老师对我攻读博士的支持及对我家庭生活的长期帮助，是我可以安心读书的重要保障。此外，我也要感谢辽宁省农业科学院、辽宁省果树科学研究所和李杏研究室的领导和同事，在这四年中替我分担了许多国内的工作和事务。

最后，我要感谢所有踏入我人生轨迹的人们，无论在过去还是未来，我的世界将因你们而变得精彩。

刘 硕

二零一九年秋，法国

Valorization and dissemination of my PhD work

Published article

Decroocq S, Cornille A, Tricon D, Babayeva S, Chague A, Eyquard J-P, Karychev R, Dolgikh S, Kostritsyna T, **Liu S**, Liu W, Geng, W, Liao K, Asma B-M, Akparov Z, Giraud T, Decroocq V (2016). New insights into the history of domesticated and wild apricot and its contribution to *Plum pox virus resistance*. *Molecular Ecology*, doi: 10.1111/mec.13772.

Liu S, Cornille A., Decroocq S., Tricon D., Chague A., Eyquard J-P, Liu W-S., Giraud T., Decroocq V. (2019) The complex evolutionary history of apricots: species divergence, gene flow and multiple domestication events. *Molecular Ecology*. doi:10.1111/mec.15296

Submitted article

Liu S, Decroocq S., Tricon D., Chague A., Decroocq V. (2019) Genetic diversity and structuration analyses in the French Alpen apricot (*Prunus brigantina*) reveal phylogenetic inconsistencies in the *Armeniaca* section. Deposited in BioRxiv as soon as the previous paper is accepted. Submission planned in *Tree Genetics & Genomes* or *BMC Evolutionary Biology*.

Oral communications

2018.06.29 - 9th International Rosaceae Genomics Conference (RGC9), Nanjing, China
Title: "Origin and Evolution of *Armeniaca* species".

2019.01.12 - XXVIIth Plant and Animal Genome Conference (PAG), San Diego, USA
Title: "New Insights into the Genome Evolution and Domestication of Apricots".

2019.10.08 – 5eme Colloque de Génomique Environnementale, La Rochelle, France
Title: "New sights into selection under genome evolution and domestication of apricot".

Poster presentation

2019.01.12 - XXVIIth Plant and Animal Genome conference (PAG), San Diego, USA
Title: "New insights into the genome evolution and domestication of apricots".

2016.04.06 - Journée de l'Ecole Doctorale, Bordeaux, France
Title: "Evaluation of genetic structuration of the apricot germplasm across Eurasia by SSR markers (microsatellites)".

Award

Earl J. Scherago Award at the XXVIIth Plant and Animal Genome conference, San Diego (CA, USA), January 2019.

Funding and grant support

PhD grant (4 years periods) awarded by the Chinese government scholarship program, China Scholarship Council (CSC).

Grant for conference from “Fédération de recherche Biologie Intégrative et Ecologie”, supported by INRA and University of Bordeaux.

Funding for the scientific exchanges from the Program Xu guangqi 2016 & 2018 by the Ministry of Europe and Foreign Affairs (MEAE) for expedition in 2016 and participation to RGC9 in 2018.

Scholarship from 5eme Colloque de Génomique Environnementale, La Rochelle, France, 8 to 10 of October 2019

PROLOGUE

In the past few decades, scientists postulated that some of the evolutionary processes differ significantly between perennial and annual species. Compared to annuals, long-lived species (trees, in the case of this PhD manuscript) experience extended juvenility periods, restricted sexual but overlapping generations, extensive outcrossing and widespread intra- and inter-specific hybridization as well as limited population structure. Based on the above life traits and mode of reproduction, slow rates of evolution in perennial species are expected. However, they still have to cope with dramatic seasonal and cyclic climate changes. Therefore, under the long-term process of natural selection, perennial tree plants have experienced complex historical events, such as migration or bottleneck, together with adaptation to new environments and climatic conditions. Little is known on how perennial populations respond to selection pressures operating on a relatively short time scale in comparison with the tree lifespan. Those selection pressures can be natural or caused by human activities (artificial). Both adaptation and selection are expected to have impacted the perennial population dynamics but also gene and genome evolution. However, the extent of this impact is still under question.

Perennial crops are important components of agricultural economies in particular in many low to mid-income countries where they represent a substantial portion of the normal dietary intake. Additionally, non-food perennials are also significant economic and ecological drivers, being important parts of many forest ecosystems worldwide and as a tree species, they are important for land preservation and carbon assimilation. Among long-lived (>5 years lifespan) species, the majority of domesticated perennials are trees, which are cultivated for their edible fruits. Here, the term 'fruit tree' will be used to refer to perennial species that are domesticated or not but in which some components of the fruit are used by humans or animals. Long-lived perennial fruit tree species were used and later domesticated by humans in all major agricultural centres but we will focus in this PhD manuscript on the apricot fruit tree, *Prunus armeniaca* L., and its related species and wild progenitors.

Historically, perennial plants were considered intractable systems for studying the genetic architecture underlying evolution and domestication. However, recently developed and constantly updated technologies (e.g., molecular genotyping, next-generation sequencing) and statistical methods (e.g., Bayesian and Approximate Bayesian Computation, new demographic inference methods), in conjunction with large breeding and botanical collections, are now facilitating detailed evolutionary analyses in perennial species. To provide a better understanding on fruit tree evolutionary history, we used apricot as a

model species because of the availability of both cultivated and wild forms and addressed a few important questions related to fruit tree origin, evolution and domestication. This PhD manuscript is organized in three main chapters, as follows. Chapter I (**Investigation on the complex evolutionary history of apricots: species divergence, gene flow and multiple domestication events**) aims at studying genetic features and demographic history of cultivated and wild apricot species across Eurasia, with an emphasis on Chinese cultivated apricots and their wild relatives. Chapter II (**Genetic diversity and structuration analysis in the French Alpen apricot (*Prunus brigantina*) reveal phylogenetic inconsistencies in the *Armeniaca* section**) is in the continuity of the previous chapter, but this time we estimate the relatedness between apricot and its only European wild relative, *P. brigantina* (or “Marmottier des Alpes”) with a focus on their use in fruit tree conservation programs. Finally, in the last and third chapter (**Distinct evolution and domestication processes in apricot (*Prunus armeniaca*) revealed by different patterns of selection**), we focus on how selection has influenced genomic architecture in apricot. To test for common or distinct signatures of selection, we took advantage of the parallel history of domestication in the European and Chinese apricots and compared with their wild, Central Asian progenitor. The outcomes of my PhD manuscript are expected to bring new insights about the process of evolution and domestication in long-lived tree species thus emphasizing potential differences or similarities between annual and perennial species.

Remarks: *In the General Introduction and Conclusion, the numbering of figures and tables will be incremented from Figure 1 to 7 and Table 1. Figures and tables will be embedded in the text. For the three chapters, they will be placed at the end of the section, after the references. In the three cases, the numbering of figures and tables will be specific to each chapter; it will always depict first the number of the chapter in capital roman letters followed by the incremental number (e.g., Figure I-1 is the first figure of Chapter I). Supplemental files and tables for each chapter are placed in annexes, at the bottom of the PhD manuscript.*

Concerning the publications cited in the text, each chapter will have its own section named ‘References’ at the end of each article while a specific bibliography for the General Introduction and Conclusion is listed at the end of the PhD manuscript.

The most voluminous tables are available at Dataverse, the INRA data portal (<https://data.inra.fr/>)

TABLE OF CONTENTS

GLOSSARY	1
List of Figures	3
List of Tables.....	6
GENERAL INTRODUCTION	1
I- Species: Concepts and definitions	10
II- Origin of genetic diversity	11
II-1 How to estimate and characterize genetic diversity?	11
II-2 The Hardy-Weinberg equilibrium.....	12
II-3 Evolutionary processes that influence genetic diversity	12
II-3-1 The particular case of interspecific hybridization	13
II-3-2 Different evolutionary processes of selection	14
III- The case-study of the domestication of long-lived, perennial crops.....	16
III-1-1 Mode(s) of reproduction	17
III-1-2 Outcrossing mating system	18
III-2 A loss of genetic diversity during perennials' domestication?	19
III-3 Common traits targeted by domestication in long-lived perennial plants	20
Problématique de thèse	23
Le(s) modèle(s) d'étude : L'abricotier et ses espèces apparentées	23
L'espèce <i>Prunus armeniaca</i> (L.): Position systématique.....	23
Répartition et écologie des espèces <i>Armeniaca</i> non-domestiquées	24
Origine de l'abricotier cultivé.....	26
Objectifs de la thèse	29
PhD context and objectives	31
Apricot, a long-lived perennial tree	31
The focus of the PhD thesis	32
CHAPTER I	34
Abstract	35
Introduction.....	36
Material and Methods	38
Sample collection and DNA extraction	38
Microsatellite markers and polymerase chain reaction (PCR) amplification.....	39
Analyses of population structure	39
Genetic variation and differentiation	40
Inference of demographic history	41
Results.....	43
Population genetic diversity and structure.....	43
Genetic variation and differentiation among the apricot genetic clusters	45

Approximate Bayesian computation for testing demographic scenarios	46
Discussion	49
Diversification of wild apricots in Central and Eastern Asia.....	49
Multiple independent domestication events of apricots across Central and East Asia...	50
Conclusion.....	52
Acknowledgements	52
Data accessibility.....	53
Author contributions.....	53
References	53
Figures.....	60
Tables.....	63
CHAPTER II.....	65
Abstract	66
Introduction.....	67
Materials and Methods	69
<i>In situ P. brigantina</i> sampling and DNA extraction	69
Microsatellite markers and polymerase chain reaction (PCR) amplification.....	69
Analyses of population subdivision and genetic relationship.....	69
Genetic diversity, differentiation and core collection constitution	70
Testing isolation by distance	70
Results and discussion.....	71
Genetic variability and structuration in <i>P. brigantina</i> population.....	71
Genetic relationship between <i>P. brigantina</i> and related <i>Armeniaca</i> species.....	72
Construction of a <i>P. brigantina</i> ex-situ core collection.....	73
Conclusion.....	74
Acknowledgements	74
Data availability	74
Reference	75
Figures.....	79
Tables.....	83
CHAPTER III.....	84
Foreword	85
Introduction on selective sweeps	86
Choosing an Adequate test for the identification of selective sweeps.....	86
Advantages and disadvantages of each test.....	87
Methods used for selective sweep detection in this study.....	88
The role of demography in selective sweep detection accuracy	89
Material and Methods	89

Sample preparation and sequencing.....	89
Raw read trimming, alignment, and filtering	90
SNP variant calling and filtering	90
Linkage disequilibrium (LD).....	91
Identification of the apricot genetic groups and the analysis of their structuration	91
Genetic diversity and demographic history	92
Identification of genomic signatures of selection	92
Potential traits/genes associated with selective sweeps	93
Results.....	93
Data filtering	93
Linkage disequilibrium in apricot	94
Inferred genetic admixture and sub-clusters within cultivated and wild apricot groups ..	94
Genetic diversity of different apricots and their inferred subclusters	95
Demographic history of apricot.....	96
Genomic signatures of selective sweeps within the apricot genome	96
Putatively selected genes and pathways	99
Discussion	102
Genomic regions under natural selection provide extensive insights into the evolutionary history of wild apricot.....	104
Selective sweeps and agronomically important genes.....	104
Conclusion.....	105
Acknowledgements	106
Reference	107
Figures.....	117
Tables.....	128
GENERAL DISCUSSION and CONCLUSION	138
I- Evolutionary processes with a role in <i>P. armeniaca</i> history	139
A common origin for the Armeniaca species?.....	139
Evolutionary processes towards diversification in wild apricot species	140
A proposed model for the evolutionary pathway of cultivated apricots.....	142
Multiple dispersion routes for the domesticated/pre-domesticated apricots.....	143
From forest to orchard, the importance of local adaptation in the domestication process	144
Wild-to-crop and/or interspecific introgression as a source of local adaptation and diversification in Central Asian and Chinese cultivated apricots?	146
II- Comparative genomics of apricot domestication and adaptation	147
A fine tuning of flowering time during apricot adaptation.....	148
Convergent genomic signatures of domestication in Chinese and European apricots	150
Outlook and outstanding questions for the future.....	152

References	154
ANNEXES	164
Annex 1: Chapter I supporting information	165
Supplemental notes.....	165
Supplementary note 1: Cautionary notes on the assignment of apricot individuals as 'wild' or 'cultivated' (landraces and/or modern cultivars).....	165
Supplementary note 2: General framework for defining focal population for ABC analyses	165
Supplementary note 3: Sets of scenarios to be compared using ABC	166
Supplementary note 4: Linear Discriminant Analyses (LDA)	167
Supplementary note 5: Parameter estimates inferred from ABC analyses.....	168
Chapter I supplemental Figures	169
Chapter I supplemental Tables	177
Annex 2: Chapter II supporting information	188
Chapter II supplemental Figures	188
Chapter II supplemental Tables	192
Annex 3: Chapter III supporting information.	196
Supplementary notes on the <i>de novo</i> assembled apricot genome.....	196
Chapter III supplemental Figures	199
Chapter III supplemental Tables	203
Résumé:	211
Abstract:	212

GLOSSARY

Allele frequency: refers to how frequent an allele is in a population. It is obtained by counting how many times the allele appears in the population then it is divided by the total number of copies of the gene in the population. It reflects the genetic diversity of a population or equivalently the richness of the gene pool of interest or the species.

Ancestor: any organism, population, or species from which some other organism, population, or species is descended by reproduction.

Artificial selection: is the selective breeding carried out by humans to alter a population.

Balancing selection: a selection that maintains different alleles within a population.

Bottleneck: an environmentally mediated point which dramatically decreases population size. By extension, a population **bottleneck** or **genetic bottleneck** is a sharp reduction in the size of a population.

F-statistics: also known as fixation indices describe the statistically expected level of heterozygosity in a population.

F_{ST}: or Wright's fixation index is the proportion of the total genetic variance contained in a subpopulation (the S subscript) relative to the total genetic variance (the T subscript). It allows estimating the proportion of overall diversity in a sample across different populations that are attributable to between-population divergence.

Gene diversity: is the proportion of polymorphic loci across the genome.

Genetic diversity: is the total number of genetic parameters (nucleotide polymorphisms, copy number variations, ploidy etc...) in the genetic makeup of a species.

Genetic variability: is the tendency of genetic parameters to vary within a population or a species, usually assessed at three levels: (a) within breeding populations, (b) between breeding populations, and (c) within species.

Gene flow: refers to the exchange of genes (in one or both directions) at a low rate between two populations, due to the dispersal gametes or of individuals from one population to another, also called migration.

Genetic drift: is the random change in allele frequency from one generation to the next. Genetic drift has much more effective in small populations, which may have an allele drift to fixation, in which all members share the same allele.

Heterozygosity: the state of a diploid locus in which different alleles are present at the two copies of that locus.

Hardy-Weinberg Equilibrium (HWE): is a principle stating that the genetic variation in a population will remain constant from one generation to the next in the absence of disturbing factors. Overtime an allele's homozygous versus heterozygous frequencies are expected to obey an independent, binomial model.

Local adaptation: designates a better average performance of individuals born in the habitat in which the measure is done, compared with the performance of immigrants.

Locus: a locus is a fixed position on a chromosome that is occupied by a given gene or one of its alleles.

Microsatellites: a polymorphism characterized by a variable number of tandem repeats often defined by the number of repeats in a row of at least two or more nucleotides.

Migration: in the genetic sense of permanent movement of genes from one location to another, into or out of a population.

Mutation: is a random heritable change in a gene or chromosome, resulting from the addition, deletion, or substitution of nitrogenous bases (nucleotides) in the DNA sequence.

Natural selection: the process by which those organisms better adapted to their environment increase in frequency relative to less well-adapted forms over a number of generations.

Non-random mating: in non-random mating, organisms may prefer to mate with others of the same genotype or of different genotypes. Non-random mating won't make allele frequencies in the population change by itself, though it can alter genotype frequencies.

Effective population size (N_e): was defined by Wright as the number of breeding individuals in an idealized population that would show the same amount of dispersion of allele frequencies under random genetic drift or the same amount of inbreeding as the population under consideration.

Outgroup: in a cladistics analysis, any taxon used to help resolve the polarity of characters, and which is hypothesized to be less closely related to each of the taxa under consideration than any are to each other.

Outlier: is a data point that differs significantly from other observations.

Polymerase Chain Reaction (PCR): a powerful method for amplifying specifically DNA segments.

Population expansion: increase in the size of a population over time. Historical population expansion can affect patterns of current polymorphism.

Population subdivision: population structure in which barriers exist to prevent random mixing and mating between members of a species.

Positive selection: selection favouring an allele that is advantageous in some or all circumstances.

Selective sweep: is the reduction or elimination of variation among the nucleotides near a mutation in DNA. It results from a beneficial allele having recently reached fixation due to strong positive selection.

Self-incompatible: a plant incapable of self-fertilization because its own pollen is prevented from germination on the stigma or because the pollen tube is blocked before it reaches the ovule.

Single nucleotide polymorphism (SNP): DNA sequence variations involving alternative single nucleotide (A, T, C, or G) residues.

List of Figures

GENERAL INTRODUCTION

Figure 1. The analogy of evolution force.

Figure 2. Modes of natural selection and molecular signatures of each selective regime.

Figure 3. Percentage of retained genetic variation in domesticated annual (green) and perennial (purple) fruit crop populations compared to their wild relatives.

Figure 4. *Armeniaca* species and morphological features of fruit, flowers and tree.

Figure 5. *Armeniaca* species worldwide distribution.

Figure 6. Map with the diverse routes of apricot dispersion, including EUR, North-America, South-America, South-Africa and Australia.

GENERAL DISCUSSION and CONCLUSION

Figure 7. The domestication process in Eurasian cultivated *P. armeniaca*.

Chapter I:

Figure I-1. Genetic structure among the 577 samples (271 cultivated and 306 wild samples) of the five *Prunus* species (*P. armeniaca*, *P. sibirica*, *P. mandshurica*, *P. mume* and *P. brigantina*) analyzed in this study with 34 microsatellite markers.

Figure I-2. Factorial correspondence analyses (FCAs).

Figure I-3. *Armeniaca* phylogenetic networks obtained from Splitstree.

Figure I-4. Most likely scenarios for the divergence and domestication of apricot inferred with approximate Bayesian computation.

Figure I-S1. The distribution of coefficient memberships for the various genetic groups.

Figure I-S2. Twelve competing scenarios of divergence history among wild *Prunus* populations tested by random forest approximate Bayesian computation (ABC analysis step 1– Wild divergence).

Figure I-S3. Eighteen competing scenarios of apricot domestication tested by random forest approximate Bayesian computation (ABC analysis steps 1 and 2).

Figure I-S4. Bayesian clustering inferred with STRUCTURE for $K=2$ to $K=10$ for the 288 wild apricots ($N=288$) sampled in 19 natural populations of Central Asia and China.

Figure I-S5. Bayesian clustering inferred with STRUCTURE for $K=2$ to $K=12$ based on the whole apricot dataset (excluding *Prunus mume*, *Prunus mandshurica* and *Prunus brigantina*, thus $n=496$) using 25 microsatellite markers.

Figure I-S6. Bayesian clustering obtained with STRUCTURE for the wild and cultivated apricots using 34 microsatellite markers, from $K=2$ to $K=12$, based on the whole dataset ($n=577$).

Figure I-S7. Delta K plotted against K values for each set of STRUCTURE analyses.

Figure I-S8. Bayesian clustering for $K=2$ to $K=10$ obtained with STRUCTURE for *Prunus sibirica* individuals ($n=84$) using 25 microsatellite markers originating from seven natural populations of Northern China.

Figure I-S9. Projection of the microsatellite apricot simulated datasets on the first two linear discriminant analysis (LDA) axes to infer the apricot speciation and domestication histories for each ABC step.

Figure I-S10. Fourteen scenarios with the occurrence of bottleneck during apricot domestication tested with approximate Bayesian computation (ABC analysis step 3).

Chapter II:

Figure II-1. Taxonomy and geographic distribution of the different species of section *Armeniaca*.

Figure II-2. *Prunus brigantina* morphological features, genetic clustering and spatial distribution over the French Alps.

Figure II-3. Principal components analysis on *P. brigantina*.

Figure II-4. Weighted neighbour-joining (NJ) tree of *P. brigantina* and other *Prunus* species.

Figure II-5. The core collection of *P. brigantina* population based on the M strategy.

Figure II-S1. Detect the appropriate number of model components in 71 *P. brigantina* accessions.

Figure II-S2. Bayesian clustering on 71 *P. brigantina* samples in French Alpen Mountains.

Figure II-S3. Isolation by distance (IBD) test for the three *P. brigantina* populations.

Figure II-S4. Bayesian analysis on European/Irano-caucasian cultivated *P. armeniaca* and wild *P. brigantina* accessions.

Figure II-S5. Bayesian analysis of current *P. brigantina* and large apricot dataset from the previous study.

Figure II-S6. Classification between *P. brigantina* and other *Prunus* species.

Figure II-S7. Detect the relevant *K* clusters from *P. brigantina* and other *Prunus* species dataset.

Annex 3:

Figure SN-1. Overview of the processing pipeline used for the *de novo* assembly of the Marouch genome.

Chapter III:

Figure III-1. Genomic signatures of hard and soft sweeps arising from one versus multiple mutations.

Figure III-2. Workflow for the detection of genomic regions under selection in wild and cultivated apricot genomes.

Figure III-3. Quality parameters of the filtered SNPs dataset used in this study.

Figure III-4. Linkage disequilibrium decay as measured by the squared correlation coefficients (r^2) between all pairs of SNPs.

Figure III-5. Structure plots for Central Asian, Chinese, European and North-American apricots (from left to right) based on the unlinked SNPs dataset.

Figure III-6. Principal components of SNP variation obtained by using SmartPCA.

Figure III-7. SMC++-inferred demographic history of wild and cultivated apricots.

Figure III-8. Circos plot of the global distribution of SNP variants and signature of selective sweeps along the eight chromosomes of Central Asian wild apricots.

Figure III-9. Circos plot of the global distribution of SNP variants and signature of selective sweeps along the eight chromosomes of Chinese cultivated apricots.

Figure III-10. Circos plot of the global distribution of SNP variants and signature of selective sweeps along the eight chromosomes of European cultivated apricots.

Figure III-11. Summary of the shared signals between the wild Central Asian and the cultivated Chinese and European apricots.

Figure III-12. Distribution of the most significant selection signals along the eight apricot chromosomes.

Figure III-13. Distribution of significant signals of selection in cultivated apricot genomes.

Figure III-14. Schematic diagrams of four selective sweeps and their closest relevant candidate genes.

Figure III-S1. Scatter plots of read depth, mapping quality, Phred-scaled quality and SNPs variants on Chromosome 1.

Figure III-S2. Linkage disequilibrium decay in wild Central Asian apricot genomes on eight chromosomes.

Figure III-S3. Linkage disequilibrium decay in cultivated Chinese apricot genomes on eight chromosomes.

Figure III-S4. Linkage disequilibrium decay in cultivated European apricot genomes on eight chromosomes.

Figure III-S5. Structure plot for Central Asian, Chinese, European and North-American apricots (from left to right) based on the entire SNPs dataset.

Figure III-S6. Distribution of CLR statistics on eight chromosomes among the wild Central Asian, cultivated European and Chinese apricots.

Figure III-S7. Distribution of Omega statistics on eight chromosomes among the wild Central Asian and the cultivated European and Chinese apricots.

Figure III-S8. Distribution of μ statistics on eight chromosomes among the wild Central Asian, and the cultivated European and Chinese apricots.

List of Tables

GENERAL INTRODUCTION

Table 1 Recent studies on the identification of selective sweeps in long-lived fruit trees.

Chapter I:

Table I-1. Geographic information and sample size of cultivated and wild *Prunus armeniaca* and close relatives from the *Armeniaca* section used in this study.

Table I-2. Genetic diversity estimates for the nine apricot clusters ($n=416$ individuals, 34 SSR markers) inferred with STRUCTURE for $K=11$.

Table I-3. Pairwise F_{ST} (upper diagonal) and *Jost's D* (lower diagonal) among the nine apricot clusters ($n=416$, 34 SSR markers) inferred with STRUCTURE for $K=11$.

Table I-S1. Description of the *Prunus* cultivated and wild trees sampled for the purpose of this study, with their geographic origin and, when applicable, provider.

Table I-S2. List of microsatellite markers used in this study and their characteristics: chromosome on which is located the marker, given name of the locus, repeat motif and whether repeats were perfect, primer sequences, number of different alleles (N_a) and number of effective alleles (N_e).

Table I-S3. Summary the model choice analyses, and the corresponding results, that were carried out successively using random forest approximate Bayesian computation to reconstruct: 1) the divergence history of wild apricots, 2) the apricot domestication history, and 3) to test for the occurrence of bottleneck in cultivated groups.

Table I-S4. Description of each scenario and hypothesis tested with approximate Bayesian computation to reconstruct the wild divergence and domestication histories of apricots.

Table I-S5. Genetic variation within each of the four clusters inferred with STRUCTURE at $K=6$ for the wild apricots ($n=234$) based on 25 microsatellite markers.

Table I-S6. Prior distributions used for approximate Bayesian computations for inferring the wild and cultivated apricot divergence histories.

Table I-S7. Numbers of wild and cultivated apricots assigned to each, or admixed between, the clusters inferred with STRUCTURE at $K=11$ ($n=577$ individuals).

Table I-S8. Percentage and number of fully assigned (*i.e.* individuals that were assigned at $\geq 85\%$ to a given cluster), or admixed, wild and cultivated apricot individuals to each cluster defined with STRUCTURE for $2 \leq K \leq 12$.

Table I-S9. Summary of the Wilcoxon signed rank tests to investigate differences in genetic diversity between cultivated and wild genetic clusters presented in Table I-2. Significant value ($FDR_Q < 0.05$) are highlighted in light blue.

Table I-S10. Summary statistics of genetic diversity in the eco-geographical / genetic group of wild and cultivated apricots ($n=577$ individuals, 34 SSR markers).

Table I-S11. Summary of the Wilcoxon signed rank tests to investigate differences in genetic diversity in the eco-geographical groups of wild and cultivated apricots presented in Table S10 ($n=577$ individuals, 34 SSR markers).

Table I-S12. Inferences of recent effective population size reductions with the BOTTLENECK software for each of the three cultivated populations (PaEUR_C1, PaSCA_C2, and China_C4).

Table I-S13. Random forest approximate Bayesian computation (ABC-RF) parameter estimates for the two most likely models of domestication *scF_GF_5_GF* and *scF_GF_8_GF*.

Chapter II:

Table II-1. Different datasets included *P. brigantina* and other apricot species in this study.

Table II-2. Pairwise population *Jost's D* of *P. brigantina*.

Table II-S1a. Sampling locations and geographic regions of *P. brigantina* and *P. cerasifera* in French Alps Mountains.

Table II-S1b: List of individuals included in the different datasets of Table II-1 together with the *P. brigantina* samples.

Table II-S2. Analysis of genetic variability from microsatellites for *P. brigantina* population.

Table II-S3. The description of representatives from the core collection of *P. brigantina*.

Table II-S4. Genetic variability from microsatellites for *P. brigantina* core collection.

Table II-S5. Mann-Whitney U tests (two-tailed) between the whole dataset *P. brigantina* population and its core collection.

Annex 3:

Table SN-1. Homozygosity over the coding sequence of Marouch #14.

Table SN-2. Brief information on Marouch v2.0 *de novo* assembly.

Chapter III:

Table III-1. Three types of selective sweep detection methods.

Table III-2. Information and number of polymorphic loci identified by GATK and after quality filtering

Table III-3. Description of apricot linkage disequilibrium and their decays for each apricot chromosome.

Table III-4. Summary of overlap between selective sweeps and annotated candidate genes or quantitative trait loci.

Table III-5. Convergent candidate, selective sweep regions and genes shared between European cultivated and Chinese cultivated apricots.

Table III-S1. Information for each accession of the study. Accession number, accession registered name, origin and available genotyping data.

Table III-S2. Nucleotide diversity (Pi , π) values for wild and cultivated apricot groups.

Table III-S3. Tajima's D values for wild and cultivated apricot groups.

Table III-S4. Genetic difference F_{st} estimates for wild and cultivated apricot groups.

Table III-S5. Summary statistics of nucleotide diversity (Pi , π) test among the 8 chromosomes of wild and cultivated apricots.

Table III-S6. Summary statistics of Tajima's D test among the 8 chromosomes of wild and cultivated apricots.

Table III-S7. Summary statistics of CLR test among the 8 chromosomes of wild and cultivated apricots.

Table III-S8. Summary statistics of Omega test among the 8 chromosomes of wild and cultivated apricots.

Table III-S9. Summary statistics of Mu (μ) test among the 8 chromosomes of wild and cultivated apricots.

Table III-S10. Summary statistics of F_{st} test among the 8 chromosomes of wild and cultivated apricots.

Table III-S11. Summary statistics of hapFLK test among the 8 chromosomes of wild and cultivated apricots.

Table III-S12. Details of selective sweeps detected from multiple tests across the 8 apricot chromosomes (multiple sheets).

Table III-S13. List and position of the 1,753 selective sweeps identified over the 8 apricot chromosomes.

GENERAL INTRODUCTION

I- Species: Concepts and definitions

Although “species” is one of the most fundamental units of evolution and biology, the definition of the term is still controversial in evolutionary biology (Hausdorf and Bernhard, 2011). To determine the variation and the limitation between species, many concepts have been proposed as follows.

Biological species concept: a species is a group of actually or potentially interbreeding natural populations that are reproductively isolated from other such groups (Mayr, 1942; 1991). Thus groups of related plants which are distinct at the level of biological species do not interbreed when growing in the same area in nature.

Ecological species concept: in this case, a species refers to a set of organisms adapted to a particular niche, in a particular environment (Andersson, 1990; Van Valen, 1976).

Evolutionary species concept: following Wiley (1981) definition, an evolutionary species “is a single lineage of ancestor-descendant populations of organisms which maintains its identity from other such lineages [in space and time] and which has its own evolutionary tendencies and historical fate”.

Morphological species concept: defines species as the smallest groups that are constantly and determinedly distinctive and distinguishable by average means (Cronquist, 1978). This concept is close to the next one, the phenetic species concept.

Phenetic species concept: refers to the definition of species as a set of organisms that look similar to each other and distinct from other sets (Ridley, 1993).

Phylogenetic species concept: simply it defines species as a group of organisms that share an ancestor. In this concept, species is the smallest diagnosable cluster of individual organisms within which there is a parental pattern of ancestry and descent (Cracraft, 1983; 1989) or it is the smallest detected set of samples of self-perpetuating organisms that have unique sets of characters (Nelsen and Platnik, 1981).

Genotypic cluster concept: a species is a morphologically or genetically distinguishable group of individuals that have few or no intermediates when in contact with other such clusters (Mallet, 1995).

Besides the above species concepts, this ambiguous term of “species” also refers to species category and species taxon (Bock, 2004). Categories are different levels recognized in the Linnaean hierarchy of classification, like order, family, genus, species, etc... Species taxon, the real unit in nature, is almost always based on criteria in addition to those used as the defining criterion for the species concepts and even for the species category (Bock,

GENERAL INTRODUCTION

2004). We can remark here that the definition of a species is rather complicated. For understanding all species living at all times, a comprehensive concept larger than any species concept indicated above and using more than one species concept should be applied (the so-called **pluralistic species concept**) (Campbell and Reece, 2002). Indeed, since evolutionary factors such as gene flow or/and selection work in different degrees in different taxa to produce the kind of phonetic clusters we call species, some species may be more ecological, others more biological. This is the case for the species that will be studied in this PhD manuscript. We will, therefore, favour a pluralistic species concept, which recognizes that no single concept accounts for all species.

II- Origin of genetic diversity

Genetic diversity is an essential component of biodiversity since it allows to describe the genetic variation within and among populations or species. Genetic diversity plays an important role in evolution because it will serve as a way for populations to adapt to ever-changing environments. With more variation, it is more likely that some individuals in a population (or species) will possess variations of alleles (the so-called beneficial alleles) that will be better suited to the new environment. It thus represents the genetic makeup on which selection will act.

II-1 How to estimate and characterize genetic diversity?

Genetic diversity of a population can be assessed by the H_E parameter (expected heterozygosity, Nei, 1973), that corresponds to the level of genetic variability within a population. Often, we will compare the observed level of heterozygosity (H_O) to what we expect under Hardy-Weinberg equilibrium (HWE), H_E . It is the (expected) probability that an individual will be heterozygous at a given locus, calculated as follows:

$$H_e = 1 - \sum_{i=1}^n p_i^2$$

Where p_i is the frequency of the i^{th} of n alleles.

Note that while genetic diversity measures (through the detection of genetic polymorphism) are estimated over several loci that are presumed to be a random sample of the genome, heterozygosity is often averaged over multiple loci to obtain an estimate of genome-wide genetic diversity.

Genetic diversity is also characterized by its structuration. Indeed, genetic diversity among populations (or species) occurs if there are differences in allele and genotype

frequencies between those populations (or species). It can be measured using several different metrics that are all based on allele frequencies in populations. In our case, we will more commonly use the F -statistics (see glossary) that allow estimating the genetic differentiation between populations (F_{ST}), also called fixation index (Weir and Cockerham, 1984; Wright, 1978).

II-2 The Hardy-Weinberg equilibrium

Allele frequencies can easily be predicted following a simple theoretical model developed for an 'idealized' population as proposed by Hardy (1908) and Weinberg (1908). The British mathematician Godfrey H. Hardy and German physician Wilhelm Weinberg independently discovered the relationship between gene and genotype frequencies, known as the Hardy-Weinberg equilibrium (Hardy, 1908; Weinberg, 1908). Hardy-Weinberg Equilibrium (HWE), also referred to as the Hardy-Weinberg principle, is used to compare allele frequencies in a given population over a period of time. Definition of equilibrium in genetic systems became part of the well-known Hardy-Weinberg principle (Hartl and Clarke, 1997) which formally states that: "If a genetic population is such that (1) organisms are diploid, (2) reproduction is sexual, (3) generations do not overlap, (4) mating is random, (5) the size of the population is significantly large, (6) allele frequencies are equal in the sexes, and (7) there is no migration, mutation, or selection, then the genotype frequencies in the population are given by weighted products of the allele frequencies (Bosco et al., 2012). If all the above conditions are met, Hardy-Weinberg principle allows predicting genotypic frequencies from the allele frequencies. This also implies that one single generation is sufficient to reach Hardy-Weinberg genotype frequencies and that allelic and genotypic frequencies remain stable from one generation to another one.

The discovery of the Hardy-Weinberg principle marked the beginning of the field of population genetics (Chen, 2010). It is a fundamental principle in population genetics because its violation indicates that one of the five forces acting on allele frequencies (see below 'evolutionary processes that influence genetic diversity') operates in the population.

II-3 Evolutionary processes that influence genetic diversity

The Hardy-Weinberg conditions are the 'null hypothesis'. However, several processes can generate new variation in a population thus resulting in variation of allele frequencies and violation of HWE. They correspond to **mutation, gene flow (or migration), genetic drift, non-random mating** and **selection** (Figure 1) (see definitions in the glossary). All five of the above mechanisms may act to some extent in any natural population (Raven and

GENERAL INTRODUCTION

Johnson, 2002). As a result, those advantageous features are passed on at a higher frequency than less advantageous traits.

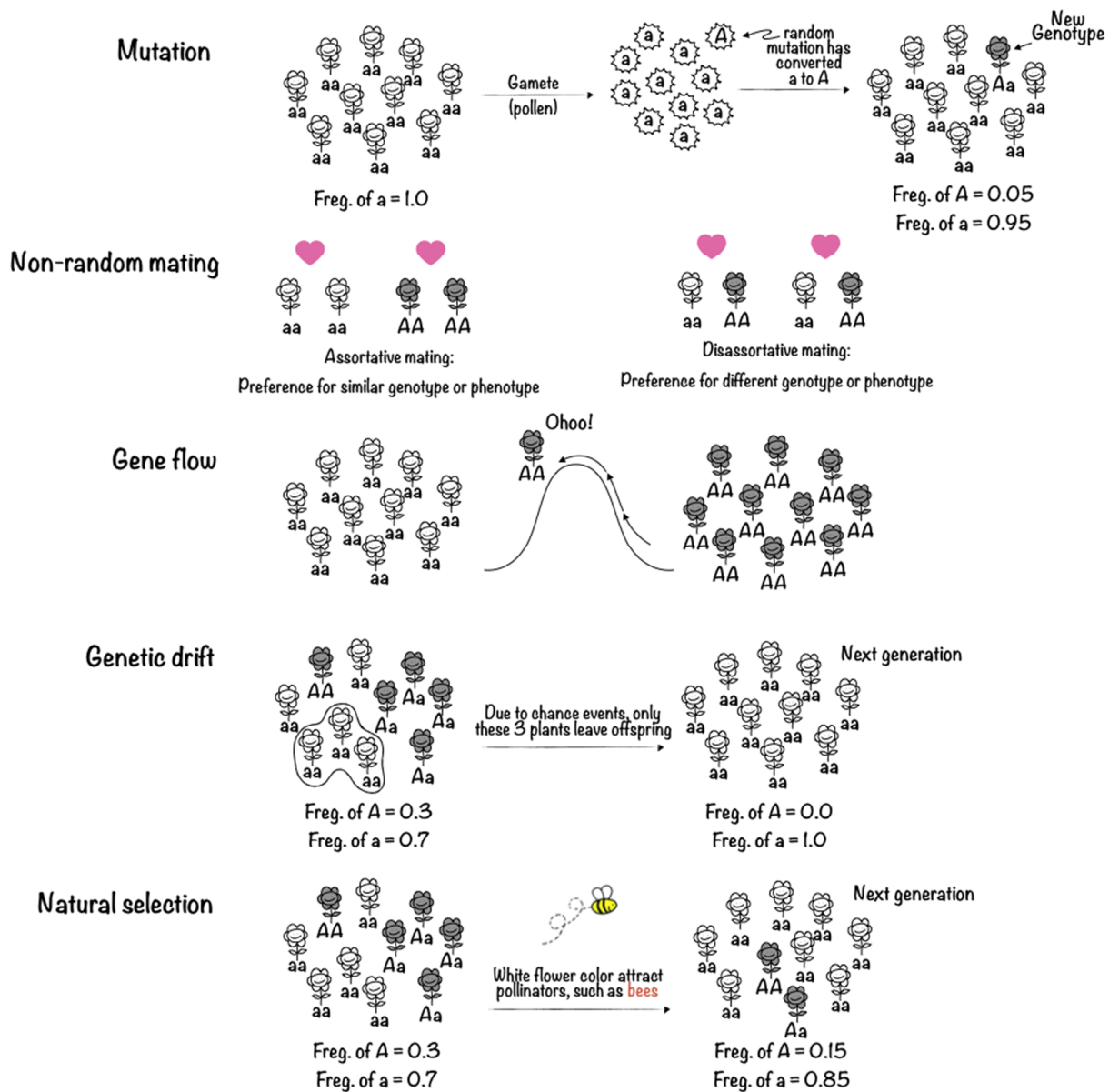


Figure 1. The analogy of evolution force.

II-3-1 The particular case of interspecific hybridization

Gene flow, often by migration, is the movement of genetic material. It can introduce novel alleles to a population. If these alleles are then integrated into the population, it will increase genetic diversity. One particular case of gene flow is the so-called inter-specific hybridization which results from gene flow between two different species. Such type of hybridization has been known at least since the time of Linnaeus and has been discussed frequently by evolutionists following Darwin's chapter 'Hybridism,' in which he demonstrated the lack of a clear boundary between varieties and species (Arnold, 1997; Darwin, 1858; Mallet, 2005).

Interspecific hybridization is a common and ongoing process in populations of land plants, with many important evolutionary consequences (Soltis and Soltis 2009). One of the challenges posed by interspecific hybridization is the difficulty of recognizing it, especially when closely related species with similar phenotypes are involved (Vit et al., 2014). Hybridization has frequently been considered as a race between fusion and speciation, with the outcome depending on the fitness of hybrids and the initial level of positive assortative mating (Harrison, 1990; 1993).

Studies of natural interspecific hybridization have addressed mainly five issues: 1) origin of hybridization (sympatric or allopatric, biogeographic pattern); 2) hybrid zone dynamics (maintenance of hybridization, stability of hybrid zones); 3) genetic and evolutionary consequences of hybridization (introgression, reticulate evolution, phylogeny); 4) speciation and genetic/reproductive isolation (species concept, assortative mating, mating barriers); and 5) fitness comparisons (habitat associations, ecological isolation) (Schwenk and Spaak, 1995).

II-3-2 Different evolutionary processes of selection

As stated by Charles Darwin in 1859 in his famous book « On the Origin of Species », **natural selection** is a key mechanism of evolution. Variation exists within all populations of organisms because of the occurrence of random mutations. From this genetic variation, natural selection will help to sort out individuals with certain variants of an adaptive trait that tend to provide a better chance of survival and reproduction, at least more than in individuals with other, less successful variants. In consequence, within-population variation allows natural selection to act upon traits that allow the population to adapt to changing environments.

Within a population, three types of natural selection are defined (Nielsen, 2005): (i) The **purifying (or negative) selection** which eliminates deleterious alleles. (ii) The **positive selection** is a process in which beneficial alleles are favoured and new advantageous genetic variants sweep a population. (iii) The **balancing selection** which maintains two or more alleles at a given locus (Figure 2).



B **Main molecular signatures of each selective regime**

Reduction of overall diversity Reduced functional variation Increase of rare variants	Reduction of overall diversity Increase of rare variants Increased LD and F_{st}	Increase of overall diversity Increase of common variants Reduced F_{st}
---	--	--

Figure 2. Modes of natural selection and molecular signatures of each selective regime. A. Purifying selection leads to the removal of deleterious alleles (in black) from the population. Positive selection favours the increase of a given allele (in red) in the population. Balancing selection can favour the presence of heterozygotes (in orange) in the population. B. The main molecular signatures of each selection type are described (LD: linkage disequilibrium). Blue circles indicate neutral mutation. Schematic diagram was modified from Quintana-Murci et al. (2013, 2016)

In contrast to natural selection, **artificial selection** is intentional because people (instead of nature) select which organisms get to reproduce, thus causing cultivated plant/animal populations to diverge morphologically and genetically from their wild progenitors (also named the ‘domestication syndrome’) (Zohary and Hoft, 2000). Domestication is a complex process along a continuum of human, plant, and animal relationships that often took place over a long-time period and was driven by a mix of ecological, biological, and human cultural factors (Harlan et al., 2012; Price and Bar-Yosef, 2011). Evolutionary biologists tend to view domestication more as species diversification, in opposition to species divergence that happens through natural selection (Darwin and Wallace, 1858). Evolution under domestication is unique in the general fields of plant evolution for three main reasons: 1) it is recent, having started not before 10,000 years ago

with the emergence of agriculture; 2) the original plant material, *i.e.* the wild progenitors of many important crop species still grow in their natural habitats; 3) human played a major role in this process. These factors enable a more reliable assessment of the impact of different evolutionary forces such as hybridization, migration, selection and drift under new circumstances such as cultivation practices, growth in a totally new environment sometimes thousands of kilometers away from its original habitat (Purugganan et al., 2009).

Of the more than 275,000 species of flowering species, less than 1% has been domesticated. Forty percent of the domesticated species belong to four families: Poaceae, Fabaceae, Rosaceae, and Solanaceae (Hilu, 1993; Li and Olsen, 2016). Different traits have been selected for in different taxa, but they commonly include increased size of edible parts (e.g., fruits or tubers), increased palatability, decreased armament, absence of dormancy, decreased toxicity, abbreviated flowering, synchronous phenology, non-shattering fruits (in Poaceae), thinner seed coats (Bennett, 2010).

III- The case-study of the domestication of long-lived, perennial crops

Our current understanding of evolution under domestication is based primarily on annual plants, often self-compatible species, which are propagated by seedlings each year. Nowadays, attention is refocusing on perennial crops, first, because the domestication of perennials is expected to follow different processes than annuals, second because they include food and oil staples that account for up to one-eighth of the world's total food-producing surface (McClure et al, 2014). Moreover, while annual crops represent a major proportion of our diet, several perennial crops are major players as well, being a key and reliable source of nutrition and a healthy diet all over the world (FAO, 2017; Gross et al., 2010). Compared with annual crops, relatively little is known about how perennial plant species evolve in response to human intervention (positive and negative selection). Understanding the evolutionary dynamics of long-lived tree species, with a special interest in domestication, is expected to emphasize potential differences or similarities between annual and perennial species but also to improve conservation strategies of perennial crops and their wild progenitors. It is also expected to facilitate breeding efforts for traits of interest in agronomy and crop improvement.

Here, we review the most recent studies that involved long-lived perennial crops and focus on the following questions: 1) how different are the evolutionary processes between annual plant and perennial plants; 2) which trait(s) is/are targeted during the domestication process of long-lived perennials? 3) How many historical events impacted the evolutionary

GENERAL INTRODUCTION

processes in long-lived perennials; 4) what is the extent to which domestication influenced genomic architecture in perennial crop species?

III-1 Different life traits between annual and perennial plants leading to different domestication processes

In plants, there are two broad categories of reproductive strategies, annuals that reproduce once every year and die and perennials that reproduce repeatedly and cycle through vegetative and reproductive phases from decades to hundreds of years (Friedman et al., 2015). While both annuals and perennials must undergo a developmental switch from vegetative growth to flowering, perennials will maintain their vegetative growth after flowering. This requires various and concomitant cell fate determination and meristem determinacy on one single plant so that some meristems either remain vegetative or switch to floral transition while others revert to vegetative growth (Amsino, 2009). This indeterminate state of the meristem allows, in perennial species, to multiply indefinitely the same individual. However, this developmental feature is not the only one that distinguishes perennials from annuals. Most scientists agree on the fact that the domestication process in perennial fruit crops departs from that observed in annuals because of major differences in their (i) mode of reproduction (clonal versus seed propagation, or sexual versus parthenocarpy), (ii) breeding systems (outcrossing versus selfing).

III-1-1 Mode(s) of reproduction

Domestication of annuals is considered as a straightforward process (Gepts, 2014). It begins with an enhanced human awareness of a wild species as a food source and then is further domesticated consciously or unconsciously, targeting agronomic traits that are of importance for genetic improvement and local adaptation of initially wild populations (Gaut, 2015). Contrasting dramatically with annuals, perennials experience long juvenile periods, most notably for olive (Lerma et al., 2014), apple (Cornille et al., 2014), peach (Gentile et al., 2002) and walnut (McCown and Brent, 2000). For example, an avocado tree (*Persea americana*) may take up to 15 years to mature before flowering (Berg and Lahav, 1996). This feature places severe limits on traditional use and breeding of perennial plants because farmers have to wait years to a decade before fruits can be evaluated and selected. For this reason, perennials' domestication has less frequently involved the selection and movement of seeds and their subsequent propagation. Instead, early farmers adopted clonal propagation, when possible, which is the primary form of reproduction in perennial fruit crops and therefore a key component of the domestication syndrome in those species (Zohary

and Spiegel-Roy, 1975). Most domesticated perennial trees are propagated primarily clonally, through cuttings, layering and grafting, three techniques that allow replicating indefinitely individuals depicting favourable traits (Miller and Gross, 2011). Since breeding of new perennial cultivars is time and space-consuming, thus expensive, this will also limit the number of elite cultivars that will be selected; the ones that are finally released will, therefore, remain for decades up to hundreds of years on the market (Peil et al., 2008). An expected consequence of both long juvenile periods and clonal propagation is the limited number of sexual cycles since the first selection of perennial, wild ancestors and thus potentially a limited loss of genetic diversity. Compared with annual crop species, most perennial crops appear to have been domesticated more recently or at least they show less divergence from their wild progenitors relative to annual plants (Meyer et al., 2012).

Beside vegetative propagation that can be human-mediated or naturally occurring (by suckering, in cherry trees *P. avium* for example; Stoeckel et al, 2006), asexual reproduction through parthenocarpy (banana, fig and pear, for example) and nucellar embryony (*Citrus* spp.) are two other modes of propagation that have been strongly selected during the domestication of some perennial crop species (Heslop-Harrison and Schwarzacher, 2007; Kislev et al. 2006; Rao et al. 2008; Zohary and Hopf, 2000). Interestingly, this was often accompanied by changes in reproductive biology, *i.e.* a switch from dioecy to hermaphroditism or from allogamous to autogamous as it happened during *Vitis* domestication (Zohary and Hopf, 2000).

III-1-2 Outcrossing mating system

Comparative analysis of the distribution of outcrossing rates among plants with different growth habit and longevity shows that the selfing rate of annual species tends to be higher than those of perennial herbs, and herbaceous perennials self more frequently than perennial woody plants (Barrett et al. 1996). Empirical studies of inbreeding depression also indicate that long-lived perennials, particularly forest trees, harbour substantial genetic loads explained, in part, by a higher outcrossing rate (Husband & Schemske 1996). Indeed, since most perennial crops are outcrossing, highly heterozygous and often propagated clonally, selection against deleterious mutations is less effective, which leads to an accumulation of recessive deleterious, mostly somatic, mutations. However, the manner in (and the extent to) which those variants might affect traits related to domestication is still unknown.

Another expected outcome of the obligate outcrossing is the role of interspecific gene flow into the origin and evolution of perennial crops. Besides increasing heterozygosity within individuals, strict allogamy also functions to increase variation within populations,

GENERAL INTRODUCTION

decrease differentiation among populations as individuals exchange genes with plants from nearby populations or wild, sympatric relatives. Therefore, interspecific hybridization will tend to increase the amount of variation on which natural and artificial selection can act (Hughes et al, 2007). And after domestication, recurrent interspecific gene flow accounts for the maintenance of genetic variation and further diversification (Cornille et al., 2012).

III-2 A loss of genetic diversity during perennials' domestication?

During crop evolution, the selective pressures are expected to result in a reduction of genetic diversity from wild progenitors to landraces. This loss of genetic variation is likely due to the selective propagation of some individuals (only a subset of the total number of individuals in a wild species is initially brought into cultivation) followed by many generations of selective breeding, thus leading to the “domestication bottleneck” (Charlesworth et al., 1997; Gepts and Papa, 2003). Have perennial crops also experienced a bottleneck (*i.e.* a reduction of genetic variation) during domestication? Comparative analyses demonstrated that perennial crops retain a greater proportion (averaged 94.8%) of the genetic variation present in their wild progenitors than annual crops (averaged 59.9%) (Miller and Gross, 2011; Figure 3). Recent genome-wide analyses of peach (*Prunus persica*) and its *Amygdalus* relative, almond (*Prunus dulcis*), showed no evidence of genetic bottlenecks associated with domestication in either species (Velasco et al., 2016)¹, and similar results have been found for grape (*Vitis vinifera*, Myles et al., 2011) and apple (*Malus domestica*, Cornille et al., 2014).

Note that comparisons between perennial and annual fruit crop domestication bottlenecks, level of heterozygosity and recombination are complicated by the fact that many outcrossing crop perennials encountered recurrent (and still-ongoing) hybridizations with their wild progenitors or related species, sometimes long after its domestication and dispersion (Cornille et al, 2012). It is indeed broadly acknowledged that the limited genetic bottleneck that accompanied the domestication of many perennial crops is likely the result of a combination of factors, including: a) few sexual cycles since domestication; b) multiple geographically and genetically distinct ancestral populations; c) past and present gene flow (including hybrid origin of cultivated species, human-mediated gene flow, and accidental

¹ In contrast with these results, two more recent studies (Yu et al. (2018) Li et al. (2019)) reported a significant reduction in genetic diversity between peach wild-related species (*P. mira*, *P. kansuensis*, *P. davidiana* etc...), landraces and modern cultivars, with only 34 to 25% of retained neutral diversity from their wild relatives, respectively. This data is indeed more in line with what is expected for the cultivated *P. persica* (peach) species, knowing its autogamous, reproductive biology. It also illustrates the importance of the samples used in such a study: 480 peach and peach-related accessions, including modern cultivars, landraces and wild relatives for Li et al (2019) and 44 for Yu et al (2018); 13 peach compared to 13 almond accessions, all cultivated, for Velasco et al (2016).

gene flow with sympatric native populations or related species. This could considerably bias the conclusion we draw, it also highlights the importance of studying cross-species gene flow and its impact on the domestication process of perennial crops.

Variation retained in domesticated populations

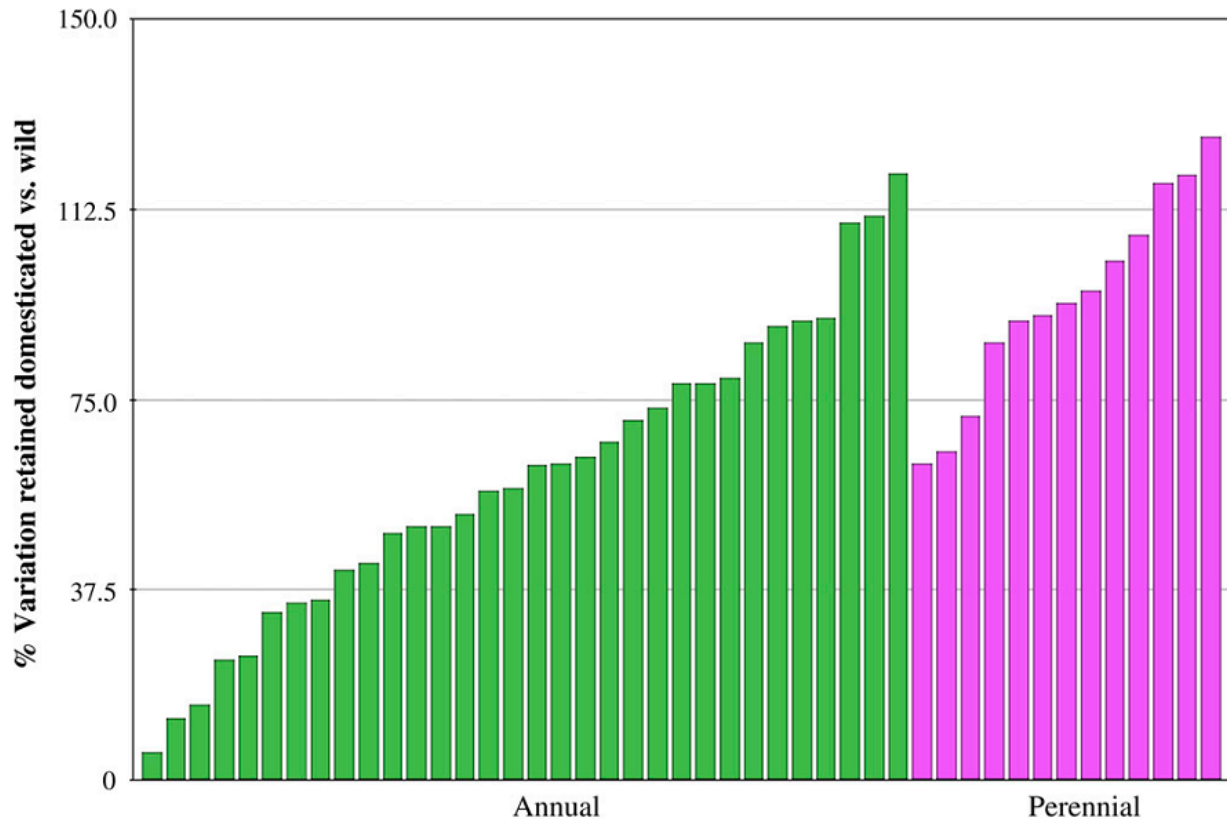


Figure 3. Percentage of retained genetic variation in domesticated annual (green) and perennial (purple) fruit crop populations compared to their wild relatives (from Miller and Gross, 2011).

III-3 Common traits targeted by domestication in long-lived perennial plants

Concerning plant domestication, it has long been known that picking and partly unconscious selection by the earliest farmers resulted in several traits found in crop species and not or rarely found in their wild progenitors. Collectively, these traits are called the “domestication syndrome” since the modification of those traits has resulted in a suite of morphological changes in cultivated stands relative to their wild progenitors (Miller and Gross, 2011). Those traits under target during domestication are usually related to the sustainable production of edible (and later palatable) plant products, which will also be storable over long periods of shortage (Fuller et al., 2015). More particularly in long-lived perennial plants, the domestication syndrome mainly consisted in acting on reproductive (self-compatibility, gynodioecy, andromonoecy or hermaphroditism, asexual reproduction, dormancy and control of flowering time) (in example, Saumitou-Laprade et al., 2017; Singh et al., 2019)

GENERAL INTRODUCTION

and vegetative traits (reduction of the juvenility phase, graft compatibility, cutting recovery, rootstock features) (Melnik and Meyerowitz, 2015; Miller and Gross, 2011; Warschefsky et al. 2016), in modifying fruit traits (fleshy fruit, weight, acidity, firmness, flavor and flesh/skin color) (Campory et al., 2015; Duan et al., 2017), in securing a better response to biotic (fungi, bacteria, insects, and weeds) (Abdala-Roberts et al., 2016; Vukicevich et al., 2016) and abiotic stresses (drought, salt and cold) (Ait Mouheb et al., 2018; Du et al., 2015; Godfrey et al., 2019). Table 1 summarizes recent studies that performed genome-wide scans to identify selection footprints within domesticated fruit perennials. However, this list might not be exhaustive. Since domestication appears to occur more slowly in perennials than in annuals because fewer sexual generations occurred since the first steps of selection and cultivation, we might underestimate those 'domestication traits', being still not so different from their wild progenitors. In some cases, it is difficult to distinguish between traits linked to local adaptation (under natural selection) and domestication (under artificial selection), especially when both the perennial crop and the wild progenitor continue to inhabit the same geographic region (sympatric species).

Table 1 Recent studies on the identification of selective sweeps in long-lived fruit trees.

Plant	Species	Number of samples	Selective sweep detection	Investigated traits	Reference	
Apple	<i>Malus domestica</i> <i>Malus sieversii</i> <i>Malus sylvestris</i> <i>Malus hupehensis</i>	58	<i>XP-CLR</i>	Acidity, color, firmness, hormone, soluble sugar, and secondary metabolites.	Duan et al. (2017)	
	<i>Malus domestica</i>	63	<i>CLR (sweed), Omega (OmegaPlus)</i>	153 candidate genes, related to photosynthesis, protein ubiquitination, metabolic pathways, biosynthesis of secondary metabolites, plant hormone signal transduction and in the purine metabolism.	Emanuela Kerschbamer. (2012)	
Chestnuts	<i>Castanea</i>	104	<i>Tajima's D, π</i>	Anthocyanidin in leaves and twigs, flowering-time regulatory, gravitropism, starch levels, desiccation tolerance, and phytohormone metabolism	LaBonte et al. (2018)	
Grape	<i>Vitis vinifera ssp. Vinifera</i>	472	<i>F_{st}, CLR, $\theta\pi$</i>	Regulating cell growth and metabolism, grape skin color, macromolecule metabolism, aromatic compound biosynthesis, and jasmonic acid metabolic pathways	Liang et al. (2019)	
	<i>Vinifera</i>	13	<i>CLR, XP-CLR</i>	Sugar accumulation during berry ripening, proanthocyanidin in accumulation, berry softening, and flowering-time	Zhou et al. (2017)	
	<i>Vitis vinifera L. subsp. Sativa</i>	137	<i>Tajima's D</i>	Grape colour (VvMybAs gene)	Fournier-Level et al. (2010)	
Date palms	<i>Phoenix dactylifera</i>	62	<i>Watterson estimator (θ_w)</i>	Immunity, fruit colour, sugar metabolism, fruit ripening	Hazzouri et al. (2015)	
Peach & almond	<i>Prunus persica</i> <i>Prunus davidiana</i> <i>Prunus ferganensis</i> <i>Prunus kansuensis</i> <i>Prunus mira</i>	84	<i>Tajima's D, reduction of diversity (ROD), F_{st}</i>	R (resistance) genes, genes related to flavonoid biosynthesis, flower development, photosynthesis, cell division, cell expansion, and carbohydrate metabolism.	Cao et al. (2014)	
	<i>Prunus persica</i> <i>Prunus mira</i> <i>Prunus davidiana</i> <i>Prunus kansuensis</i> <i>Prunus tangutica</i> <i>Prunus webbii</i>	95	<i>π, Tajima's D, iHS, XP-EHH</i>	Fruit size, maturity date, salt/osmotic tolerance, flowering/bud-burst timing biosynthesis of lutein, self-incompatibility	Akagi et al. (2016)	
	<i>Prunus persica</i>	418	<i>π, XP-EHH</i>	Fruit weight, sugars and acids, health-related compounds	Cao et al. (2019)	
	<i>Prunus mira</i> <i>Prunus davidiana</i> <i>Prunus kansuensis</i> <i>Prunus potaninii Batal</i>	480	<i>ROD, π ratio (i.e. $\pi_{Wild}/\pi_{Landraces}$), CLR</i>	Vegetative growth, delayed flowering time, low chilling requirement	Li et al. (2019)	
	Pear	<i>Pyrus pyrifolia</i>	41	<i>F_{st}, π ratio</i>	Plant cell division, auxin synthesis and efflux, lignin synthesis, photosynthesis, and stress resistance	Li et al. (2019)
		<i>Pyrus pyrifolia</i> <i>Pyrus bretschneideri</i> <i>Pyrus sinkiangensis</i>	113	<i>$\theta\pi$, Tajima's D, ROD, F_{st}</i>	Growth, response to cold, meristem and flower development, sugar accumulation, and few single-organism metabolic processes (stone cell formation and fruit size development)	Wu et al. (2018)
		Papaya	<i>Carica papaya L., Caricaceae</i>	48	<i>extended haplotype homozygosity (EHH)</i>	Red flesh color (CYC-b locu)
Jujube	<i>Ziziphus jujuba</i> Mill.	31	<i>F_{st}, XP-CLR, π ratio, Tajima's D</i>	Sweetness/acidity, self-incompatibility (S-locus)	Huang et al. (2016)	
Olive	<i>Olea europaea L. subsp. europaea var. europaea</i>	68	<i>ROD</i>		Muriel et al. (2019)	
		90	<i>Tajima's D, Fu and Li's D and Fu and Li's F</i>	Acyl carrier protein (ACP) genes. OeACP1 and OeACP2 gene encoding for the sucrose transporter 1 (SUT1), lupeol synthase (LUS)	Cultrera et al. (2018)	

Problématique de thèse

Cette partie a été rédigée en français selon les recommandations de l'école doctorale de Bordeaux. Elle présente le modèle d'étude et le contexte de la recherche, la démarche adoptée, les objectifs ainsi que les principaux résultats attendus. Cette partie a été préparée en étroite concertation avec l'étudiant Shuo Liu, il a lui-même proposé et réalisé les figures et l'ensemble a été traduit en français par la directrice de thèse.

Le(s) modèle(s) d'étude : L'abricotier et ses espèces apparentées

*L'espèce *Prunus armeniaca* (L.): Position systématique*

L'abricotier cultivé, que certains nomment 'commun', appartient au genre *Prunus*, de la sous-famille des *Prunoideae* qui comprend la totalité des arbres fruitiers à noyaux de la famille des *Rosaceae*. *Prunus* est un genre économiquement important avec approximativement 200 espèces, la plupart dans l'hémisphère Nord (Klakman, 1988). Parmi les espèces cultivées, citons le prunier domestique (*P. domestica* L.), le cerisier doux (*P. avium* L.), le cerisier acide (*P. cerasus* L.), le pêcher (*P. persica* L. Batsch), l'amandier (*P. dulcis* Mill. D. A. Webb.) et l'abricotier (*P. armeniaca* L.). Le cerisier en fleurs (*Pseudocerasus* Koehne, sous-genre *Cerasus*) et l'abricotier du Japon (*P. mume*) sont eux plus connus en tant que ornementaux, tandis que certaines espèces comme le cerisier noir (*P. serotina* Ehrh. ou merisier d'Amérique) sont recherchées pour la qualité de leur bois.

C. Bauhin (1560-1624) dans son *Kraüterbuch* publié en 1687 fut le premier à désigner l'abricotier comme une espèce à part entière. Tournefort (1700), botaniste de Louis XIV et inventeur de noms latins pour identifier les espèces botaniques, distingue l'abricotier du pêcher et lui donne le nom de « Armeniaca ». Enfin, Linnaeus (1737) intègre l'abricotier avec les cerisiers et pruniers dans le genre *Prunus*. De nos jours, la communauté scientifique s'accorde à classer l'abricotier dans la sous-famille des *Prunoideae*, dans le genre *Prunus* L., sous-genre *Prunophora* (Neck.) Focke, et section *Armeniaca* (Mill.) Koch (Rehder, 1940). Le sous-genre *Prunophora* (abricotiers et pruniers) est caractérisé par des jeunes feuilles enroulées dans le bourgeon (convolutées) ce qui est très facilement observable lors des stades précoces du débourrement végétatif (Bailey, 1927).

Parmi les *Prunophora*, la section *Armeniaca* (Lam.) Koch. se présente comme un complexe d'espèces diploïdes et inter-fertiles à $2n=16$ chromosomes, morphologiquement et écologiquement variées, réparties pour les trois-quarts en Asie du Nord-Est (Chine et Asie Centrale) et une seule espèce en Europe. Les barrières reproductives semblent d'ordres écologique et phénologique (Li et al., 2011). Le nombre d'espèces composant ce

taxon n'est pas connu avec exactitude du fait de cette interfertilité potentielle (parfois réalisée *in situ* et même entre taxons de sections différentes comme *P. dasycarpa*, par exemple) et des variations intra-spécifiques qui masquent parfois les limites entre espèces. Ainsi l'existence des espèces *P. holosericea* (ou *P. holosericea*, (Batalin) Kostina) et *P. ansu* (Maxim.) Kom est régulièrement remise en question, ces espèces ne pouvant être distinguées de l'abricotier commun selon des critères morphologiques. Elles présentent cependant des aires de distribution très spécifiques, Sud-Ouest (Tibet) et Sud-Est de la Chine respectivement. Elles sont plus couramment classées comme des sous-espèces ou des variants de *P. armeniaca*. Par souci de simplification, j'utiliserai, dans ce manuscrit, la classification proposée par Rehder, 1940 (Figure 4).

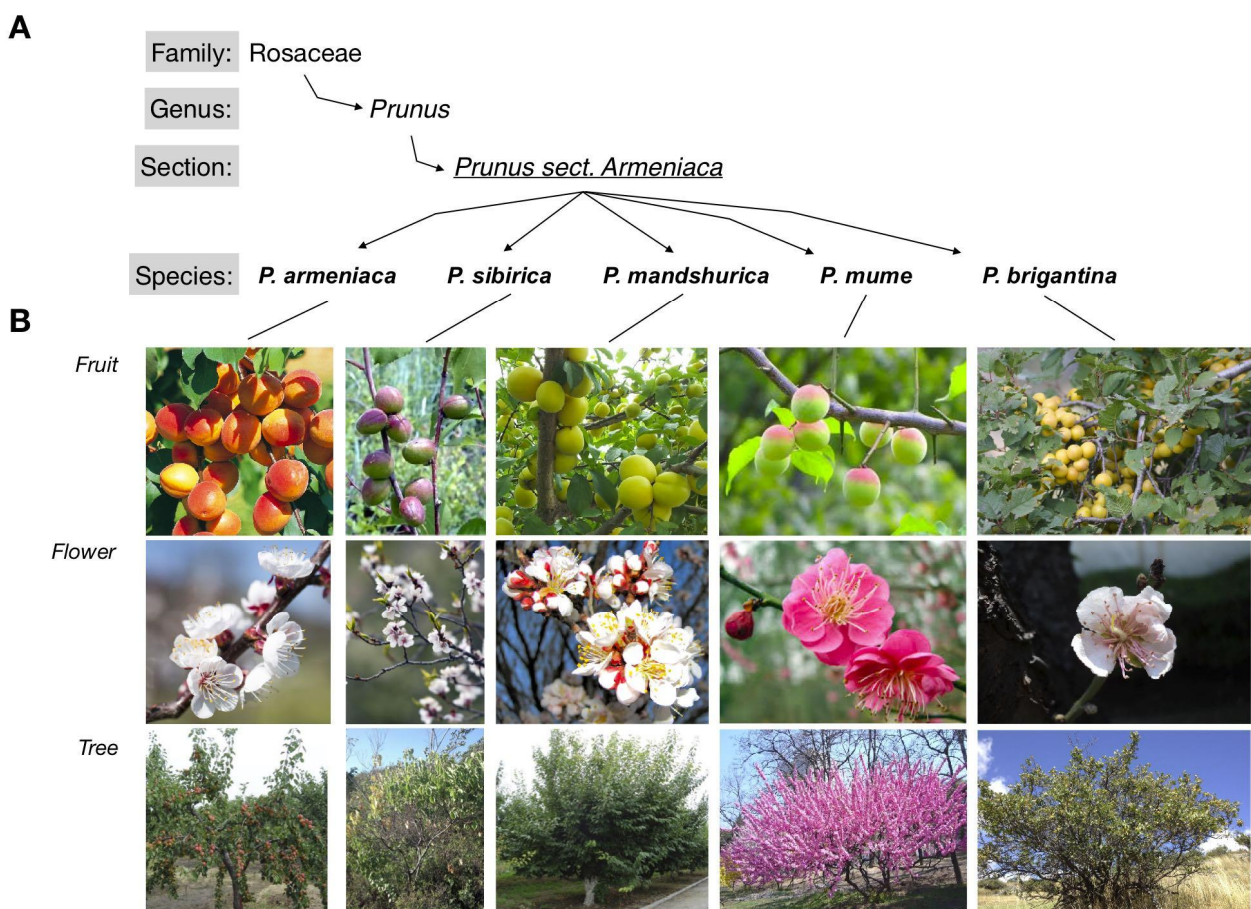


Figure 4. *Armeniaca* species and morphological features of fruit, flowers and tree.

Répartition et écologie des espèces *Armeniaca* non-domestiquées

L'espèce sauvage *P. armeniaca* (également appelée *P. armeniaca vulgaris* pour la distinguer de celle cultivée) est endémique en Asie Centrale, dans les mêmes forêts d'altitude (1500-2000m d'altitude) que l'ancêtre du pommier (*Malus sieversii*), du noyer (*Juglans regia*), du houblon et... du cannabis (Figure 5-A). La répartition actuelle de l'espèce résulte du morcellement de l'aire de distribution initiale. Les populations isolées et souvent

de tailles limitées se rencontrent principalement dans les massifs montagneux du Tian Shan (Tien Shan) et du Pamir, c'est-à-dire du Nord au Sud, du Kazakhstan au Pakistan, et d'Ouest en Est, du Turkménistan à la province du Xinjiang, en Chine (extrême Ouest de la Chine) (Kostina, 1964; Zaurov et al., 2013).

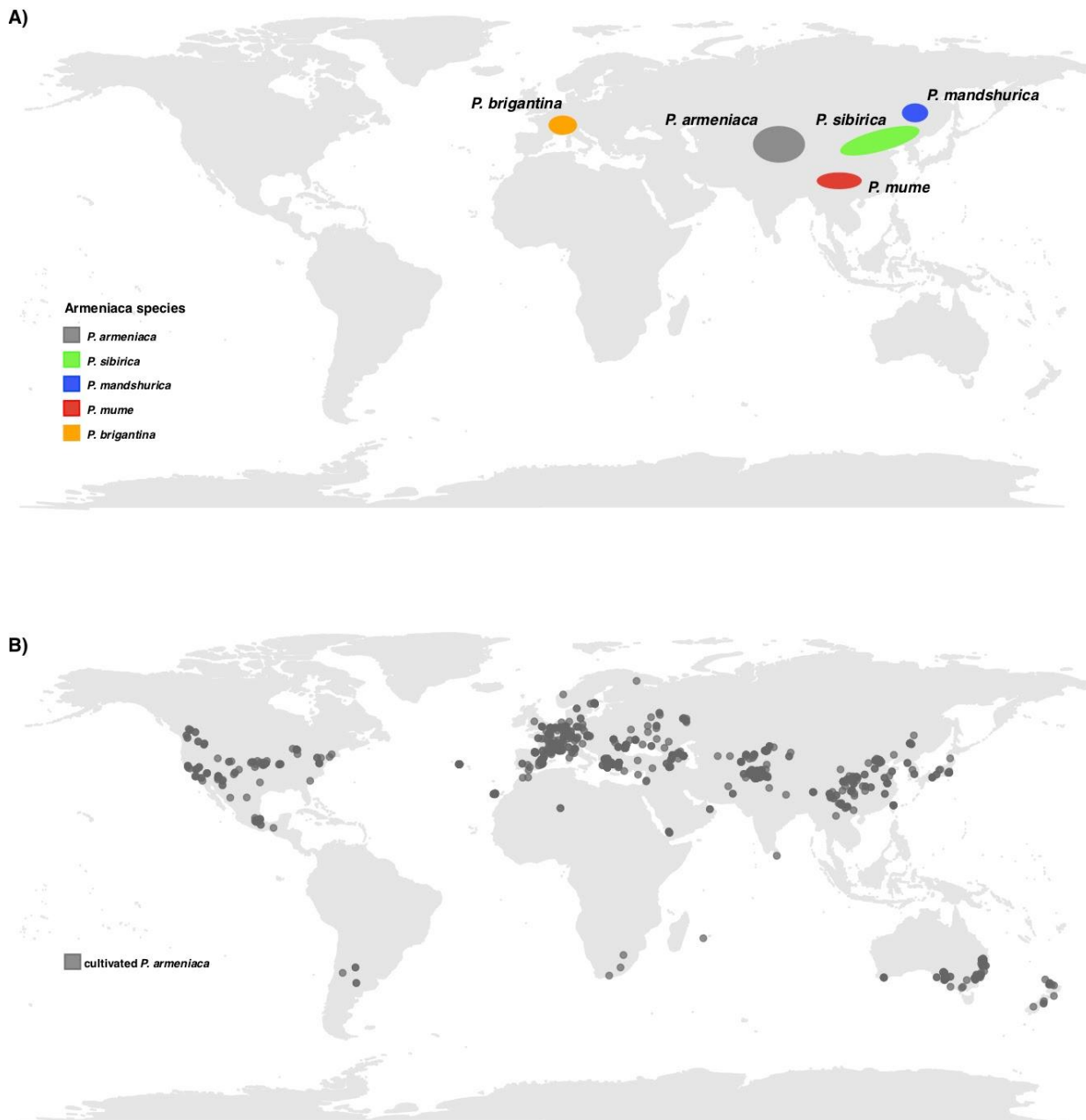


Figure 5. *Armeniaca* species worldwide distribution. A) Schematic geographic regions of wild species distribution. B) Distribution of apricot cultivation, all over the world. Data was retrieved from GBIF (Global Biodiversity Information Facility) database (<https://www.gbif.org/>).

De l'autre côté de l'Himalaya, trois espèces distinctes sont endémiques et peuvent encore être trouvées à l'état sauvage: *P. mume* (Siebold) Siebold & Zucc. (sur les contreforts Sud du Tibet), *P. sibirica* L. (au Nord-Ouest et Est de la Chine, jusqu'en Mongolie et Sibérie) et *P. mandshurica* (Maxim.) Koehne (à l'extrême Nord-Est de la Chine, jusqu'en

Corée et Russie orientale) (Figure 5-A) (Wang et al., 2014; Zhang et al., 2018). Ces populations naturelles (celles de *P. armeniaca* incluses), dont le maintien reste fortement tributaire de la dynamique des peuplements forestiers qui les hébergent, sont devenues rares. Elles ont subi au cours des siècles de nombreux dommages souvent dus à l'action des hommes: incendies, déforestation, extensification de l'agriculture et des villes... Cependant, il faut remarquer qu'en Asie Centrale et de l'Est, l'abricotier cultivé partage sa zone de répartition avec les populations naturelles, la seule barrière étant souvent l'altitude (les populations naturelles dans les massifs montagneux comme le Tian Shan, le Pamir et le Tibet et les abricotiers cultivés dans les vallées) (Figure 5-B). Tandis que les peuplements naturels de *P. mandshurica* et de *P. mume* sont devenus quasi inexistantes, *P. sibirica* conserve une aire de distribution assez large, du Nord au Sud et d'Est en Ouest (Wang et al., 2014). Les populations naturelles de *P. armeniaca* sont également très rares en Chine et n'existent que dans la province du Xinjiang, à l'extrême Nord-Ouest, à la frontière avec le Kazakhstan (Vallée Illy) et un peu plus au Sud, à la frontière avec le Kirghizistan et le Pakistan (Figure 5-A) (Maynard, 1999 ; Wang et al., 2017). Cette structuration géographique (et écologique) des espèces *Armeniaca* en Asie interroge sur l'histoire évolutive de ces espèces, sur l'origine de leur distribution, de leur adaptation et des relations réciproques existant entre espèces sauvages d'une part et entre espèces sauvages et cultivées, d'autre part. Ceci sera abordé dans le chapitre I de ce manuscrit de thèse, au travers d'un échantillonnage plus conséquent des espèces asiatiques en comparaison avec l'article publié en 2016 (Decroocq et al, 2016).

Enfin, au sein de la section *Armeniaca*, *P. brigantina* Vill. est la seule espèce native du continent européen (Bailey and Hough, 1975), ce qui soulève nombre de questions, d'autant plus qu'elle se distingue du reste des espèces par l'aspect glabre de ses ovaires. La question de l'origine de *P. brigantina* ainsi que de sa classification dans la section *Armeniaca* sera abordée dans le chapitre II de ce manuscrit de thèse.

Origine de l'abricotier cultivé

D'un point de vue phylogénétique, l'origine de l'abricotier commun serait l'Asie. Selon le botaniste russe Vavilov (1951) (Vavilov, 1951), il y aurait trois centres d'origine pour l'abricotier: (a) la Chine et le Tibet, (b) l'Asie Centrale (du Tien-Shan au Kashmir), (c) le Proche-Orient (Iran, Caucase, Turquie). Alors que la plupart des écrits, occidentaux et chinois, s'accordait très largement sur une origine chinoise, des travaux récents réalisés par l'équipe de Bordeaux sur la base de marqueurs nucléaires ont montré que *P. armeniaca vulgaris*, l'abricotier sauvage d'Asie Centrale, est l'ancêtre de l'abricotier cultivé, tout au

moins celui cultivé en Europe de l'Ouest et de l'Est, sur le pourtour méditerranéen et dans l'aire géographique irano-caucasienne (Decroocq et al, 2016). En effet, Bourguiba et al (2012) ont montré que tous les abricotiers cultivés européens et nord-africains ont une origine commune, irano-caucasienne (d'où son nom de *P. armeniaca*, 'celui qui vient de la Grande Arménie'), avec deux routes d'introduction : Une au Nord, par les balkans et l'autre au Sud, par le Proche Orient, l'Afrique du Nord et enfin le Sud de l'Espagne (Murcia notamment). Ainsi, après sa domestication en Asie Centrale, l'abricotier a suivi une voie de dispersion vers l'Ouest au travers de la Perse puis au travers du Caucase et de l'Anatolie de l'Est d'un côté et du Proche Orient et du Maghreb de l'autre, vers l'Europe. Une alternative serait que l'abricotier européen ait connu un processus de domestication lent et continu (accompagné de diversification), à partir de son centre d'origine, l'Asie Centrale, vers le continent Européen. Plus récemment, il a poursuivi sa dispersion vers l'Ouest et l'hémisphère Sud, c'est-à-dire vers le continent américain et les nouvelles colonies britanniques et espagnoles (Figure 6).

Malgré un nombre limité de loci étudiés (les mêmes que dans l'étude de Bourguiba et al, 2012), Decroocq et al (2016) avait également émis l'hypothèse d'au moins deux événements distincts de domestication de l'abricotier cultivé : un ayant donné lieu à l'abricotier chinois et le second à l'origine de l'abricotier européen. Cette hypothèse sera également testée et vérifiée dans le chapitre I de ce manuscrit, grâce un nombre plus conséquent de marqueurs et un échantillonnage plus représentatif des espèces de la section *Armeniaca* présentes en Chine (Figure 5-A).

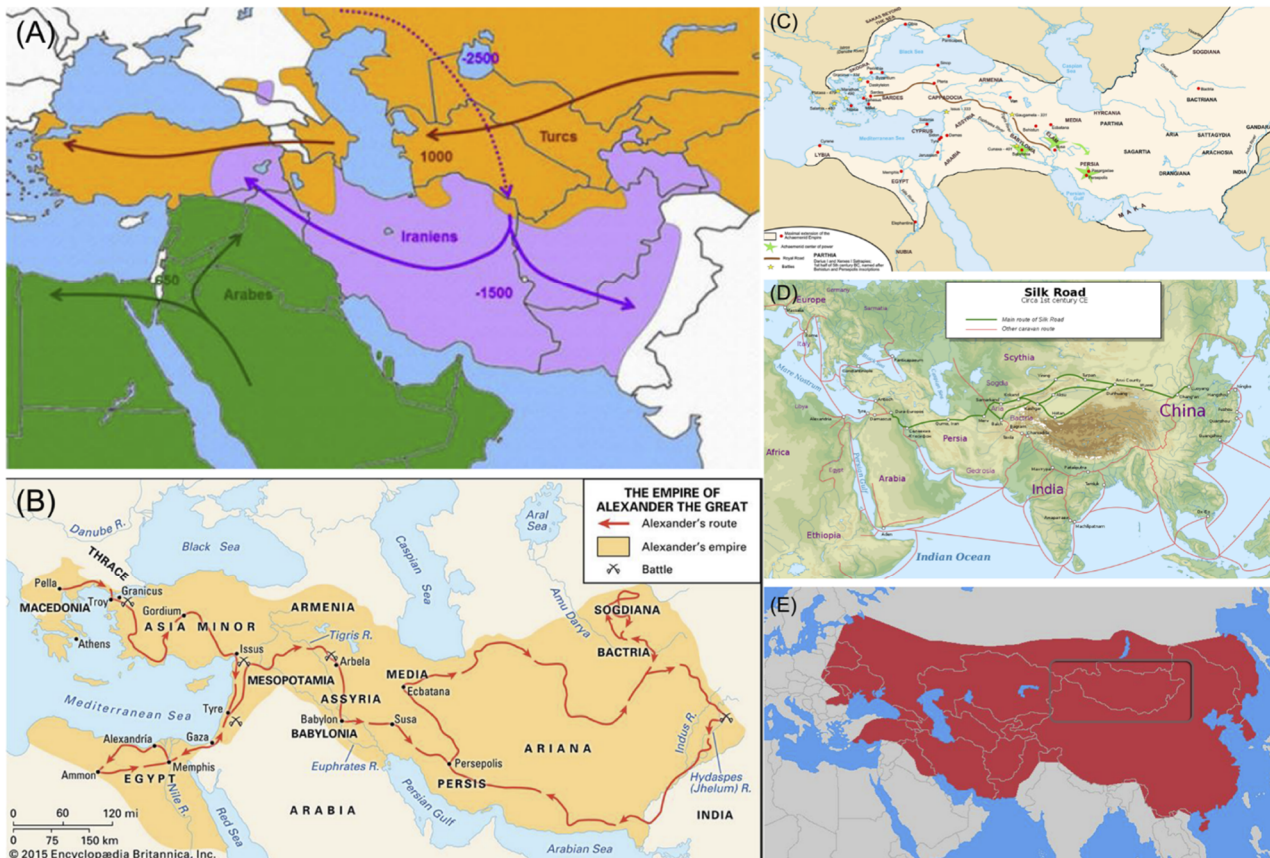


Figure 6. Map with the diverse routes of apricot dispersion, including EUR, North-America, South-America, South-Africa and Australia. In pre-islamic and -christian times, Central Asia was predominantly Iranian (Persian), populated by Eastern Iranian-speaking tribes such as the Sogdians (the region being sometimes referred as Sogdiana and local wild plum as *P. sogdiana*) who came initially from the North-Eastern territories of Caspian sea (Scythia region, 2500BC) and spread all over Central Asia and Iran (2000-1500BC) (Figure 6-A). The region would eventually and temporary be annexed by the Macedonian ruler Alexander the Great in 328 BC (Figure 6-B, Frank (2019)). This would mark the first supposed introduction of apricot on the European continent, through Greece. Apricots being cultivated in Persia since antiquity, it could also have been introduced through the earlier Silk Road traders. The Silk Road is referred as a network of trade routes connecting China with the Middle East and Europe (Figure 6-D). It opened up with the Han dynasty (130 BC) until 1453 when the Ottoman Empire boycotted trade with China (About the Silk Road: UNESCO, <https://en.unesco.org>). However the transport of goods and services along these routes are supposed to have started much earlier, through the Royal road (Figure 6-C) (Graf, 1994). This route connected Susa (in present-day Iran) more than 2,500 Km west to Sardis (Turkey) and was established by the Persian ruler Darius I, 300 years before the opening of the Silk Road (Figure 6-C and D). Trades through the Royal and the Silk roads were concomitant with the migration of Turkic people, coming from the actual Xinjiang and part of the Mongol area (Figure 6-A). This occurred mostly between the 5th and 10th centuries AC, when they spread across Central Asia, Near-East and up to Turkey ((A) map of human migration). Therefore, expansion of the Ottoman Empire westwards to Eastern Europe, together with the Mongol invasions (1206-1279, Figure 6-E), would indicate the latest routes of apricot introduction from Central Asia to the European continent, through Armenia and Eastern Anatolia in the North and the Fertile Crescent and North-Africa, in the South. Acknowl for map 6-E: Astrokey44 [CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=885439>]. Acknowl for map 6-D: Kaidor [CC BY-SA 4.0 (<https://creativecommons.org/licenses/by-sa/4.0>)].

Objectifs de la thèse

L'objectif de cette thèse est de caractériser la diversité et la variabilité génétique chez l'abricotier afin de décrire les processus évolutifs qui en sont à l'origine. En effet, la diversité des espèces cultivées résulte d'une série d'évènements de domestication, de flux de gènes entre compartiments sauvages et cultivés, d'effets de la sélection adaptative naturelle et aussi de la sélection humaine et des dynamiques de diffusion à de larges échelles, souvent sur de longues périodes. Dans le cas de l'abricotier, l'origine multiple et complexe de cette espèce ainsi que son aire de répartition en partie partagée avec les espèces sauvages apparentées m'a obligé à mener cette étude à la fois au niveau intraspécifique mais également interspécifique. Dans les deux premiers chapitres, nous nous focaliserons donc sur la section taxonomique *Armeniaca* qui constitue un bon modèle puisque des ressources, à la fois cultivées et sauvages, sont disponibles. En effet, si l'étude de la diversité génétique intra-spécifique s'avère utile pour la compréhension de la capacité d'adaptation d'une espèce aux variations de son environnement et donc de sa pérennité dans le temps, l'étude de la diversité génétique interspécifique permet, elle, d'étudier des événements évolutifs plus anciens à l'échelle des espèces ou même de la section entière.

Pour ces études, présentées dans la première partie de ce manuscrit de thèse (Chapitre I), trente-quatre marqueurs nucléaires ont été utilisés chez plus de 500 individus de la section *Armeniaca* et combinés à des approches Bayésiennes afin de reconstruire l'histoire évolutive de ces espèces à partir des niveaux de diversité et de différenciation génétique interspécifique. L'une des forces de cette étude est de porter sur un grand échantillonnage, caractérisé génétiquement par un grand nombre de marqueurs SSR. La seconde de nos forces fut la collaboration étroite avec Amandine Cornille (CNRS, GQE Le Moulon) et Tatiana Giraud (CNRS, ESE) ce qui m'a permis de réaliser un stage de 4 mois à l'Université de Paris-Saclay (GQE le Moulon) avec à la clef des analyses statistiques bayésiennes robustes.

La même approche et les mêmes marqueurs ont été utilisés dans le chapitre II afin de décrire la diversité et la structuration génétique de *P. brigantina* à partir d'un échantillonnage plus vaste des populations alpines françaises et ainsi mettre en évidence la relation entre cette espèce européenne et le reste de la section *Armeniaca*, Ceci m'a permis également de préciser la classification de *P. brigantina* dans le genre *Prunus* et de proposer une core-collection représentative de la diversité de l'espèce afin d'en assurer sa conservation *ex-situ*.

Dans le Chapitre III, sur la base des informations obtenues dans le Chapitre I sur l'abricotier sauvage et cultivé (*P. armeniaca*), en excluant cette fois les espèces

apparentées, je me suis posé la question de l'effet de la sélection (naturelle et artificielle) sur l'architecture génomique de l'abricotier. Pour cela, j'ai pu bénéficier d'un assemblage *de novo* du génome abricotier réalisé par le laboratoire d'accueil à Bordeaux en collaboration avec le Centre de Bioinformatique de l'Université de Bordeaux (CBIB). Grâce à des financements successifs de l'ANR (ABRIWG), de l'Université de Bordeaux (G2P SWAGMAN et ATT ABXING) et de France Génomique (SWAG), le laboratoire d'accueil a également eu la possibilité de faire séquencer en fragments courts de 150 bp (ILLUMINA, 15X de profondeur de lecture) l'entièreté de la collection botanique *Armeniaca* de l'INRA de Bordeaux ainsi que la collection de variétés patrimoniales cultivées de l'INRA d'Avignon (UGAFL). Plus de 600 génomes *Armeniaca* sont maintenant disponibles mais je me focalise ici sur les génomes *P. armeniaca*, cultivés et sauvages. L'ensemble de ces données génomiques m'a permis dans le Chapitre III de m'interroger sur « *quelle réduction de diversité lors de la domestication de l'abricotier et pour quel(s) gène(s), quel(s) caractère(s)?* ». D'après les données publiées par Bourguiba et al (2012) et confirmées en partie par Decroocq et al (2016), il semblerait que la réduction de diversité liée à la domestication (c'est-à-dire le goulot d'étranglement génétique) ait été relativement faible, en tout cas plus faible chez les abricotiers Chinois que chez les Européens. Dans le Chapitre I, j'ai pu tester cette hypothèse mais également l'existence de multiples événements de domestication, indépendants, dont les plus anciens ont donné lieu à l'abricotier Chinois puis à l'abricotier Européen. Ceci m'a conduit à tester l'hypothèse dans le Chapitre III de l'existence de signatures convergentes ou divergentes de domestication au sein des génomes abricotiers Chinois et Européens et d'explicitier les spécificités de chacun.

Enfin dans une conclusion générale, les principaux résultats seront rappelés et mis en relation pour qu'émerge une vision globale des forces pouvant générer les niveaux et les profils de diversité génétique chez les abricotiers, sauvages et cultivés, et dans une moindre mesure, chez les espèces de la section *Armeniaca*.

En résumé, ce travail de thèse vise à mieux comprendre les différents processus de l'histoire évolutive d'une espèce fruitière pérenne et comment ceux-ci influent sur la variabilité et la structuration génétique de l'espèce cultivée. Ceci inclut son adaptation à de multiples et changeantes conditions environnementales mais également à l'action de l'Homme au travers de la domestication, de la sélection et de l'amélioration génétique.

PhD context and objectives

This part summarizes, in English, the above section named 'Problématique de la thèse'.

Apricot, a long-lived perennial tree

Apricot temperate fruit tree (*Prunus armeniaca* L.) belongs to the family *Rosaceae*, subfamily *Prunoideae*, genus *Prunus* and section *Armeniaca* (Rehder, 1940) (Figure 3-A, above). The species *P. armeniaca* refers to both the wild progenitor and the cultivated species (also called 'common apricot'). It is a deciduous tree grown for its edible fruits with an annual worldwide production of ~ 4.2 million tons (FAO, 2017), mostly cultivated in the Mediterranean region (Turkey as the first producer, mainly of dried apricots), the Middle East, in Armenia, India, Pakistan and China. Natural populations of *P. armeniaca* are still available, solely in Central Asia (Hormaza et al., 2007; Lingdi and Bartholomew 2003). Other related species belonging to the section *Armeniaca* are: *P. sibirica* L., *P. mandshurica* (Maxim.) Koehne, *P. mume* (Siebold) Siebold & Zucc. and *P. holosericea* all endemic in Eastern Asia (mostly China), *P. brigantina* Vill. in the French and Italian Alps (Bailey, 1916; Rehder, 1940). The existence of *P. holosericea*, together with *P. ansu*, is still questioned (Faust et al., 1998). All these apricot species are diploid ($2n = 16$) and have a genome size of approximately 220-230 Mbp. Among those species of the section *Armeniaca*, only *P. armeniaca* and *P. mume* were domesticated, the first one for its fruits, the second one for its flowers mostly (and secondly, for its fruits consumed as preserved).

P. sibirica, also called Siberian apricot, is distributed widely across the mountainous areas of north and northeast China, eastern Siberia and Mongolia (Maynard, 1999) (Figure 4-A above). Siberian apricot trees are able to thrive under many types of harsh environmental conditions such as low temperature, strong wind, low rainfall and poor soil. All along with their long-term evolution, the wild Siberian apricot populations generated a large number of variations (Wang et al., 2017). Another wild species endemic in nearby northeast China is *P. mandshurica*, also called Manchurian apricot. It occupies a small area in Eastern Manchuria and is sometimes considered a subspecies of the common apricot because of the strength and size of its trunk (Figure 4-B). Trees of *P. mandshurica* are very cold hardy, tolerating temperatures of -40 to -45 °C (Mehlenbacher et al., 1991). *P. mume*, also known as the flowering apricot, is most famous for its ornamental flowers and its fruits used in preserves and liquors in East Asian countries. This species is still present as both domesticated and wild forms (Zhang et al., 2018). The distribution of the wild progenitor is centred around the borders of northwestern Yunnan Province, southwestern Sichuan Province, and southeastern Tibet Autonomous Region. Wild *P. mume* can be found across

a wide region south of Changjiang River (from southwest to southeast China) (Zhang et al., 2010). Flowers and young fruits of *P. mume* can tolerate low temperatures (-4 to -2 °C) in early spring (Zhang et al., 2012). In contrast with the other *Armeniaca* species, *P. brigantina* is native from Europe (Villars, 1789) (Figure 4-B and 5-A). Representatives of this species are still growing wild within the Alpen Mountains between France and Italy in southern Europe, where it is regarded either as apricot or a plum species (Hagen et al, 2002; Pignatti, 1982).

The focus of the PhD thesis

In this thesis, we focus on apricot species, *P. armeniaca*, and its related species from section *Armeniaca*, to address a few important questions related to fruit tree origin, evolution and domestication. The main goal is to characterize apricot genetic diversity and variability to describe processes of its evolutionary history under both natural and artificial selections. Our results will be presented in three chapters as follows: I) Investigation on the complex evolutionary history of apricots: species divergence, gene flow and multiple domestication events, II) Genetic diversity and structuration analysis in the French Alpen apricot (*Prunus brigantina*) reveal phylogenetic inconsistencies in the *Armeniaca* section, and III) Distinct evolution and domestication of apricot (*P. armeniaca*) revealed by genomic signatures under selection.

In Chapter I, the objective was to study the evolutionary history of apricot with an emphasis on the cultivated and wild *Prunus* apricots endemic to China. In this chapter, based on microsatellite genotyping data, we addressed the questions of the genetic diversity and structure of wild and cultivated apricots in Central/East Asia, of the divergence history of wild lineages across this region and the number of domestication events that resulted from those lineages. We also tried to elucidate the contribution of each wild species to the currently cultivated apricot gene pool and the extent of interspecific gene flow and bottlenecks during apricot evolutionary history.

In Chapter II, the objective was to evaluate the extent of *P. brigantina* diversity and differentiation. Through this molecular characterization, we investigated the relationship between *P. brigantina* and other *Armeniaca* species. Finally, we identified a collection of unique genotypes and selected the best subset that allows to define a *P. brigantina* core collection, with maximum maintenance of allelic diversity, which will be further characterized and genotyped for stone fruit crop improvement.

In Chapter III, we conducted a population-level analysis of genetic variation of apricots based on the resequencing of genomes of three genetic clusters of wild and

cultivated apricots. We focused here on how selection has influenced genomic architecture in apricot. To test for common or distinct signatures of selection, we took advantage of the parallel history of domestication in the European and Chinese apricots and compared with their wild, Central Asian progenitor.

The complex evolutionary history of apricots: species divergence, gene flow and multiple domestication events

Published in Molecular Ecology, doi:10.1111/mec.15296

The complex evolutionary history of apricots: species divergence, gene flow and multiple domestication events

Liu Shuo^{1,2*}, Cornille Amandine^{3*}, Decroocq Stéphane¹, Tricon David¹, Chague Aurélie¹, Eyquard Jean-Philippe¹, Liu Wei-Sheng², Giraud Tatiana^{4‡}, Decroocq Véronique^{1 ‡}

¹ UMR BFP, INRA-Université de Bordeaux, Villenave d'Ornon, France

² Liaoning Institute of Pomology, Tiedong Street, Xiongyue Town, Bayuquan District, Yingkou City, Liaoning, 115009, China

³ GQE– Le Moulon, INRA, Univ. Paris-Sud, CNRS, AgroParisTech, Université Paris-Saclay, 91190, Gif-sur-Yvette, France

⁴ Ecologie Systematique Evolution, Univ. Paris-Sud, CNRS, AgroParisTech, Université Paris-Saclay, Orsay, France

* Co-first authors: Those authors have contributed equally to this work.

‡ Corresponding authors: veronique.decroocq@inra.fr, tatiana.giraud@u-psud.fr

Abstract

Domestication is an excellent model to study diversification and this evolutionary process can be different in perennial plants such as fruit trees compared to annual crops. Here, we inferred the history of wild apricot species divergence and apricot domestication history across Eurasia, with a special focus on Central and Eastern Asia, based on microsatellite markers and approximate Bayesian computation. We significantly extended our previous sampling of apricots in Europe and Central Asia towards Eastern Asia, resulting in a total sample of 271 cultivated samples and 306 wild apricots across Eurasia, mainly *Prunus armeniaca* and *P. sibirica*, with also some *P. mume* and *P. mandshurica*. We recovered wild Chinese species as genetically differentiated clusters, with *P. sibirica* being divided into two clusters, one possibly resulting from hybridization with *P. armeniaca*. Central Asia also appeared as a diversification centre of wild apricots. We further revealed at least three domestication events, without bottlenecks, that gave rise to European, Southern Central Asian and Chinese cultivated apricots, with ancient gene flow among them. The domestication event in China possibly resulted from ancient hybridization between wild populations from Central and Eastern Asia. We also detected extensive footprints of recent admixture in all groups of cultivated apricots. Our results thus show that apricot is an excellent model for studying speciation and domestication in long-lived perennial fruit trees.

Keywords: ABC-RF (random forest), model testing, gene flow, domestication, admixture, introgression.

Introduction

Domestication of living organisms is one of the most significant cultural and evolutionary transitions over the past 12,000 years (Gross and Olsen, 2010; Larson et al., 2014; Meyer and Purugganan, 2013). The original traits of the wild ancestors of our modern crop species have been considerably modified during the protracted processes of plant domestication and subsequent variety breeding (Purugganan et al. 2009). Not all desired traits could, however, be selected for all varieties (Collard et al., 2008; Nagaraju et al., 2002). In addition, bottlenecks often occurred in domesticated lineages, reducing genetic diversity (Voss-Fels et al., 2019). As a consequence, many important traits, such as resistance to major crop pest or diseases, may be lacking in the cultivated germplasm while present in wild relatives. This is the case for the resistance to sharka in apricot, to mildew in grape or to apple scab in apple (Belfanti et al., 2004; Decroocq et al., 2016; Riaz et al., 2013). The potential contribution of wild relatives for crop improvement has long been recognized and today forms an important component of breeding and conservation programs for most cultivated species in the face of emerging diseases and climate changes (Kovach and McCouch 2008; McCouch et al., 2013; Tanksley and McCouch 1997; Zhang et al., 2017). Investigating the genetic and phenotypic diversity of wild relatives of crop species allows identifying interesting traits that can be introgressed into elite lines for future breeding programs (Zhang et al. 2017). Furthermore, the comparison of genetic diversity between crops and their wild relatives allows understanding the process of domestication, and more generally the mechanisms of adaptation and diversification (Meyer and Purugganan, 2013).

The domestication process differs between annual and long-lived perennial crops (Gaut et al., 2015; Gross and Olsen, 2010). Compared to annuals, the domestication of perennial crop species is more recent in terms of generation number, with the occurrence of weaker bottlenecks, if any, and a less marked domestication syndrome (Besnard et al., 2017; Cornille et al., 2014; Cornille et al., 2012; Gaut et al., 2015; Gross and Olsen, 2010). The specificities of the domestication process in perennials are partly due to clonal propagation and long juvenile phases, that both reduce the number of sexual cycles separating domesticated individuals from their wild progenitors (Gaut et al., 2015; Miller and Gross, 2011). Recent studies have corroborated such expectations by documenting the domestication history of several perennial crops, in particular of tree species cultivated for fruit consumption (Besnard et al., 2013; Besnard et al., 2017; Cornille et al., 2012; Decroocq et al., 2016; Delplancke et al., 2013; Gros-Balthazard et al., 2017; Gross and Olsen, 2010; Hazzouri et al., 2015; Wincker, 2013) or ornamental purposes (Iwata et al., 2000; Liorzou et al., 2016; Yuan et al., 2014). These studies also revealed that domestication in fruit trees

often involved gene flow from multiple wild species in different geographic areas, such as in citrus (Wu et al., 2018), apple (Cornille et al., 2012), olive (Besnard et al., 2017), peach (Yu et al., 2018), banana (Martin et al., 2017) and date palm (Flowers et al., 2019). The study of the distribution and population structure of wild species that are closely related to crop species is therefore essential for a thorough understanding of domestication.

Apricot, *Prunus armeniaca* L., is an emblematic fruit tree species, that includes both the domesticated form, cultivated worldwide and the wild form endemic to the Tian Shan Mountains in Central Asia (Decroocq et al., 2016). Apricot has recently become an attractive model species for studying fruit tree evolution because of the naturally occurring populations displaying substantial genetic and phenotypic diversities, especially regarding resistance to sharka disease (Decroocq et al., 2016). Morphological evidence supported the view that the cultivated apricot has been domesticated in Central Asia (Vavilov, 1930; 1951), followed by subsequent diffusion toward the Caucasus (Kostina, 1960) and Europe (Bourguiba et al. 2012), along the trade routes. Genetic data combined with coalescent-based simulations (approximate Bayesian computation) further supported this hypothesis and detected weak bottlenecks in cultivated apricots (Decroocq et al., 2016). Vavilov (1951) also suggested a second origin for apricot domestication in China based on morphological observations. Genetic data supported the existence of genetic differentiation of Chinese cultivated apricots from both cultivated Irano-Caucasian and wild Central Asian apricots (Decroocq et al., 2016). However, the low number of Chinese apricot cultivars analysed ($n=18$) prevented a definitive conclusion about the existence of a second apricot domestication centre in China.

Additional wild *Prunus* species from the *Armeniaca* section that are endemic to China may also have contributed to apricot domestication in a secondary centre in Eastern Asia (Decroocq et al., 2016). In China, the cultivation of apricot started at least 3,000 years ago (Hormaza et al., 2007). There, the cultivated apricot shares its distribution with several wild endemic *Prunus* species belonging to the *Armeniaca* section, mainly *P. sibirica* L., *P. mandshurica* (Maxim.) Koehne and *P. mume* (Siebold) Siebold & Zucc. (Ledbetter, 2008). In addition to apricot domestication for fruit consumption, wild species from the *Armeniaca* section have been also cultivated for ornamental purposes, such as *P. mume* that is currently cultivated all over East Asia (Zhang et al., 2018). Previous studies have established the phylogeny of species within the *Armeniaca* section (Rehder, 1940) and their geographic distribution (Bailey and Hough, 1975). However, the contribution of the various wild *Prunus* species to the current Chinese cultivated germplasm has been little studied, or with insufficient sample sizes so far (Ai et al., 2011; Geuna et al., 2003; Zhebentyayeva et al., 2008). Moreover, previous studies on the genetic diversity of Chinese species focused

separately on cultivated (Li et al., 2014; Li et al., 2010; Liu et al., 2015; Zhang et al., 2014) or wild apricots (He et al., 2007; Li et al., 2013; Wang et al., 2014), without inferences on Chinese apricot domestication history.

We therefore studied here the domestication history of apricot, *P. armeniaca*, with an emphasis on the cultivated and wild *Prunus* apricots endemic to China. We acquired 172 additional Chinese samples and 29 additional South Central Asian cultivars compared to our previous study (Decroocq et al., 2016), amounting to a total of 577 trees collected across Central and East Asia. The new Chinese samples included 80 cultivated apricot trees (71 landraces and nine *P. mume* trees) and 92 wild apricot trees (*P. sibirica* and *P. mandshurica*). We also used a larger number of microsatellite markers than previously (Decroocq et al., 2016). With this extensive dataset, we aimed to address the following questions: (1) What is the genetic diversity and structure of wild and cultivated apricots in Central/East Asia? (2) What is the divergence history of wild lineages across Central/East Asia? (3) Were there a single domestication or multiple domestication events in apricots? (4) Which wild species did contribute to the current cultivated apricot genepool? (5) Did apricot domestication occur with ancient gene flow, with recent admixture, and/or with bottlenecks?

Material and Methods

Sample collection and DNA extraction

In order to complement the collection obtained in a previous study (Decroocq et al., 2016), we acquired further samples across the Eastern range of *Prunus* species. In total, the plant material analyzed in the current study included 271 cultivated samples and 306 wild trees (Table I-S1 and Figure I-1, see supplementary note 1 for details). We collected wild *P. armeniaca* material ($n=212$) from 2010 to 2014 in natural populations in Kazakhstan ($n=125$), Kyrgyzstan ($n=79$) and the Ili valley in the Chinese Xinjiang province ($n=8$). We sampled cultivated *P. armeniaca* individuals ($n=57$) in Northern Central Asia and the Caucasus during the same period (see details in Decroocq et al., 2016). We retrieved leaf material from European ($n=49$) and Southern Central Asian ($n=47$, mainly Pakistan, Afghanistan, Turkmenistan) cultivars and landraces from Czech, French and US national repositories (Lednice Pomology Faculty-Mendelova Univerzita V Brne, UGAFL-INRA and ARS-USDA, respectively). Landraces are traditional cultivated apricots that do not originate from breeding programs, in contrast to cultivars. The Chinese landraces ($n=71$) came from the Chinese national apricot repository (Liaoning Institute of Pomology). We collected samples of *P. sibirica* ($n=84$) in 2016 in seven natural sites from mountainous areas in Northern China (Table I-S1), while *P. mandshurica* individuals ($n=8$) originated from a single site located in

Northeast China (Heilongjiang province, Table I-1). We completed the collection with samples of cultivated *P. mume* trees ($n=9$) originating from Southern China and used for ornamental purposes, and with wild trees of *P. brigantina* Vill. ($n=2$), the only *Armeniaca* species endemic in Europe (more specifically in the French and Italian Alps); *P. mume* and *P. brigantina* samples were kindly provided by the curators of the Chinese and French national repositories (Dr. Z Gao, Nanjing Agricultural University; J-M Audergon, UGAFL-INRA). We extracted genomic DNA as described in Decroocq et al. (2016).

Microsatellite markers and polymerase chain reaction (PCR) amplification

We used 48 microsatellite markers spread across the eight *P. armeniaca* chromosomes, with two to 13 loci per chromosome. We performed PCR amplification in an Eppendorf Mastercycler and scored fragment sizes on an ABI PRISM 3730 (Applied Biosystems) as previously described (Decroocq et al., 2016). We scored alleles with GENEMAPPER v.4.0 (Applied Biosystems). From the raw dataset, we retained the 34 microsatellite markers with less than 5% missing data. Detailed information on the microsatellite markers, including their repeat motifs, sequences, as well as conditions of amplification, are given in Supplementary Table I-S2.

Analyses of population structure

To identify siblings or clonemates in our dataset, we used GenoDive (Meirmans and Van Tienderen, 2004) to assess the probability of observing unrelated individuals with the detected similar genotypes given the population allelic frequencies (corrected *Nei's* diversity calculated in GenoDive with a threshold of 50). We retained only one individual of each pair detected as clonemates or siblings for further analyses.

We identified population structure with the STRUCTURE software v.2.3.3 (Pritchard et al. 2000), without the use of *a priori* grouping information and assuming individuals had mixed ancestry with correlated allele frequencies among populations. STRUCTURE assumes sexual reproduction and random mating, but is robust to violations of these assumptions (François et al. 2010). The clustering method implemented in STRUCTURE is based on Monte Carlo Markov Chain (MCMC) simulations and is used to infer the proportion of ancestry of genotypes in K distinct clusters. We simulated K values ranging from 2 to 12 for analyses on the whole dataset ($n=577$), and from 2 to 10 for the subdatasets including only Central Asian wild *P. armeniaca* and *P. sibirica* samples ($n=288$), or *P. sibirica* alone ($n=84$). For each K , the STRUCTURE runs consisted of 10 replicates of 10,000 'burn in'

steps followed by 100,000 MCMC iterations. The resulting matrices of estimated cluster membership coefficients (Q) were permuted with CLUMPP v.1.1.2 (Jakobsson and Rosenberg, 2007). We displayed STRUCTURE barplots with DISTRUCT v.1.1 (Rosenberg, 2004). We determined the strongest level of genetic structure using ΔK (Evanno et al. 2005), as implemented in the online post processing software Structure Harvester (Earl and Vonholdt, 2012). However, the K identified by this criterion often does not correspond to the finest biologically relevant population structure; we therefore also inspected visually the barplots and chose the K value at which all clusters had well assigned individuals while no further well delimited clusters and biogeographically relevant could be identified at higher K values.

We considered an individual to be assigned to a cluster when its assignment probability was $\geq 85\%$ to this cluster. As shown by modeling studies (e.g., Vähä and Primmer, 2006), it is important to set assignment thresholds in order to detect recent hybridization and a threshold around the one we set seems optimal (Vähä and Primmer, 2006). Admixture beyond a few generations will not be detectable by methods like structure minimizing deviations from Hardy-Weinberg structure as this later is reached rapidly under panmixia (Hartl and Clark, 1997). In order to set the threshold for assigning individuals to the cluster or hybrid class, we plotted the distribution of coefficient memberships for the various genetic groups (Figure I-S1). Assignment memberships displayed Gaussian distributions beginning after the 0.85 limit for all the groups, indicating that individuals beyond this limit stand out the normal distribution and could thus be considered as hybrids. We ran analyses on datasets with the 34 microsatellite markers, and also with a restricted dataset of the 25 microsatellite markers with perfect repeats for the inference of demographic history (Table I-S2, supplementary note 2).

We explored the relationships among clusters using a factorial correspondence analysis (FCA) and SplitsTree. We performed the FCA with Genetix v4.05 (Belkhir et al. 2004) and visualized it using the 'scatterplot3d' R package ([Ligges and Mächler, 2002](#)) (R Development Core Team, URL <http://www.R-project.org>). We used the program SplitTree with the neighbornet method (Huson and Bryant, 2006).

Genetic variation and differentiation

We calculated the number of different alleles (N_a), the number of effective alleles (N_e), the observed heterozygosity (H_o) and the unbiased expected heterozygosity (H_E) with GenAlEx v6.5 (Peakall and Smouse, 2012) and ARLEQUIN v3.5 (Excoffier and Lischer, 2010). We

calculated allelic richness (A_r) and private allele frequencies (A_p) after adjustment for sample size differences among groups through the rarefaction procedure implemented in ADZE (Szpiech et al., 2008). We assessed the significance of pairwise genetic differentiation estimated using F_{ST} and $Jost'D$ in exact tests carried out with GenoDive (Meirmans and Van Tienderen, 2004). Recent events of effective population size reduction were investigated with BOTTLENECK v.1.2.02 (Piry et al., 1999), which compares the expected heterozygosity estimated from allele frequencies with that estimated from the number of alleles and the sample size, which should be identical for a neutral locus in a population at mutation-drift equilibrium.

Inference of demographic history

We used approximate Bayesian computation (ABC) to unravel the evolutionary history of apricot in Central/East Asia, with a focus on the speciation and domestication histories in China, including the question of the occurrence of ancient gene flow and bottlenecks (Decroocq et al., 2016). We built the scenarios to be tested by ABC based on the genetic clustering obtained with the STRUCTURE software, combined with genetic, historical and geographical information (see supplementary notes 2 and 3).

We used the newly developed ABC method based on a machine learning tool named “random forest” (ABC-RF) to perform model choice and parameter estimation (Pudlo et al., 2016; Raynal et al., 2019). This approach allows disentangling complex demographic models (Pudlo et al., 2016), by comparing groups of scenarios with a specific type of evolutionary events to other groups with different types of evolutionary events (instead of considering all scenarios separately) (Estoup et al., 2018), in what we will hereafter call “ABC rounds”. We used sequential rounds to compare a group of scenarios with gene flow to a group of scenarios without gene flow, a group of scenarios assuming independent domestication events from different wild lineages to a group of scenarios assuming repeated domestication events from the same wild lineage, and a group of scenarios assuming bottlenecks to a group of scenarios assuming no bottleneck during domestication. Such a grouping approach in scenario choice is more powerful than testing individually all scenarios to disentangle the main evolutionary events characterizing speciation and domestication histories (Estoup et al., 2018).

We processed ABC analyses using a nested approach with three steps, including multiple rounds within the first two steps (Estoup et al., 2018) (Figures I-S2 and I-S3, Tables I-S3 and I-S4). In total, we performed seven nested rounds of competing scenarios that we analyzed sequentially in the three steps using ABC model choice methodologies. Each

round allowed retaining one type of evolutionary event (e.g., gene flow, order of divergence, tree topology, bottleneck, Tables I-S3 and I-S4) in the subsequent scenarios to be tested. In the first step, we established the most likely scenario of the wild species divergence history (Step 1: “Wild apricot divergence”, with two rounds). In the second step, we used the selected scenario of wild apricot divergence history as a backbone to place the cultivated apricot genetic groups and thus assess the most likely scenario of domestication (Step 2: “Domestication”, with four rounds). The third step used the selected scenario of domestication to test the occurrence of bottlenecks during domestication (Step 3: “Bottleneck”, with one round).

We used ABCtoolbox (Wegmann et al., 2010) with fastsimcoal 2.5 (Excoffier and Foll, 2011) to simulate datasets keeping only the 25 perfect microsatellite markers (i.e. without interruption of the repeat motif, Table I-S2) to fit the mutational model assumed in the simulated scenarios. In order to optimally choose the clusters and individuals to include in the ABC analyses, we ran two additional STRUCTURE clustering analyses, using the same parameters as described above (i.e. K values ranging from 2 to 12, 10,000 steps of ‘burn in’, 100,000 MCMC steps for analysis, and 10 repeats for each K value), with the 25 ‘perfect’ SSR loci and samples of interest for each of the two steps of ABC analyses, keeping only individuals that were assigned at $\geq 85\%$ to a given cluster, and only clusters with at least 10 individuals (see supplementary note 2.1). We removed admixed individuals ($q > 0.85\%$) to limit potential biases due to recent admixture events and limit the number of scenarios to be tested; recent admixture events can be detected directly from STRUCTURE barplots, unlike ancient events. Ancient gene flow events during the evolutionary history of apricots were thus inferred with ABC (see below). For step 1 (“Wild divergence”), we thus only included the $n=234$ wild genotypes (Tables I-S3, I-S4 and I-S5, Figure I-S4, see supplementary note 2.1). For steps 2 and 3 (“Domestication” and “Bottleneck”), we added the $n=116$ cultivated genotypes (Tables I-S3 and I-S4, Figure I-S5, $n=350$ individuals, see supplementary note 2.2).

For each scenario, we generated 10,000 genetic datasets using coalescent simulations with model parameters drawn from prior distributions. We set prior distributions (Table I-S6) for historical and demographic parameters taking into account historical and available information from previous studies on apricots and other *Prunus* species (Chin et al., 2014; Decroocq et al., 2016; Yazbek and Oh, 2013). We assumed a non-overlapping generation time of 10 years (Zaurov et al., 2013) and we estimated the divergence times between groups x and y (T_{x-y}), the effective size of each group (N_x), and, when included in the model, the migration rate per generation between population x and y (m_{x-y}) (Table I-S6).

We calculated a set of summary statistics for microsatellite markers describing within and among population genetic diversity and differentiation with *arlsumstats* (Excoffier and Foll, 2011) for each simulated and observed dataset: the mean number of alleles per locus, the expected heterozygosity, the Garza-Williamson statistics across loci (Garza and Williamson, 2001), F_{ST} (Wright, 1949) and $(\delta\mu)^2$ Goldstein's distance (Goldstein et al., 1995). We assumed a generalized stepwise model of microsatellite evolution (Estoup et al., 2002). The mutation rate was allowed to vary across markers, with locus-specific mutation rates drawn from a gamma distribution (α , α/μ) in which μ is the mutation rate per generation and α is a shape parameter. We assumed a uniform prior distribution for μ [0.000001, 0.0001] and a uniform distribution for α [1, 30]. The parameter values were adapted from Cornille et al. (2012) and Decroocq et al. (2016).

The ABC-RF analysis provides a classification vote representing the number of times a scenario is selected as the best one among n trees in the constructed random forest. For each round, we selected the group of scenarios with the highest number of classification votes as the best group of scenarios among a total of 500 classification trees (Breiman, 2001). We computed the posterior probabilities and prior error rates (*i.e.* the probability of choosing a wrong group of scenarios when drawing model index and parameter values into the priors of the best scenario) over 10 replicate analyses (Estoup et al., 2002) for each ABC round. We used the *abcrf* v.1.7.0 R statistical package (Pudlo et al., 2016) to conduct ABC-RF. We also checked visually that the simulated models were compatible with the observed dataset by projecting the simulated and the observed datasets on the two first linear discriminant analysis (LDA) axes with the *abcrf* R statistical package (Pudlo et al., 2016) and checking that the observed dataset fell within the clouds of simulated datasets. We then performed parameter inferences using the final selected model following the three-round ABC procedure. Note that the ABC-RF approach includes the model checking step that was done *a posteriori* in previous ABC methods.

Results

Population genetic diversity and structure

We analyzed 577 apricot samples, including 271 cultivated and 306 wild samples (Tables I-1 and I-S1), genotyped with 34 microsatellite markers yielding less than 5% missing data and a total of 220 different alleles (mean allele number $N_A=6.5$; mean effective allele number $N_E=4.2$; Table I-S2). We visually inspected the barplots to identify all well delimited clusters that could be biologically relevant. We thus identified nine well-delimited clusters (Figure I-S6; Table I-2), corresponding to species and/or geographical regions (Figure I-1 and Table

I-1). We therefore considered these nine clusters as the most relevant genetic structure. Oddly, these nine clusters were not identified before $K=11$ (Figure I-1), likely due to the fact that some were represented by only a few individuals and/or had low differentiation levels. ΔK was the highest at $K=10$ (Figure I-S7a). However, the nine clusters appeared better delimited at $K=11$, which we retained for further analyses.

Four genetic clusters occurred across Europe, the Irano-Caucasian region and Central Asia (Figure I-1). Regarding wild apricots in Central Asia, we retrieved the previously identified large yellow cluster (PaNCA_W2, also including cultivated apricots from North Central Asia (Decroocq et al., 2016) and the red Central Asian cluster (PaKGZ_W1, Figure I-1). We detected four genetic clusters of cultivated apricots across Europe and Central Asia, including the previously identified grey cluster (PaEUR_C1), corresponding to the European and Irano-Caucasian cultivars, and the yellow cluster in Central Asia (PaNCA_W2) that included both wild and cultivated apricots (Figure I-1). We detected here two additional clusters of cultivated apricots, the light blue cluster in Southern Central Asia (PaSCA_C2) and the small purple cluster (PaXJ_C3) in Xinjiang province in Western China (Figure I-1). Cultivated trees with purple assignment were also found in Central Asia, with however there high levels of admixture, mostly with the yellow cluster.

We also documented the genetic structure of apricots in China. The brown cluster (China_C4) corresponded to Chinese cultivated apricots (Figure I-1c), and appeared differentiated from the Central Asian *P. armeniaca* clusters described above, and also from *P. mume* (orange cluster in Figure I-1), and from *P. sibirica*, which was subdivided into two geographically separated clusters, green and dark blue (Psib_nw_W3 and Psib_ne_W4, respectively, Figure I-1a). These findings raise questions about the classification of the Chinese cultivated apricots; we therefore named this cluster China_C4 without indication of species name. *Prunus mandshurica* and *P. brigantina* appeared admixed, likely due to the low number of samples available for these species (Figure I-1), which can prevent STRUCTURE from recognizing specific clusters (Neophytou, 2014). A second STRUCTURE analysis retaining only the 84 wild Chinese *P. sibirica* samples (Table I-1) confirmed at $K=2$ the presence of two major genetic groups in *P. sibirica*: the green one (Psib_nw_W3, Northwest China, $n=34$) and the dark blue one (Psib_ne_W4, Northeast China, $n=50$) (Figure I-S8).

The cultivated apricots presented substantial footprints of recent admixture compared to wild apricots (Figure I-1; Table I-S7). Using again the 85% assignment threshold, 34.5% of European/Irano-Caucasian cultivated apricots showed signs of recent introgression (Table I-S7). The North Central Asian cultivated apricots (PaNCA_C6, Figure I-1) included

45.6% of individuals fully assigned to the yellow wild Central Asian cluster, and 43.9% of trees admixed, mostly with the grey European cultivated apricots (PaEUR_C1) and/or with the purple Asian cultivated cluster (PaXJ_C3). The North Central Asian cultivated apricots from Xinjiang (Western province of China, PaXJ_C3) showed 28.6% of trees with high assignment to the purple cluster, and 57.1% admixed, with either the South Central Asian (PaSCA_C2), the grey European cultivated apricots and/or the yellow wild Central Asian cluster (Figure I-1). Regarding the South Central Asian cultivated apricots (PaSCA_C2), 40.4% of the trees were admixed, mainly with the European/Irano-Caucasian cultivated apricots. Among Chinese cultivated apricots, 47.4% were admixed, mostly with the clusters originating from Southern China and of the wild *P. sibirica* (Psib_nw_W3 and/or Psib_ne_W4, Figure I-1).

Genetic variation and differentiation among the apricot genetic clusters

We considered hereafter a genotype to be unequivocally assigned to a population when its assignment probability was $\geq 85\%$ to one of the nine clusters at $K=11$, which was the case for 72.1% of individuals ($n=416$, Table I-S8). We explored the genetic relationships among these nine “pure” populations with a factorial correspondence analysis (FCA, Figure I-2a); *P. mume* (orange) and Northeast *P. sibirica* (dark blue) were well separated from other apricots, with the later being distant from the Northwestern *P. sibirica* cluster (green). Excluding the two differentiated clusters (orange *P. mume* and dark blue *P. sibirica*), a second FCA (Figure I-2b) showed a clearer genetic differentiation among the seven other populations. The cultivated apricot samples then appeared the most differentiated from other populations, forming two well separated clusters (grey for Europe and brown for China), in contrast to the cultivated apricots from the yellow and purple clusters from Central Asia that fell within wild samples. All the wild forms were clumped in this second FCA (Figure I-2b), including the Northwestern *P. sibirica* cluster (green) that were mixed with the *P. armeniaca* clusters. In a third PCA (Figure I-2c) without cultivated apricots from Europe and China, we distinguished three major groups: the two Central Asian *P. armeniaca* clusters (yellow PaNCA_W2 and red PaKGZ_W1 clusters) and the Chinese *P. sibirica* cluster (Psib_nw_W3). A SplitsTree recovered similar relationships among clusters (Figure I-3). The reticulations suggested the occurrence of recent and ancient gene flow among apricots (Figure I-3): recent gene flow is indicated by reticulations reaching the terminal part of branches and disappearing on the networks without admixed individuals while ancient gene flow is suggested by reticulations closer to the base of the network.

Between-population F_{ST} and Jost's D parameters provide information about the

degree of differentiation between genetic clusters, F_{ST} being a classical index and Jost's D being less sensitive to diversity. F_{ST} and Jost's D estimates among the genetic clusters obtained with STRUCTURE analysis and filtering out admixed individuals revealed highly significant genetic differentiation among clusters (Table I-3, P -values < 0.001). The genetic diversity in the wild *P. armeniaca* apricots (mean \pm standard deviation: $H_E=0.710 \pm 0.021$) was higher than in the cultivated genetic clusters ($H_E=0.675 \pm 0.024$), although the difference was significant only for some of the pairwise comparisons (Tables I-2 and I-S9).

We further investigated the genetic diversity in apricots considering all samples, even admixed trees, and delimiting groups on the basis of the inferred genetic clusters and the eco-geographical groups previously described (Kostina, 1964) (Table I-S10). In this case, the cultivated apricots from North Central Asia showed a significantly higher genetic diversity ($H_E=0.753 \pm 0.165$) than the wild *P. armeniaca* group from South Central Asia ($H_E=0.70 \pm 0.16$, P -values <0.05). Other pairwise comparisons were not significant except for the North-Western *P. sibirica* group within consistently significantly higher genetic diversity than cultivated groups (Table I-S11). The cultivated *P. armeniaca* from North Central Asia and from the Chinese Xinjiang also showed significantly higher allelic richness ($A_R=1.756 \pm 0.167$ and 1.764 ± 0.133) than wild *P. armeniaca* populations from North and South Central Asia ($A_R=1.724 \pm 0.172$ and $A_R=1.716 \pm 0.166$, respectively; Tables I-S10 and I-S11, all P -values <0.01). The European/Irano-Caucasian cultivated apricots showed the lowest genetic diversity ($H_E=0.663 \pm 0.209$) and allelic richness ($A_R=1.661 \pm 0.211$).

Approximate Bayesian computation for testing demographic scenarios

We used approximate Bayesian computation to infer the order of lineage divergence events and the number of independent domestication events. For this goal, we focused on trees assigned to $\geq 85\%$ to a cluster to avoid biases due to recent admixture events, which can be seen from the STRUCTURE barplots. The results of the ABC-RF analyses are presented in Table I-S3, supplementary note 3 and Figures I-S2, I-S3 and I-S10. The observed summary statistics fell within the cloud of the simulated summary statistics, which did not overlap across groups of models (Figure I-S10 and supplementary note 4), indicating high power to discriminate among groups of scenarios. Only for the third step ("Bottleneck"), the simulations under different groups of scenarios strongly overlapped, indicating a lack of power to identify the number and strength of bottlenecks, and the cultivated lineages affected (Figure I-S9, Step 3 Bottleneck).

For step 1 of ABC analyses ("Wild divergence", Table I-S3 and Figures I-4a, I-4d and I-S2), the classification votes were by far the highest for the group of scenarios assuming

that the green *P. sibirica* lineage (Psib_nw_W3) resulted from an admixture event between the blue *P. sibirica* lineage (Psib_ne_W4) and the red wild *P. armeniaca* lineage (PaKGZ_W1) (72% RF-trees voting for this group of scenarios, posterior probability $P=0.84$, prior error rate=0.16). The group of scenarios assuming gene flow between the wild yellow *P. armeniaca* lineage (PaNCA_W2) and each of the red wild *P. armeniaca* (PaKGZ_W1) and green wild *P. sibirica* (Psib_nw_W3) lineages was also favored with less confidence (37% of votes, posterior probability $P=0.54$, prior error rate=0.41).

From the step 2 of ABC analyses, we inferred the domestication history of cultivated apricots using as a backbone the most likely scenario of wild species divergence history as inferred above (Table I-S3, Figure I-4b, and supplementary note 2). The classification votes were by far the highest for the group of scenarios assuming gene flow between the grey European/Irano-Caucasian (PaEUR_C1) and brown Chinese cultivated apricot (China_C4) lineages, and between the light blue South Central Asian (PaSCA_C2) and the brown Chinese cultivated apricot (China_C4) lineages (90% of votes, posterior probability $P=0.96$, prior error rate=0.0001). Then, assuming such gene flow, the classification votes were the highest for the group of scenarios assuming three independent domestication events. The fourth round of ABC analyses indicated a domestication of the light blue South Central Asian cultivated apricots (PaSCA_C2) from the red wild *P. armeniaca* lineage (PaKGZ_W1), and a domestication of the grey European cultivated apricot lineage (PaEUR_C1) from the yellow wild *P. armeniaca* lineage (PaNCA_W2) (63% of votes, posterior probability $P=0.72$, prior error rate=0.19). The fifth and sixth rounds of ABC analyses suggested a third domestication event, leading to the brown Chinese cultivated apricot lineage (China_C4). However, we lacked the power to identify the progenitor of the Chinese cultivated apricot lineage (Table I-S3): China_C4 could thus originate either from the yellow wild *P. armeniaca* lineage (PaNCA_W2) or the green *P. sibirica* lineage (Psib_nw_W3) (Figure I-4b).

In the ABC step 3, we tested the occurrence of bottlenecks for the European, the Chinese and the South Asian cultivated apricots based on the two most likely scenarios as inferred in the previous step (Table I-S3). The classification votes estimated for the observed wild and cultivated apricot microsatellite dataset returned similar probabilities for scenarios with or without bottlenecks, confirming a lack of power to test the strength and the occurrence of bottlenecks during domestication. However, further analysis using another coalescent-based simulator based on allele frequency (BOTTLENECK) revealed no footprint of a recent reduction in effective population size in either of the three cultivated clusters (Table I-S12; one-tailed Wilcoxon signed rank test, all P -values>0.36). For estimating parameters, we therefore kept the two scenarios selected in the second ABC

step, assuming no bottleneck during domestication.

Parameter estimates are presented in Table I-S13 and supplementary note 5 for the two most likely scenarios, and were averaged over the two scenarios hereafter. The ancient divergence of the red *P. armeniaca* (PaKGZ_W1) and the dark blue *P. sibirica* (Psib_ne_W4) lineages from an ancestral population was estimated to have occurred 12,138,225 ya (90% credibility interval (CI), q5% and q95%: 8,119,315-15,864,840, Table I-S13), and 11,346,407 ya (90% CI: 8,034,770- 15,804,315, Table I-S13), respectively. The origin of the green *P. sibirica* lineage (Psib_nw_W3) from an admixture event between the dark blue *P. sibirica* lineage (Psib_ne_W4) and the red *P. armeniaca* lineage (PaKGZ_W1) was inferred to have occurred ~ 339,054 ya (90% CI: 12,615 – 474,265, Table I-S13), followed by a split of the yellow wild *P. armeniaca* lineage (PaNCA_W2) from the red *P. armeniaca* lineage (PaKGZ_W1) 246,699 ya (90% CI: 6,262 – 379,215, Table I-S13, Figure I-3e).

According to one of the most likely scenarios for domestication, the brown Chinese cultivated apricot lineage (China_C4) derived from the green wild *P. sibirica* lineage (Psib_nw_W3) ~92,972 ya (90% CI: 2,730 – 281,730, Table I-S13). Assuming the alternative scenario, the brown Chinese cultivated apricot lineage (China_C4) originated from the yellow wild *P. armeniaca* lineage (PaNCA_W2) ~65,264 ya (90% CI: 1,770 – 272,190, Table I-S13 and Figure I-3f). The domestication events yielding the light blue South Central Asian cultivated apricots (PaSCA_C2) from the red wild *P. armeniaca* lineage (PaKGZ_W1) was inferred to have occurred 9,543 ya (90% CI: 420–27,810, Table I-S13, and the domestication of the grey European cultivated apricot lineages (PaEUR_C1) from the yellow wild *P. armeniaca* lineage (PaNCA_W2) 3,981 ya (90% CI: 185–19,440, Table I-S13, Figure I-3f). Mutation rates were estimated on average across the two scenarios at $6.9e^{-05}$ per bp and generation (90% CI: $2.4e^{-05}$ – $9.8e^{-05}$, Table I-S13). It should be noted, however, that the credibility intervals were very large for all these parameter estimates. In addition, CIs for the Chinese domestication events were much older than expected for crop domestication by humans, i.e. typically around 10,000 years ago. Even for the European domestication events, most CI upper bounds fell well beyond reasonable expectations. Therefore, parameter estimates should be considered with extreme caution. While ABC is a highly powerful method to compare demographic scenarios, caution is required to interpret estimates with large credibility intervals.

Discussion

Our study significantly extended sampling of wild and cultivated apricots in Eastern Asia compared to previous studies to infer the history of wild apricot species divergence and apricot domestication processes. We found genetically differentiated clusters of wild apricot species in China, with contrasted geographical distributions, that may have been shaped by the Himalayan orogeny. Central Asia also appears as a diversification centre of wild apricots, with genetically differentiated clusters, that may result from population isolation in refugia during glaciation periods. We further revealed the existence of at least three domestication events that gave rise to European, Southern Central Asian and Chinese cultivated apricots, respectively, with ancient gene flow among them, as well as massive recent admixture (Figure I-4c). We also found footprints of a fourth domestication event and extensive recent admixture events among virtually all cultivated clusters. Altogether, our study shows that apricot is an excellent model to unravel evolutionary processes at play during divergence and domestication of long-lived perennial fruit trees.

Diversification of wild apricots in Central and Eastern Asia

We confirmed in this study previous results showing genetic structure within the wild *P. armeniaca* populations in Central Asia, with a highly differentiated Kyrgyz population in the South (red cluster, Decroocq et al. 2016). Such genetic structure is probably related to glacial refugia for *P. armeniaca* in South Central Asia (Aradhya et al. 2017), followed by a recolonization of the Northern territories, including the Kazakh Tian Shan ranges, upon global warming. ABC analyses further confirmed that the yellow wild *P. armeniaca* populations in Central Asia recently diverged from the red wild *P. armeniaca* populations in Kyrgyzstan. Our estimate of the divergence time (6 to 379 Kya) included the dates suggested by historical and archaeological evidence for the last glacial maximum (LGM) (Decroocq et al. 2016), but given the large confidence interval, it should be taken with caution. A similar scenario of contraction during LGM was proposed as a driver of diversification in other perennial plant species, such as *Amygdalus* (almond; Zeinalabedini et al. (2010) and *Juglans* (nuts; Pollegioni et al. (2014).

Thanks to more extensive sampling compared to previous studies, we could infer the wild apricot evolutionary history in China. We found high genetic differentiation between the different *Armeniaca* species, *P. armeniaca*, *P. sibirica* and *P. mume*. Surprisingly, we found two highly differentiated genetic clusters (green and dark blue clusters) within the samples identified in the field as the *P. sibirica* species, with the green cluster inferred to have resulted from hybridization between *P. armeniaca* and the “pure” dark blue *P. sibirica*. As

this inference was done by retaining only individuals without footprints of recent admixture, this likely represents ancient hybridization. The taxonomy should likely be revised in the light of these results, and by adding morphological data and denser genetic markers, but this falls outside of the scope of our study. ABC analyses inferred the occurrence of gene flow between this putative hybrid species and its *P. armeniaca* parent, while they appear to occur in different geographic areas. The Himalaya separates *P. armeniaca* (yellow cluster) and the hybrid species (green cluster), except along the Xinjiang province (China) where the yellow cluster expands from the Tian Shan ranges to the Ily valley.

Regarding the divergence between wild *P. armeniaca* and *P. sibirica*, the most likely evolutionary scenario inferred with ABC analyses pictured a speciation event 8 to 16 million years ago, which includes the divergence time of ~8 million years ago between peach and almond (Velasco et al., 2016). A compelling hypothesis for the origin of divergence between peach and almond is that climatic changes after Tibetan Plateau uplift and Himalayan orogeny led to isolation and subsequent divergence of peach (on the Eastern part of Himalaya) and almond (on the Western part) lineages from a common ancestral species (Chin et al., 2014; Velasco et al., 2016). The uplift of the Himalayan Mountains is indeed known to have resulted in the formation of geographical barriers and climate changes, which fostered plant diversification (Beer et al., 2008; Yu et al., 2018; Zhisheng et al., 2001). These paleogeographic events likely also played an important role in wild apricot demographic history.

Multiple independent domestication events of apricots across Central and East Asia

Our study addressed the domestication history of cultivated apricots, showing that cultivated apricots were genetically highly differentiated from wild apricots, except the Central Asian yellow cluster that encompassed both cultivated and wild forms. We revealed that at least three independent domestication events occurred, likely four, with ongoing extensive gene flow among the cultivated groups, likely fostered by commercial exchanges along the Silk Roads and modern ever-increasing globalization. We inferred two independent domestication events from Central Asian apricots: i) the yellow lineage yielded the European/Irano-Caucasian cultivars (grey cluster), as previously shown (Q1 cluster in Decroocq et al., 2016), and ii) the Southern Central Asian red cluster gave rise to the Central Asian cultivated apricots (light blue cluster); this light blue lineage of cultivated apricots could not be distinguished from the brown Chinese lineage in our previous study due to the low number of samples (Decroocq et al., 2016). Our more extensive sampling in the present study thus allowed obtaining a more comprehensive view of apricot domestication history.

The ABC analyses further showed that Chinese cultivated apricots represented a third domestication event. In our previous study, we inferred that the Chinese cultivated apricots had originated from the yellow cluster (Q2-4 lineage in Decroocq et al. 2016). Here, ABC analyses based on a larger sample from Eastern Asia could not distinguish between scenarios in which Chinese cultivated apricots would have derive independently from the yellow Central Asian wild genetic cluster or would result from hybridization between this yellow *P. armeniaca* cluster and the green *P. sibirica* cluster. Our results are in accordance with historical evidence from the ancient Chinese literature that indicate the cultivation of apricots during the Tang and Song dynasties (739-1,400 ya), which is posterior to our estimates of apricot domestication in China (Chen, 2009; Zhu et al. 2016), together with other stone fruits, such as peach (Zheng et al. 2014), almond and cherry (Spengler et al. 2018). Again, the inferred dates of domestication had very large confidence intervals and should be considered with great caution. Domestication events in Eastern and/or Central Asia from local wild populations have been documented in other fruit crops, such as peach (*Prunus persica* and *Prunus ferganensis*) (Verde et al. 2013), apple (*Malus domestica* Borkh.) (Cornille et al. 2014), walnut (*Juglans regia* L. and *Juglans sigillata* Dode) (Zhao et al. 2018), as well as in ornamental trees, such as *P. mume* (Zhang et al. 2018). In the current study, we also identified an additional genetic cluster of cultivated apricots compared to previous studies, i.e. the purple cluster corresponding to cultivated apricots from Xinjiang near-Western Chinese province and Central Asia. We could not use this purple cluster in ABC analyses due to its low number of non-admixed samples and future studies may be able to test whether it represents an additional domestication event followed by hybridization with apricots from Central Asia. Finally, *P. mume* used for ornamental purposes appeared as a distinct genetic cluster, but further cultivated and wild samples will be needed to infer its origin.

Overall, our findings support Vavilov's hypotheses of i) two domestication centers of apricot, in China and in Central Asia (Vavilov, 1951), and ii) that cultivated fruit species originated from hybridization between a number of different species (Vavilov, 1930). We also confirmed Vavilov's hypothesis that the Irano-Caucasian area, expanding from the North-Eastern Iran to the Caucasus and Central Turkey, is a secondary center of diversity for cultivated apricots. However, the inferred dates of speciation and domestication should be taken with high caution given the large credibility intervals. Moreover, the inferred mean dates of domestication in China were much older than expectations for domestication by humans, which may be due to low power of inference or to the fact that domestication occurred from a yet unidentified Chinese wild species; in this later case, the inferred

domestication dates would actually be those of the divergence between these unidentified wild species and the sampled wild species rather than domestication dates.

Conclusion

Our study provides novel insights into the divergence history of wild apricot species across Central and Eastern Asia and of apricot domestication history. We showed that, as for many other fruit trees, Central and Eastern Asia were diversification centers for the wild apricots. We also inferred the existence of at least three domestication events that gave rise to European, Southern Central Asian and Chinese cultivated apricots, respectively; the domestication event in China possibly resulted from hybridization between wild populations from Central and Eastern Asia. Importantly, we found such extensive recent admixture among all cultivated apricots, likely due to human transport along the Silk Roads and more recently to modern breeding. Our study thus shows that apricot can be an excellent model to study speciation with gene flow and domestication in long-lived perennial fruit trees. Future studies including genome sequencing will likely provide more information and allow distinguishing between the scenarios of domestication that could not be teased apart based on microsatellite data. Genome sequencing may also reveal key genomic regions under selection during apricot domestication.

Acknowledgements

We acknowledge the European FP7 IRSES-246795 “STONE” and the ANR-13-KBBE-0006 “COBRA” grants while W-S.L. acknowledges support by the National Key Research and Development Program of China (grant SQ2018YFD100028). S.L. is a recipient of a Chinese Scholarship Council PhD grant. V.D. and S.L. thank the French Embassy in Beijing for accommodation and research support in China (Xu Guangqi program 2016-2018) and the Research Federation on Integrative Biology and Ecology of Bordeaux University for a travel grant. A.Co thank the Génétique Quantitative et Evolution-Le Moulon lab for accommodation at Univ. Paris Sud. We are grateful to Thierry Mauduit and Pascal Briard for help in the management of the *Armeniaca* collection. We are grateful to the INRA MIGALE bioinformatics platform (<http://migale.jouy.inra.fr>) for providing help and support, in particular to Véronique Martin, Eric Montaubon and Valentin Loux. We also thank Adrien Falce, Olivier Langella and Benoit Johannet for help and support on the the INRA-Génétique Quantitative et Evolution- Le Moulon lab cluster. We also thank Arnaud Estoup, Jean Michel Marin and Louis Raynal for their advice for ABC analyses. Molecular analysis was performed at the

Genomic and Sequencing Facility of Bordeaux (grants from the Conseil Regional d'Aquitaine n°20030304002FA and 20040305003FA and from the European Union, FEDER n°2003227 and from the Investissements d'avenir, ANR-10-EQPX-16-01).

We dedicate this work to the memory of the late Dr Raul Karychev, forever enthusiastic pomologist of the Research Institute of Horticulture and Viticulture, Almaty, Kazakhstan.

Data accessibility

The datasets generated during the current study, *i.e.* the SSR genotyping, are available at the INRA data portal (<https://data.inra.fr/>) where they can be freely retrieved under the link <https://doi.org/10.15454/560LSH>

Author contributions

VD conceived and designed the project with the help of TG and ACo. SL and SD planned and designed genotyping experiments. SL, SD, ACh genotyped the plants. SL and ACo performed the analyses, ACo ran the ABC analyses. DT and JPE managed the *Armeniaca* collection and W-SL the Chinese *Armeniaca* repository. SL and W-SL shared effort in collecting Chinese *Armeniaca* species. SL, VD, TG and ACo drafted the manuscript. All authors read and approved the final version of the manuscript. They declare no conflict of interest.

References

- Ai, P. F., Zhen, Z. J., & Jin, Z. Z. (2011). Genetic diversity and relationships within sweet kernel apricot and related *Armeniaca* species based on sequence-related amplified polymorphism markers. *Biochemical Systematics and Ecology*, 39(4-6), 694-699. doi:10.1016/j.bse.2011.05.026
- Aradhya, M., Velasco, D., Ibrahimov, Z., Toktoraliev, B., Maghradze, D., Musayev, M., . . . Preece, J. E. (2017). Genetic and ecological insights into glacial refugia of walnut (*Juglans regia* L.). *PLoS One*, 12(10), e0185974. doi:10.1371/journal.pone.0185974
- Bailey, C. H., & Hough, L. F. (1975). Apricots. In J. M. Janick, J. N. (Ed.), *Advances in Fruit Breeding*. (pp. 367-384.). West Purdue University: Press Lafayette.
- Beer, R., Kaiser, F., Schmidt, K., Armann, B., Carraro, G., Grisa, E., & Tinner, W. (2008). Vegetation history of the walnut forests in Kyrgyzstan (Central Asia): natural or anthropogenic origin? *Quaternary Science Reviews*, 27(5-6), 621-632. doi:10.1016/j.quascirev.2007.11.012
- Belfanti, E., Silfverberg-Dilworth, E., Tartarini, S., Patocchi, A., Barbieri, M., Zhu, J., . . . Sansavini, S. (2004). The HcrVf2 gene from a wild apple confers scab resistance to a transgenic cultivated variety. *Proceedings of the National Academy of Sciences*, 101(3), 886-890. doi:10.1073/pnas.0304808101
- Belkhir, K., Borsa, P., Chikhi, L., Raufaste, N., & Bonhomme, F. (2004). GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. <http://www.genetix.univ-montp2.fr/genetix/genetix.htm>.

CHAPTER I

- Besnard, G., El Bakkali, A., Haouane, H., Baali-Cherif, D., Moukhli, A., & Khadari, B. (2013). Population genetics of Mediterranean and Saharan olives: geographic patterns of differentiation and evidence for early generations of admixture. *Annals of Botany*, *112*(7), 1293-1302. doi:10.1093/aob/mct196
- Besnard, G., Terral, J.-F., & Cornille, A. (2017). On the origins and domestication of the olive: a review and perspectives. *Annals of botany*, *121*(3), 385-403. doi:doi.org/10.1093/aob/mcy002
- Bourguiba, H., Audergon, J. M., Krichen, L., Trifi-Farah, N., Mamouni, A., Trabelsi, S., . . . Khadari, B. (2012). Loss of genetic diversity as a signature of apricot domestication and diffusion into the Mediterranean Basin. *BMC Plant Biology*, *12*(49), 1471-2229. doi:Artn 49 10.1186/1471-2229-12-49
- Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5-32. doi:10.1023/A:1010933404324
- Chen, H. Y. (2009). A buddhist classification of animals and plants in early Tang times. *Journal of Asian History*, *43*(1), 31-51.
- Chin, S. W., Shaw, J., Haberle, R., Wen, J., & Potter, D. (2014). Diversification of almonds, peaches, plums and cherries-molecular systematics and biogeographic history of *Prunus* (Rosaceae). *Molecular Phylogenetics and Evolution*, *76*, 34-48. doi:10.1016/j.ympev.2014.02.024
- Collard, B. C., & Mackill, D. J. (2008). Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *363*(1491), 557-572. doi:10.1098/rstb.2007.2170
- Cornille, A., Giraud, T., Smulders, M. J., Roldan-Ruiz, I., & Gladieux, P. (2014). The domestication and evolutionary ecology of apples. *Trends in Genetics*, *30*(2), 57-65. doi:10.1016/j.tig.2013.10.002
- Cornille, A., Gladieux, P., Smulders, M. J., Roldan-Ruiz, I., Laurens, F., Le Cam, B., . . . Giraud, T. (2012). New insight into the history of domesticated apple: secondary contribution of the European wild apple to the genome of cultivated varieties. *PLOS Genetics*, *8*(5), e1002703. doi:10.1371/journal.pgen.1002703
- Decroocq, S., Cornille, A., Tricon, D., Babayeva, S., Chague, A., Eyquard, J. P., . . . Decroocq, V. (2016). New insights into the history of domesticated and wild apricots and its contribution to Plum pox virus resistance. *Molecular Ecology*, *25*(19), 4712-4729. doi:10.1111/mec.13772
- Delplancke, M., Alvarez, N., Benoit, L., Espindola, A., H, I. J., Neuenschwander, S., & Arrigo, N. (2013). Evolutionary history of almond tree domestication in the Mediterranean basin. *Molecular Ecology*, *22*(4), 1092-1104. doi:10.1111/mec.12129
- Earl, D. A., & Vonholdt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, *4*(2), 359-361. doi:10.1007/s12686-011-9548-7
- Estoup, A., Jarne, P., & Cornuet, J. M. (2002). Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Molecular Ecology*, *11*(9), 1591-1604. doi:10.1046/j.1365-294X.2002.01576.x
- Estoup, A., Raynal, L., Verdu, P., & Marin, J. M. (2018). Model choice using Approximate Bayesian Computation and Random Forests: analyses based on model grouping to make inferences about the genetic history of Pygmy human populations. *Journal of the Société française de Statistique*, *159*(3), 167-190.
- 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. doi:10.1111/j.1365-294X.2005.02553.x

- Excoffier, L., & Foll, M. (2011). fastsimcoal: a continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics*, 27(9), 1332-1334. doi:10.1093/bioinformatics/btr124
- Excoffier, L., & Lischer, H. E. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, 10(3), 564-567. doi:10.1111/j.1755-0998.2010.02847.x
- Flowers, J. M., Hazzouri, K. M., Gros-Balthazard, M., Mo, Z., Koutroumpa, K., Perrakis, A., . . . Purugganan, M. D. (2019). Cross-species hybridization and the origin of North African date palms. *Proceedings of the National Academy of Sciences*, 116(5), 1651-1658. doi:10.1073/pnas.1817453116
- Francois, O., & Durand, E. (2010). Spatially explicit Bayesian clustering models in population genetics. *Molecular Ecology Resources*, 10(5), 773-784. doi:10.1111/j.1755-0998.2010.02868.x
- Garza, J. C., & Williamson, E. G. (2001). Detection of reduction in population size using data from microsatellite loci. *Molecular Ecology*, 10(2), 305-318. doi:10.1046/j.1365-294x.2001.01190.x
- Gaut, B. S., Diez, C. M., & Morrell, P. L. (2015). Genomics and the contrasting dynamics of annual and perennial domestication. *Trends in Genetics*, 31(12), 709-719. doi:10.1016/j.tig.2015.10.002
- Geuna, F., Toschi, M., & Bassi, D. (2003). The use of AFLP markers for cultivar identification in apricot. *Plant Breeding*, 122(6), 526-531. doi:10.1111/j.1439-0523.2003.00897.x
- Goldstein, D. B., Linares, A. R., Cavallisforza, L. L., & Feldman, M. W. (1995). An evaluation of genetic distances for use with microsatellite loci. *Genetics*, 139(1), 463-471.
- Gros-Balthazard, M., Galimberti, M., Kousathanas, A., Newton, C., Ivorra, S., Paradis, L., . . . Wegmann, D. (2017). The discovery of wild date palms in Oman reveals a complex domestication history involving centers in the Middle East and Africa. *Current Biology*, 27(14), 2211-2218 e2218. doi:10.1016/j.cub.2017.06.045
- Gross, B. L., & Olsen, K. M. (2010). Genetic perspectives on crop domestication. *Trends in Plant Science*, 15(9), 529-537. doi:10.1016/j.tplants.2010.05.008
- Hazzouri, K. M., Flowers, J. M., Visser, H. J., Khierallah, H. S. M., Rosas, U., Pham, G. M., . . . Purugganan, M. D. (2015). Whole genome re-sequencing of date palms yields insights into diversification of a fruit tree crop. *Nature Communications*, 6, 8824. doi:10.1038/ncomms9824
- He, T. M., Chen, X. S., Xu, Z., Gao, J. S., Lin, P. J., Liu, W., . . . 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. doi:10.1007/s10722-006-0013-5
- Hormaza, J., Yamane, H., & Rodrigo, J. (2007). Apricot. In K. Chittaranjan (Ed.), *Fruits and Nuts* (pp. 171-187): Springer.
- Huson, D. H., & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, 23(2), 254-267. doi.org/10.1093/molbev/msj030
- Iwata, H., Kato, T., & Ohno, S. (2000). Triparental origin of Damask roses. *Gene*, 259(1-2), 53-59. doi:10.1016/S0378-1119(00)00487-X
- Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, 23(14), 1801-1806. doi:10.1093/bioinformatics/btm233
- Kostina, K. (1960). The use of varietal resources of apricots for breeding. *Trud. nikit. bot. Sad. (Trans. Nikita bot. Gdn.)*, 40, 45-63.

CHAPTER I

- Kostina, K. (1964). Application of phytogeographical method to apricot classification. *Trud Nikit Bot Sad*, 37.
- Kovach, M. J., & McCouch, S. R. (2008). Leveraging natural diversity: back through the bottleneck. *Current Opinion in Plant Biology*, 11(2), 193-200. doi:10.1016/j.pbi.2007.12.006
- Larson, G., Piperno, D. R., Allaby, R. G., Purugganan, M. D., Andersson, L., Arroyo-Kalin, M., . . . Fuller, D. Q. (2014). Current perspectives and the future of domestication studies. *Proceedings of the National Academy of Sciences*, 111(17), 6139-6146. doi:10.1073/pnas.1323964111
- Ledbetter, C. A. (2008). Apricots. In J. F. Hancock (Ed.), *Temperate fruit crop breeding: germplasm to genomics* (pp. 39-82): Springer Science & Business Media.
- Li, M., Zhao, Z., Miao, X., & Zhou, J. (2013). Genetic diversity and population structure of Siberian apricot (*Prunus sibirica* L.) in China. *International Journal of Molecular Sciences*, 15(1), 377-400. doi:10.3390/ijms15010377
- Li, M., Zhao, Z., & Miao, X. J. (2014). Genetic diversity and relationships of apricot cultivars in North China revealed by ISSR and SRAP markers. *Scientia Horticulturae*, 173, 20-28. doi:10.1016/j.scienta.2014.04.030
- Li, X., Wang, Y., Wang, B., Wang, C., Shangguan, L., Huang, Z., & Fang, J. (2010). Genetic relationships between fruiting and flowering mei (*Prunus mume*) cultivars using SNP markers. *Journal of Horticultural Science & Biotechnology*, 85(4), 329-334. doi:10.1080/14620316.2010.11512676
- Ligges, U., & Mächler, M. (2002). *Scatterplot3d - an R package for visualizing multivariate data*. Retrieved from <https://EconPapers.repec.org/RePEc:zbw:sfb475:200222>
- Liorzou, M., Pernet, A., Li, S., Chastellier, A., Thouroude, T., Michel, G., . . . Grapin, A. (2016). Nineteenth century French rose (*Rosa* sp.) germplasm shows a shift over time from a European to an Asian genetic background. *Journal of Experimental Botany*, 67(15), 4711-4725. doi:10.1093/jxb/erw269
- Liu, M. P., Du, H. Y., Zhu, G. P., Fu, D. L., & Tana, W. Y. (2015). Genetic diversity analysis of sweet kernel apricot in China based on SSR and ISSR markers. *Genetics and Molecular Research*, 14(3), 9722-9729. doi:10.4238/2015.August.19.4
- Martin, G., Carreel, F., Coriton, O., Hervouet, C., Cardi, C., Derouault, P., . . . D'Hont, A. (2017). Evolution of the banana genome (*Musa acuminata*) is impacted by large chromosomal translocations. *Molecular Biology and Evolution*, 34(9), 2140-2152. doi:10.1093/molbev/msx164
- McCouch, S., Baute, G. J., Bradeen, J., Bramel, P., Bretting, P. K., Buckler, E., . . . Zamir, D. (2013). Agriculture: Feeding the future. *Nature*, 499(7456), 23-24. doi:10.1038/499023a
- Meirmans, P. G., & Van Tienderen, P. H. (2004). GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes*, 4(4), 792-794. doi:10.1111/j.1471-8286.2004.00770.x
- Meyer, R. S., & Purugganan, M. D. (2013). Evolution of crop species: genetics of domestication and diversification. *Nature Reviews Genetics*, 14(12), 840-852. doi:10.1038/nrg3605
- Miller, A. J., & Gross, B. L. (2011). From forest to field: perennial fruit crop domestication. *American Journal of Botany*, 98(9), 1389-1414. doi:10.3732/ajb.1000522
- Nagaraju, J., Kathirvel, M., Kumar, R. R., Siddiq, E. A., & Hasnain, S. E. (2002). Genetic analysis of traditional and evolved Basmati and non-Basmati rice varieties by using fluorescence-based ISSR-PCR and SSR markers. *Proceedings of the National Academy of Sciences*, 99(9), 5836-5841. doi:10.1073/pnas.042099099

- Neophytou, C. (2014). Bayesian clustering analyses for genetic assignment and study of hybridization in oaks: effects of asymmetric phylogenies and asymmetric sampling schemes. *Tree Genetics & Genomes*, 10(2), 273-285. doi:10.1007/s11295-013-0680-2
- Peakall, R., & Smouse, P. E. (2012). GenAIEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research--an update. *Bioinformatics*, 28(19), 2537-2539. doi:10.1093/bioinformatics/bts460
- Piry, S., Luikart, G., & Cornuet, J. M. (1999). BOTTLENECK: A computer program for detecting recent reductions in the effective population size using allele frequency data. *Journal of Heredity*, 90(4), 502-503. doi:10.1093/jhered/90.4.502
- Pollegioni, P., Woeste, K. E., Chiocchini, F., Olimpieri, I., Tortolano, V., Clark, J., . . . Malvolti, M. E. (2014). Landscape genetics of Persian walnut (*Juglans regia* L.) across its Asian range. *Tree Genetics & Genomes*, 10(4), 1027-1043. doi:10.1007/s11295-014-0740-2
- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155(2), 945-959.
- Pudlo, P., Marin, J. M., Estoup, A., Cornuet, J. M., Gautier, M., & Robert, C. P. (2016). Reliable ABC model choice via random forests. *Bioinformatics*, 32(6), 859-866. doi:10.1093/bioinformatics/btv684
- Purugganan, M. D., & Fuller, D. Q. (2009). The nature of selection during plant domestication. *Nature*, 457(7231), 843-848. doi:10.1038/nature07895
- Raynal, L., Marin, J. M., Pudlo, P., Ribatet, M., Robert, C. P., & Estoup, A. (2019). ABC random forests for Bayesian parameter inference. *Bioinformatics*, 35(10), 1720-1728. doi:10.1093/bioinformatics/bty867
- Rehder, A. (1940). Manual of cultivated trees and shrubs hardy in North America: exclusive of the subtropical and warmer temperate regions. In: The Macmillan Company.
- Riaz, S., Boursiquot, J. M., Dangl, G. S., Lacombe, T., Laucou, V., Tenschler, A. C., & Walker, M. A. (2013). Identification of mildew resistance in wild and cultivated Central Asian grape germplasm. *BMC Plant Biology*, 13(1), 149. doi:10.1186/1471-2229-13-149
- Rosenberg, N. A. (2004). DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes*, 4(1), 137-138. doi:10.1046/j.1471-8286.2003.00566.x
- Spengler, R. N., Maksudov, F., Bullion, E., Merkle, A., Hermes, T., & Frachetti, M. (2018). Arboreal crops on the medieval Silk Road: Archaeobotanical studies at Tashbulak. *PLoS One*, 13(8), e0201409. doi:10.1371/journal.pone.0201409
- Szpiech, Z. A., Jakobsson, M., & Rosenberg, N. A. (2008). ADZE: a rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics*, 24(21), 2498-2504. doi:10.1093/bioinformatics/btn478
- Tanksley, S. D., & McCouch, S. R. (1997). Seed banks and molecular maps: unlocking genetic potential from the wild. *Science*, 277(5329), 1063-1066. doi:10.1126/science.277.5329.1063
- Vavilov, N. I. (1930). Wild progenitors of the fruit trees of Turkistan and the Caucasus and the problem of the origin of fruit trees. *Report. and Proceedings. 9th International Horticultural Congress*, 271-286.
- Vavilov, N. I. (1951). Phytogeographic basis of plant breeding. The origin, variation, immunity and breeding of cultivated plants (tr. K.S. Chester). *Chron Bot*, 13: 366 p.
- Velasco, D., Hough, J., Aradhya, M., & Ross-Ibarra, J. (2016). Evolutionary genomics of peach and almond domestication. *G3 (Bethesda)*, 6(12), 3985-3993. doi:10.1534/g3.116.032672

CHAPTER I

- Verde, I., Abbott, A. G., Scalabrin, S., Jung, S., Shu, S., Marroni, F., . . . Rokhsar, D. S. (2013). The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. *Nature Genetics*, *45*(5), 487-494. doi:10.1038/ng.2586
- Voss-Fels, K. P., Stahl, A., & Hickey, L. T. (2019). Q&A: modern crop breeding for future food security. *BMC Biology*, *17*(1), 18. doi:10.1186/s12915-019-0638-4
- Wang, Z., Kang, M., Liu, H., Gao, J., Zhang, Z., Li, Y., . . . Pang, X. (2014). High-level genetic diversity and complex population structure of Siberian apricot (*Prunus sibirica* L.) in China as revealed by nuclear SSR markers. *PLoS One*, *9*(2), e87381. doi:10.1371/journal.pone.0087381
- Wegmann, D., Leuenberger, C., Neuenschwander, S., & Excoffier, L. (2010). ABCtoolbox: a versatile toolkit for approximate Bayesian computations. *BMC Bioinformatics*, *11*(1), 116. doi:10.1186/1471-2105-11-116
- Wincker, P. (2013). Genomics and fruit crop selection. *Nat Genet*, *45*(1), 9-10. doi:10.1038/ng.2498
- Wright, S. (1949). The Genetical Structure of Populations. *Annals of Eugenics*, *15*(1), 323-354. doi:10.1111/j.1469-1809.1949.tb02451.x
- Wu, G. A., Terol, J., Ibanez, V., Lopez-Garcia, A., Perez-Roman, E., Borreda, C., . . . Talon, M. (2018). Genomics of the origin and evolution of Citrus. *Nature*, *554*(7692), 311-316. doi:10.1038/nature25447
- Yazbek, M., & Oh, S. H. (2013). Peaches and almonds: phylogeny of *Prunus* subg. *Amygdalus* (Rosaceae) based on DNA sequences and morphology. *Plant Systematics and Evolution*, *299*(8), 1403-1418. doi:10.1007/s00606-013-0802-1
- Yu, Y., Fu, J., Xu, Y., Zhang, J., Ren, F., Zhao, H., . . . Xie, H. (2018). Genome re-sequencing reveals the evolutionary history of peach fruit edibility. *Nature Communications*, *9*(1), 5404. doi:10.1038/s41467-018-07744-3
- Yuan, J. H., Cornille, A., Giraud, T., Cheng, F. Y., & Hu, Y. H. (2014). Independent domestications of cultivated tree peonies from different wild peony species. *Molecular Ecology*, *23*(1), 82-95. doi:10.1111/mec.12567
- Zaurov, D., Molnar, T., Eisenman, S., M. Ford, T., Mavlyanova, R., M. Capik, J., . . . C. Goffreda, J. (2013). Genetic resources of apricots (*Prunus armeniaca* L.) in Central Asia. *48*, 681-691. doi:10.21273/hortsci.48.6.681
- Zeinalabedini, M., Khayam-Nekoui, M., Grigorian, V., Gradziel, T. M., & Martinez-Gomez, P. (2010). The origin and dissemination of the cultivated almond as determined by nuclear and chloroplast SSR marker analysis. *Scientia Horticulturae*, *125*(4), 593-601. doi:10.1016/j.scienta.2010.05.007
- Zhang, H., Mittal, N., Leamy, L. J., Barazani, O., & Song, B. H. (2017). Back into the wild-Apply untapped genetic diversity of wild relatives for crop improvement. *Evolutionary Applications*, *10*(1), 5-24. doi:10.1111/eva.12434
- Zhang, Q., Zhang, H., Sun, L., Fan, G., Ye, M., Jiang, L., . . . Cheng, T. (2018). The genetic architecture of floral traits in the woody plant *Prunus mume*. *Nature Communications*, *9*(1), 1702. doi:10.1038/s41467-018-04093-z
- Zhang, Q. P., Liu, D. C., Liu, S., Liu, N., Wei, X., Zhang, A. M., & Liu, W. S. (2014). Genetic diversity and relationships of common apricot (*Prunus armeniaca* L.) in China based on simple sequence repeat (SSR) markers. *Genetic Resources and Crop Evolution*, *61*(2), 357-368. doi:10.1007/s10722-013-0039-4
- Zhao, P., Zhou, H. J., Potter, D., Hu, Y. H., Feng, X. J., Dang, M., . . . Woeste, K. (2018). Population genetics, phylogenomics and hybrid speciation of *Juglans* in China determined from whole chloroplast genomes, transcriptomes, and genotyping-by-sequencing (GBS). *Mol Phylogenet Evol*, *126*, 250-265. doi:10.1016/j.ympev.2018.04.014

-
- Zhebentyayeva, T. N., Reighard, G. L., Lalli, D., Gorina, V. M., Krska, B., & Abbott, A. G. (2008). Origin of resistance to plum pox virus in Apricot: what new AFLP and targeted SSR data analyses tell. *Tree Genetics & Genomes*, 4(3), 403-417. doi:10.1007/s11295-007-0119-8
- Zheng, Y., Crawford, G. W., & Chen, X. (2014). Archaeological evidence for peach (*Prunus persica*) cultivation and domestication in China. *PLoS One*, 9(9), e106595. doi:10.1371/journal.pone.0106595
- Zhisheng, A., Kutzbach, J. E., Prell, W. L., & Porter, S. C. (2001). Evolution of Asian monsoons and phased uplift of the Himalaya-Tibetan plateau since Late Miocene times. *Nature*, 411(6833), 62-66. doi:10.1038/35075035
- Zhu, R., Zhang, B., Cai, C., & Wang, Z. (2016). The Food of Song. In *A Social History of Medieval China* (pp. 74-76): Cambridge University Press.

Figures

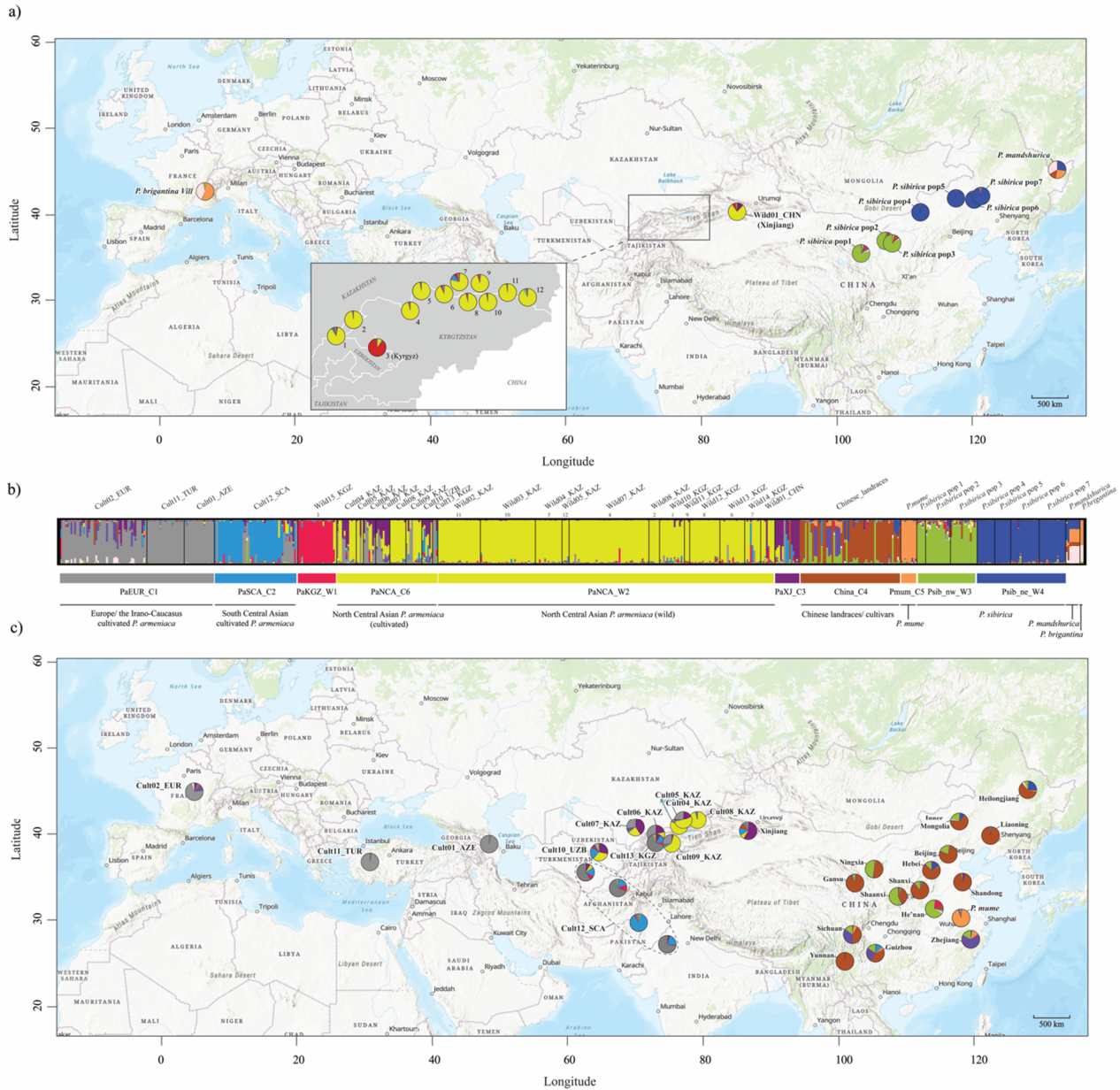


Figure I-1. Genetic structure among the 577 samples (271 cultivated and 306 wild samples) of the five *Prunus* species (*P. armeniaca*, *P. sibirica*, *P. mandshurica*, *P. mume* and *P. brigantina*) analyzed in this study with 34 microsatellite markers. a) Map of the wild sample origins. The partial enlarged view represented on the map corresponds to natural populations of *P. armeniaca* from Central Asia. b) Genetic structure as inferred with STRUCTURE for $K=11$, where nine well-delimited clusters were identified (Figure I-S6) and sorted from West (Europe) to East (China). c) Map of cultivated apricot origins. Pie chart colors correspond to the colors of Bayesian clustering assignment.

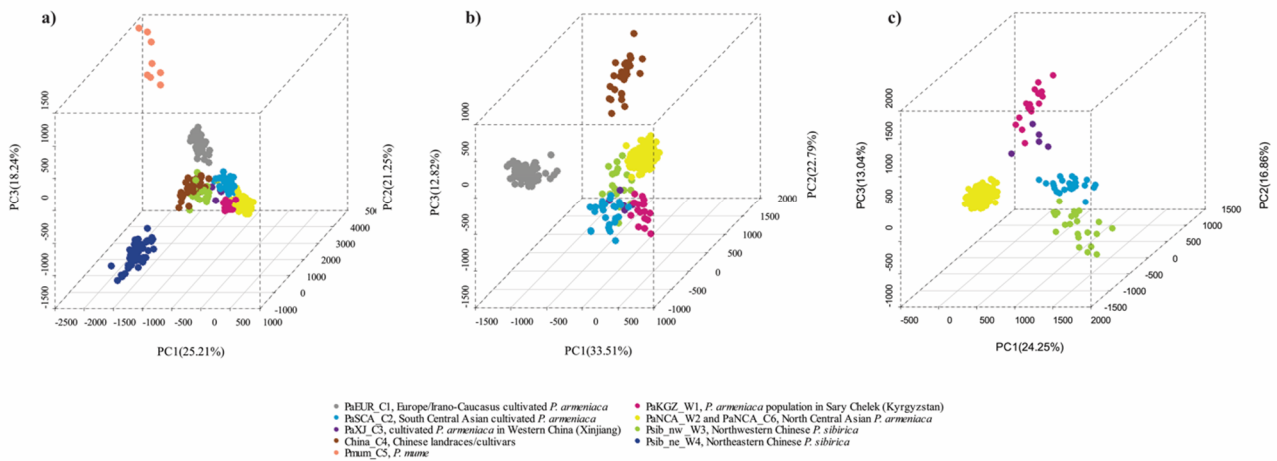


Figure I-2. Factorial correspondence analyses (FCAs). (a) including the 416 individuals with membership coefficient $\geq 85\%$ to a cluster in the STRUCTURE analysis at $K=11$, colored as in Figure I-1; (b) excluding *P. mume* and northeast wild apricot *P. sibirica*; (c) excluding European (PaEUR_C1) and Chinese (China_C4) cultivated apricots. Individuals are colored according to the cluster to which they are assigned at $K=11$ in the STRUCTURE analysis on the subdataset (Figure I-S6).

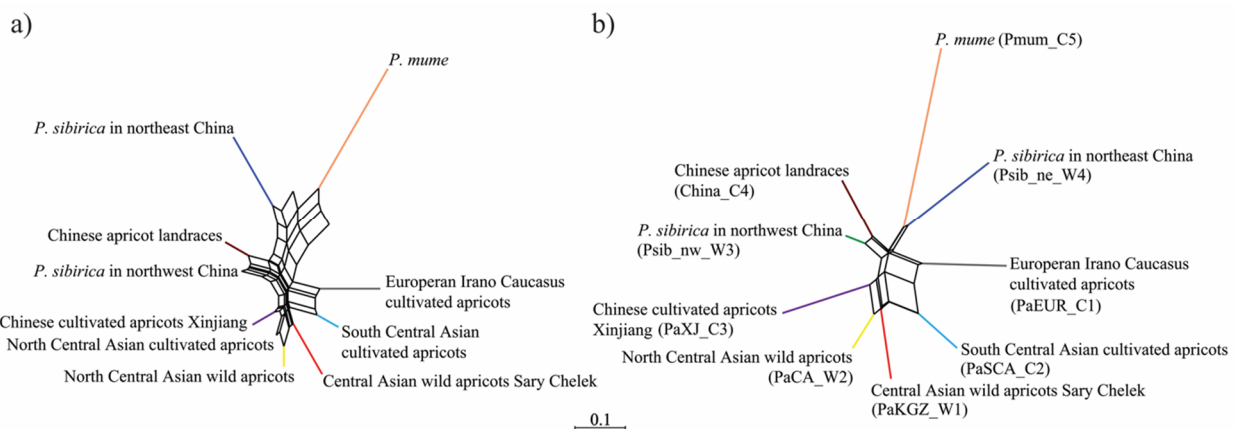


Figure I-3. Armeniaca phylogenetic networks obtained from Split tree. a) Reticulate network issued from all *P. armeniaca*, *P. sibirica* and *P. mume* individuals and b) from the nine genetic clusters depicted in Figure I-1. The internal nodes correspond to ancestry, and the edges represent patterns of descent.

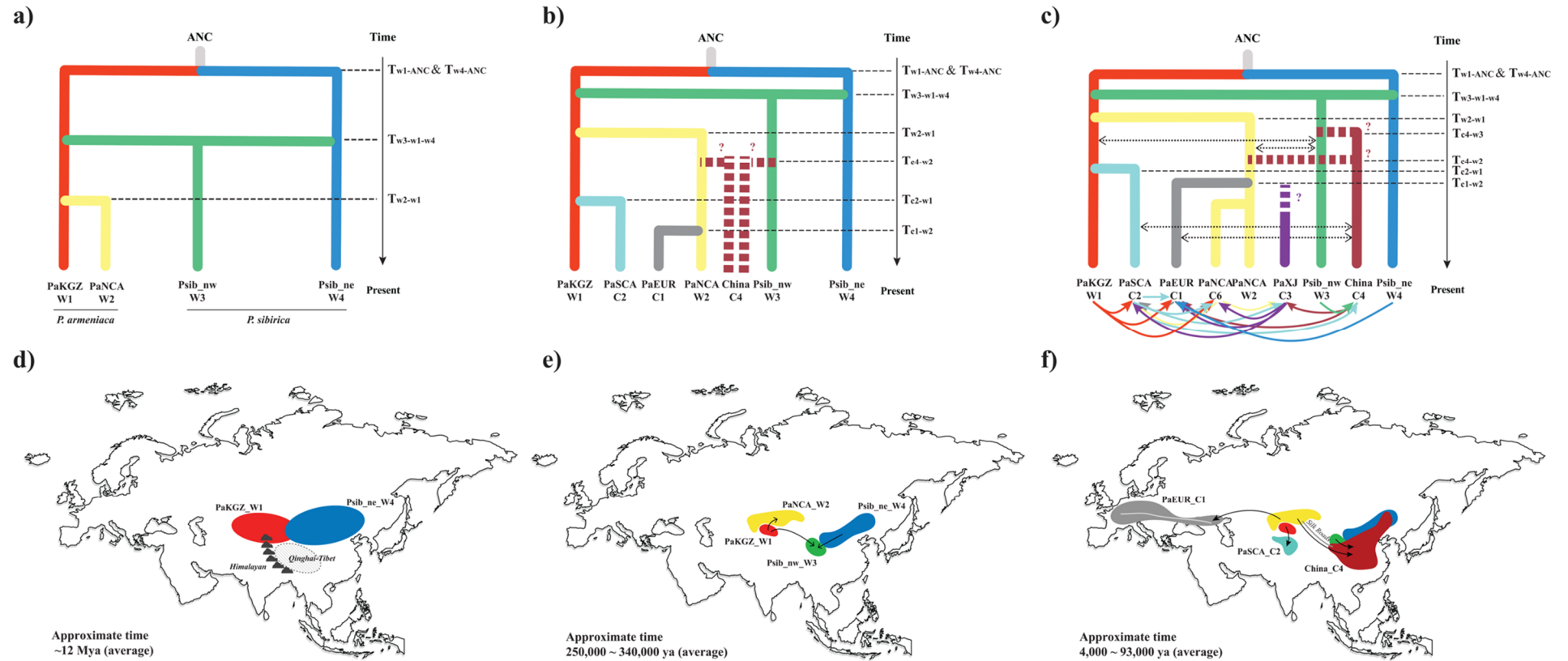


Figure I-4. Most likely scenarios for the divergence and domestication of apricot inferred with approximate Bayesian computation. a) The most likely scenario of divergence of wild apricots. b) Scenario of apricot domestication based on the two most likely models (*scF_GF_5_GF* and *scF_GF_8_GF*). c) The most likely scenario of species diversification and domestication highlighting ancient gene flow (dotted arrows) and recent admixture (plain arrow at the bottom) events. Figures I-4d to I-4f depict the biogeographic history of wild and cultivated apricots, with the approximate periods of time, drawn from ABC inferences.

Tables

Table I-1. Geographic information and sample size of cultivated and wild *Prunus armeniaca* and close relatives from the *Armeniaca* section used in this study.

Code	Size	Type of material	Habitat	Origin	Collection locality or area	Eco-geographical group
Breeding varieties, local and ancient cultivars and landraces						
Cult01_AZE	17	landraces	cultivated	Azerbaijan	Azerbaijan landraces	Irano-caucasian
Cult02_EUR	49	cultivars	cultivated	EU, US	Occidental cultivars (EU, US)	European
Cult04_KAZ	11	landraces/wild	cultivated	Kazakhstan	Almaty Pomological garden	Dzhungar-Zailig
Cult05_KAZ	4	landraces	semi-wild	Kazakhstan	Ak-Kain	Dzhungar-Zailig
Cult06_KAZ	7	landraces	cultivated	Kazakhstan	Chymkent Dendro Park	Central Asian
Cult07_KAZ	8	landraces	semi-wild	Kazakhstan	Sayram	Dzhungar-Zailig
Cult08_KAZ	9	landraces	semi-wild	Kazakhstan	Wine yard irrigation canal	Dzhungar-Zailig
Cult09_KAZ	9	local cultivars	cultivated	Kazakhstan	Almaty market	Dzhungar-Zailig
Cult10_UZB	6	landraces	cultivated	Uzbekistan	Boukhara	Central Asian
Cult11_TUR	21	landraces	cultivated	Turkey	Eastern Anatolian landraces	Irano-caucasian
Cult12_SCA	47	landraces	cultivated	South Central Asia	South Central Asian countries	Central Asian
Cult13_KGZ	3	landraces	cultivated	Kyrgyzstan	Arslan Bob	Central Asian
Chinese_landraces	71	landraces	cultivated	China	Xiongyue apricot repository	China
<i>Prunus mume</i>	9	landraces	cultivated	China	Southern China	-
Total cultivated	271					
<i>P. armeniaca</i> natural populations						
Wild01_CHN	8	wild	montane forest	China	Ily valley	Dzhungar-Zailig
Wild02_KAZ	24	wild	montane forest	Kazakhstan	Belbulak Canyon	Dzhungar-Zailig
Wild03_KAZ	31	wild	montane forest	Kazakhstan	Esik Lake	Dzhungar-Zailig
Wild04_KAZ	15	wild	montane forest	Kazakhstan	Medeu valley	Dzhungar-Zailig
Wild05_KAZ	4	wild	montane forest	Kazakhstan	Turgen Valley	Dzhungar-Zailig
Wild07_KAZ	45	wild	montane forest	Kazakhstan	Big Almaty Lake	Dzhungar-Zailig
Wild08_KAZ	6	wild	montane forest	Kazakhstan	Aksu Zhabagyly National Park	Dzhungar-Zailig
Wild10_KGZ	14	wild	montane forest	Uzbekistan	Urukty river valley	Dzhungar-Zailig
Wild11_KGZ	3	wild	montane forest	Kyrgyzstan	Issyk Kul Anan'Yevo village	Dzhungar-Zailig
Wild12_KGZ	17	wild	montane forest	Kyrgyzstan	Issyk Kul Orto Byrosun river	Dzhungar-Zailig
Wild13_KGZ	14	wild	montane forest	Kyrgyzstan	Chuy River / Boom canyon	Dzhungar-Zailig
Wild14_KGZ	9	wild	montane forest	Kyrgyzstan	Ala Archa National Park	Dzhungar-Zailig
Wild15_KGZ	22	wild	montane forest	Kyrgyzstan	Sary Chelek National Park	Central Asian
Total <i>Prunus armeniaca</i>	212					
<i>P. sibirica</i> (wild species)						
<i>Prunus sibirica</i> pop 1	14	wild	montane forest	China	Gansu province	--
<i>Prunus sibirica</i> pop 2	5	wild	montane forest	China	Shaanxi province	--
<i>Prunus sibirica</i> pop 3	15	wild	montane forest	China	Shaanxi province	--
<i>Prunus sibirica</i> pop 4	10	wild	montane forest	China	Shanxi province	--
<i>Prunus sibirica</i> pop 5	9	wild	montane forest	China	Inner Mongolia province	--
<i>Prunus sibirica</i> pop 6	16	wild	montane forest	China	Liaoning province	--
<i>Prunus sibirica</i> pop 7	15	wild	montane forest	China	Liaoning province	--
Total <i>Prunus sibirica</i>	84					
Other species of the <i>Armeniaca</i> section sampled in natural populations						
<i>Prunus brigantina</i>	2	wild	montane forest	France	Alpen	--
<i>Prunus mandshurica</i>	8	wild	montane forest	China	Heilongjiang province	--
Whole dataset (wild+cult)	577					

Footnotes: "Cult#" refers to geographical groups of apricot landraces and cultivars, followed by country codes; "Wild#" refers to natural populations sampled in Kazakhstan (KAZ), Uzbekistan (UZB) and Kyrgyzstan (KGZ). The Chinese landrace group is a repository of apricot samples collected across Eastern and Western China. Local cultivars and landraces from the Central Asian eco-geographical group (Cult04 to Cult10) were described in a previous study (Decroocq *et al.* 2016). Landraces in the Cult12_SCA group originate in majority from Pakistan but also from Turkmenistan, Afghanistan, India (see details in Table I-S1). The eco-geographical groups correspond to the classification of Kostina (1960).

Table I-2. Genetic diversity estimates for the nine apricot clusters ($n=416$ individuals, 34 SSR markers) inferred with STRUCTURE for $K=11$.

Genetic cluster	N	H_O	H_E	A_R	A_P
<i>P. armeniaca</i> cultivated clusters					
PaEUR_C1	64	0.568 (0.199)	0.639 (0.214)	2.713	0.362
PaSCA_C2	26	0.645 (0.189)	0.685 (0.171)	2.947	0.298
PaXJ_C3	5	0.679 (0.256)	0.681 (0.168)	2.864	0.325
China_C4	28	0.675 (0.214)	0.694 (0.200)	3.010	0.537
Mean		0.642 (0.052)	0.675 (0.024)	2.883 (0.128)	0.381 (0.108)
<i>P. armeniaca</i> wild clusters					
PaKGZ_W1	18	0.652 (0.209)	0.695 (0.167)	2.984	0.329
PaNCA_W2	198	0.663 (0.163)	0.724 (0.170)	3.160	0.385
Mean		0.658 (0.008)	0.710 (0.021)	3.072 (0.124)	0.357 (0.040)
Other <i>Armeniaca</i> species					
Psib_nw_W3	22	0.706 (0.200)	0.783 (0.198)	3.510	0.610
Psib_ne_W4	47	0.629 (0.252)	0.738 (0.262)	3.371	1.043
Pmum_C5	8	0.568 (0.229)	0.720 (0.186)	3.097	1.222

Footnotes: N =Number of individuals; A_R and A_P : allelic richness and private allele richness averaged across loci, respectively, estimated by rarefaction using a sample size of six; H_O : observed heterozygosity; H_E : expected heterozygosity. In brackets, standard deviations.

Table I-3. Pairwise F_{ST} (upper diagonal) and Jost's D (lower diagonal) among the nine apricot clusters ($n=416$, 34 SSR markers) inferred with STRUCTURE for $K=11$.

Cluster	PaEUR_C1	PaSCA_C2	PaXJ_C3	China_C4	Pmum_C5	PaKGZ_W1	PaNCA_W2	Psib_nw_W3	Psib_ne_W4
PaEUR_C1	--	0.171	0.208	0.178	0.282	0.224	0.179	0.135	0.188
PaSCA_C2	0.381	--	0.181	0.171	0.238	0.128	0.104	0.125	0.181
PaXJ_C3	0.457	0.468	--	0.183	0.230	0.147	0.116	0.099	0.164
China_C4	0.391	0.436	0.456	--	0.225	0.157	0.133	0.062	0.153
Pmum_C5	0.723	0.705	0.666	0.627	--	0.225	0.209	0.162	0.182
PaKGZ_W1	0.537	0.329	0.376	0.404	0.680	--	0.089	0.100	0.167
PaNCA_W2	0.466	0.286	0.322	0.369	0.667	0.246	--	0.084	0.160
Psib_nw_W3	0.345	0.396	0.320	0.176	0.591	0.319	0.270	--	0.100
Psib_ne_W4	0.484	0.554	0.501	0.441	0.594	0.519	0.517	0.350	--

Footnotes: The pairwise F_{ST} values (below the diagonal) were calculated with GENODIVE. All pairwise F_{ST} and Jost's D values were significant ($P<0.05$, Number of permutations = 1,000). Cluster names are as follows: PaEUR_C1, *Prunus armeniaca* European and Irano-Caucasian cultivars; PaSCA_C2, *P. armeniaca* South Central Asian cultivated apricots; PaXJ_C3, *P. armeniaca* Xinjiang cultivated apricots; China_C4, Chinese landraces; Pmum_C5, *Prunus mume* landraces; PaKGZ_W1, wild *P. armeniaca* from Sary Chelek Kyrgyz sampling site; PaNCA_W2, wild Central Asian *P. armeniaca*; Psib_nw_W3, Chinese North Western *Prunus sibirica*; Psib_ne_W4, Chinese North Eastern *P. sibirica*

CHAPTER II

Genetic diversity and structuration analyses in the French Alpen apricot (*Prunus brigantina* Vill.) reveal phylogenetic inconsistencies in the *Armeniaca* section

Article to be submitted to BioRxiv and later to a peer-reviewed journal

Genetic diversity and structuration analyses in the French Alpen apricot (*Prunus brigantina* Vill.) reveal phylogenetic inconsistencies in the *Armeniaca* section

Liu Shuo^{1,2}, Decroocq Stephane¹, Tricon David¹, Chague Aurelie¹, Decroocq Veronique^{1,‡}

¹ UMR BFP 1332, INRA-Université de Bordeaux, 71 Avenue Edouard Bourlaux, CS20032, 33883 Villenave d'Ornon, France.

² Liaoning Institute of Pomology, Tiedong Street, Xiongyue Town, Bayuquan District, Yingkou City, Liaoning, 115009, China

[‡] Corresponding authors: veronique.decroocq@inra.fr

Abstract

In-depth characterization of crop relatives is a prerequisite for genetic improvement and for fruit germplasm management, in particular when the wild species are endangered. In this study, we sampled *Prunus brigantina* Vill. (or 'Briançon apricot') along with its main natural distribution in the French Alps, where *P. brigantina* populations are severely fragmented, and population sizes are influenced by humans. We examined 71 wild accessions with 34 nuclear markers and evaluated their genetic diversity, population structure to reveal the genetic relationships with other *Armeniaca* and Prunophora species, and to construct a core collection for long-term ex-situ conservation. Our results revealed a low to moderate genetic diversity in *P. brigantina*, with expected heterozygosity H_E of 0.43 and observed heterozygosity H_O of 0.32. A Bayesian model-based clustering approach revealed a weak but significant structuration in three genetic clusters (*Jost's D* = 0.12) consistent with three geographical Alpine regions. No significant gene flow was detected between *P. brigantina* and European *P. armeniaca* cultivars. Meanwhile, we further performed a phylogenetic analysis in which *P. brigantina* grouped apart from the other *Armeniaca* species, together with diploid plum species. Our results revealed a misclassification of *P. brigantina* within the *Armeniaca* section. Based on our analysis, a subset of 32 accessions were selected for ex-situ conservation in a core-collection that encompasses 100 % of the total *P. brigantina* allelic diversity.

Keywords: Apricot, Prunus, classification, genetic structure, core collection

Introduction

Within the genus *Prunus* L. (stone fruit species), the subgenus *Prunophora* includes European and North-American plums (section *Prunus* and *Prunocerasus* Koehne, respectively) and apricots (section *Armeniaca* (Mill.) K. Koch), which are all native from the Northern hemisphere (Rehder, 1940). Within the *Prunophora* subgenus, the section *Armeniaca* comprises only diploid species, ranging from 3 to 12 species depending on the classification system. Six distinct species are usually recognized: *P. armeniaca* L. (common apricot), *P. sibirica* L. (wild apricot in Northeastern Asia), *P. mandshurica* Maxim. (Northeast China and Eastern Russia), *P. mume* (Sieb.) Sieb. & Zucc. (South China and Japan), *P. holosericeae* Batal (South-West China) and *P. brigantina* Vill. (Figure II-1). While the five first species all originate from Asia, ranging from Central to North-East Asia, the last one, *P. brigantina* (synonym *P. brigantiaca*, <http://www.theplantlist.org>), is native from Europe (Villars, 1786). Representatives of this species still grow wild within the Alps between France and Italy in southern Europe, where it is regarded either as apricot or a plum species (Pignatti, 1982, Hagen et al., 2002).

Prunus brigantina, also called Briançon apricot, was first reported in French literature <*Histoire des Plantes de Dauphiné*> by Villars et al (1786). It grows in arid places in shrub thickets in the Alps, above 1,400 m altitude. Like other *Prunus* species, *P. brigantina* is hermaphrodite and is pollinated by insects. It flowers in May and fruits ripen from August to September (Noble et al., 2015; Tison et al., 2014). In natural stands, trees grow 2 to 5 meters high with non-spiny branches and have the heart leaves with double-serrated teeth (Figure II-2a). Full-fledged drupe from *P. brigantina* tree has a small size and appears glabrous with a yellowish fruit skin (Figure II-2a). In the Alps, *P. brigantina* fruits are collected by locals to be processed as jam (Couplan 2009), and their seeds were also used, in the past, for oil production instead of olive or almond (Don, 1834; Dupouy, 1959). It is locally called 'Marmottier' or 'Afatoulier' and is acknowledged as the highest endemic fruit tree in Europe. However, there is currently insufficient information available to evaluate the current population size and its population structure and to determine the potential threats to this species and its *in situ* conservation status (IUCN Red List, Branca et al., 2011).

Previous phylogenetic studies questioned the classification of *P. brigantina* in section *Armeniaca* (Hagen et al., 2002; Reales et al., 2010; Takeda et al., 1998). However, in the former three studies, few samples of *P. brigantina* were tested (one to 2); they were not directly collected *in situ* but instead maintained in germplasm repositories such as the Kew Royal Botanical Garden (UK), the Czech national genetic resources of Lednice, the French Centre de Ressources Génétiques of Montfavet and the Japanese Chiyoda experimental

station. Because of its ability to be propagated by grafting and its interfertility with species from both the section *Prunus* and *Armeniaca*, they could also be replicates or hybrids but this was never tested. Nevertheless, their true origin is unknown. Moreover, sampling only one or two individuals per species is expected to lower the accuracy of phylogenetic analyses (Wiens and Servedio, 1997; Heled and Drummond, 2009). This bias was somewhat suspected in more global diversity and structuration analysis of the *Armeniaca* species (Liu et al., 2019, Chapter I of this manuscript). In consequence, the genetic relationship between *P. brigantina* and other apricot species in the section *Armeniaca* still remains unclear.

Studying the genetic diversity of crop-related species is not only significant for biodiversity conservation but also for the sustainable use of valuable genetic resources through the set-up of *ex-situ* germplasm collections. Developing such a collection of a sufficient number of individuals to be representative of its diversity is space-, time- and money-consuming, especially in perennial species (Brown, 1989; Van Hintum et al., 2000). In order to minimize this cost, the development of a core-collection is recommended, it corresponds to a germplasm diversity panel representative of the global species genetic variability (Frankel, 1984; Glaszmann et al, 2010; Govindaraj et al., 2015). In addition, core collections in woody perennial species have the extra advantages of being able to be propagated vegetatively and maintained for decades, as clones, in field genebanks (Escribano et al., 2008).

To provide useful guidelines for *P. brigantina* conservation, a critical task is obviously to clarify the *P. brigantina* genetic divergence and its taxonomic standing. However, nowadays, neither morphological descriptors nor phylogenetic data are able to provide sufficient insight into the genetic diversity and population structure of this species. In consequence, in the current study, we conducted extensive *in-situ* sampling of *P. brigantina* in its natural habitat and performed its fingerprinting with a *Prunus* set of 34 nuclear markers (Liu et al., 2019. Chapter I of this manuscript). The primary goal of our study was to evaluate the extent of *P. brigantina* diversity and population structure. Second, we investigate the relationship between *P. brigantina* and other *Armeniaca* species. For this purpose, we performed a population genetic analysis on members of the section *Armeniaca* and two outgroups species from the section *Prunus*. Finally, we identified a collection of unique genotypes and selected the best subset that allows to define a *P. brigantina* core collection, with maximum maintenance of allelic diversity, which will be further characterized and genotyped for stone fruit crop improvement.

Materials and Methods

In situ P. brigantina sampling and DNA extraction

A total of seventy-one wild *P. brigantina* trees were collected in 2017 from three sampling sites, in southeast France, along the Alpen Mountains (Figure II-2a and Table II-S1). Young leaves and mature fruits from each tree were collected for DNA extraction and seedling growth, respectively. At least one seedling from each sampled tree was selected and transferred to a collection. Leaf samples from the section *Prunus* were added to the study as follows: wild *P. salicina* ($n=5$, China) and *P. cerasifera* ($n=1$, French Alps). Genomic DNA was extracted as described in Decroocq et al. (2016).

Microsatellite markers and polymerase chain reaction (PCR) amplification

We used 34 microsatellite loci distributed across the eight *P. armeniaca* chromosomes (see details in Liu et al, 2019 Chapter I of this manuscript). PCR amplification and fragment size genotyping were performed on an ABI PRISM 3730 (Applied Biosystems) as described previously (Decroocq et al., 2016). Alleles were scored with the GENEMAPPER 4.0 software (Applied Biosystems). Those 34 loci were selected based on their amplification and polymorphism within the different species of the section *Armeniaca* (Liu et al, 2019, Chapter I). Detailed information on these microsatellite markers, including their repeat motifs, sequences, as well as conditions of amplification are available in Liu et al (2019, Chapter I).

Analyses of population subdivision and genetic relationship

We identified population subdivision with the STRUCTURE software v. 2.3.3 (Pritchard et al., 2000), without the use of *a priori* grouping information and assuming individuals had mixed ancestry with correlated allele frequencies among populations. The clustering method implemented in STRUCTURE is based on Monte Carlo Markov Chain (MCMC) simulations and is used to infer the proportion of ancestry of genotypes in K distinct clusters. We simulated K values ranging from 2 to 10 for the *P. brigantina* population and three additional datasets (Table II-1). Those datasets were obtained in the same conditions as described here but on *Armeniaca* species originating from Central and Eastern Asia (Liu et al., 2019, Chapter I). For each K , we ran 10,000 generations of 'burn-in' and 100,000 MCMC. Simulations were repeated 10 times for each K value; the resulting matrices of estimated cluster membership coefficients (Q) were permuted with CLUMPP (Jakobsson and Rosenberg, 2007). STRUCTURE barplots were displayed with DISTRUCT 1.1 (Rosenberg, 2004). The strongest level of the genetic subdivision was determined using ΔK (Evanno et al., 2005), as implemented in the online post-processing software Structure Harvester

(<http://taylor0.biology.ucla.edu/structureHarvester/>) (Earl and vonHoldt, 2012). Besides, principal components analysis (PCA) was also used to investigate the structure of *P. brigantina* in scatterplot3d R package (Ligges and Miachler, 2003). Further genetic differentiation and relationships were estimated using a weighted neighbour-joining tree as implemented in the DARwin software package v6.0.017 (Perrier and Jacquemoud-Collet, 2006).

We performed a two-step structuration analysis, the first one started with only *P. brigantina* samples collected *in situ* in the French Alps implemented in both STRUCTURE and PCA, and the second one by merging *P. brigantina* SSR fingerprinting with former *Armeniaca* dataset (Liu et al., 2019, Chapter I). This extra set includes SSR fingerprinting of *P. armeniaca*, *P. sibirica*, *P. mandshurica* and *P. mume* wild and cultivated samples (Liu et al., 2019, Chapter I). We also added in the second analysis few samples of *Prunus salicina* (Japanese plum) and *Prunus cerasifera* (myrobolan, cherry plum), both being diploid plum species.

Genetic diversity, differentiation and core collection constitution

We used GENALEX 6.501 (Peakall and Smouse, 2006) to estimate the marker variants using the number of alleles (N_a) and the effective alleles (N_e), and calculate the observed heterozygosity (H_o), the unbiased expected heterozygosity (H_E) and the Shannon index (I) (Shannon, 1948) to evaluate genetic diversity. Genetic differences (*Jost'* D) among genetic clusters were estimated in Genodive (Meirmans et al., 2004).

Estimations for the core collection were performed with the COREFINDER program based on molecular data (Cipriani et al., 2010). The maximization (M) strategy (Schoen and Brown, 1993) implemented in the software COREFINDER was used to generate a core *P. brigantina* tree collection that maximized the number of observed alleles in our dataset. The M-strategy consisted of detecting the sample size that best captured 100% of the genetic diversity present within the entire germplasm collection. We further used the Mann-Whitney U test to check the genetic diversity difference between the core collection and the entire *P. brigantina* population.

Testing isolation by distance

We additionally implemented a Mantel test between a matrix of Edwards' distances and a matrix of Euclidean geographic distances of *P. brigantina* genetic clusters through the R adegenet package (Jombart et al., 2011).

Results and discussion

Genetic variability and structuration in *P. brigantina* population

Thirty-four introduced microsatellites performed well in the former study (Liu et al., 2019) and were thus selected to the *P. brigantina* population study. After allele scoring, four markers (AMPA109, ssr02iso4G, BPPCT008 and BPPCT038) failed to amplify or generated over 50% of missing data; they were consequently eliminated from the study. Additional six markers of BPPCT030, CPPCT022, CPSCT004, UDP98-412, UDA-002 and PacB26 performed poor amplification in *P. brigantina*, by accounting more than 10% missing data. Most of the above microsatellites were developed from *Prunus* species such as peach, almond, apricot and Japanese plum genomic data. This might reflect a poor transferability of *Prunus* markers to *P. brigantina*, it was at least more limited than for apricot and other *Armeniaca* species (Liu et al., 2019, Chapter I). The remaining 24 microsatellite markers that performed well in *P. brigantina* (Table II-S2) were used to estimate the genetic variability at the species level. Fingerprinting allowed producing a total of 121 different alleles (N_A) in our *P. brigantina* sampling and 59.57 effective alleles (N_E), with a mean value of 5.04 and 2.48 respectively (Table II-S2).

The most relevant genetic clustering of *P. brigantina* was strongly suggested at $K=3$ (Figure II-S1). Three inferred genetic clusters (blue, yellow and orange colours, in Figure II-2b) correspond to three French national parks: “Queyras”, “Ecrins” and “Mercantour”, respectively (Figures II-2b and II-S2). Weak genetic differentiation (mean *Jost’s D*=0.117) was further found among three *P. brigantina* populations (Table II-2). Both the *Jost’s D* and PCA indicated that *P. brigantina* populations from “Queyras” cluster, located on the positive PC1-axes, was apart from both that (on the negative PC1-axes) from “Ecrins” (*Jost’s D*=0.12) and “Mercantour” (*Jost’s D*=0.14) clusters, and a close genetic relationship (*Jost’s D*=0.097) was found between “Ecrins” and “Mercantour” clusters (Table II-2, Figure II-3).

Additional population (dark blue in Figure II-S3) was found at $K=8$, indicating a sub structuration in “the Ecrins” region. Genetic admixture within the accessions from the “Mercantour” region appeared without clear sub structuration since $K=4$ (Figure II-S2, Figure II-3). Such admixture might depict past and current genetic exchange through gene flow with other inter-fertile *Prunus* species sharing the same habitat (*i.e.* cultivated apricot *P. armeniaca* and plum, *P. salicina* as well as the wild myrobolan, *P. cerasifera*). Monte-Carlo test on three *P. brigantina* clusters indicated no significant relationship between genetic differentiation and geographic distance ($P=0.308$, Figure II-S3a), while two existing

discontinuity clouds indicate genetic differentiation instead of isolation by distance (Figure II-S3b).

Genetic relationship between *P. brigantina* and related *Armeniaca* species

To obtain a better understanding of the classification of *P. brigantina* in the genus *Prunus*, we combined the current *P. brigantina* data with former *Armeniaca* dataset in Liu et al., 2019 (Tables II-1 and II-S1b).

We first speculated the occurrence of introgression between *P. brigantina* and cultivated *P. armeniaca* in European/Irano-caucasian region (Figure II-S4). Two well-differentiated clusters appeared at $K=2$, which corresponded to *P. armeniaca* (in blue in Figure II-S4) and *P. brigantina* (orange) accessions, respectively. The European cultivated apricots showed a high level of admixture from $K=3$ to $K=10$, while the *P. brigantina* population displayed an obvious intraspecific subdivision, especially within the “Mercantour” region. We also speculated that while both species are partly sharing habitats, no significant gene flow was detected in between wild *P. brigantina* and cultivated apricots.

To evaluate the genetic relationship between *P. brigantina* and other *Prunus* species within section *Armeniaca*, we repeated the Bayesian clustering analysis on the second dataset described in Table II-1. We obtained a similar structuration as the one described in Liu et al (2019, Chapter I), except for *P. brigantina* that differentiated clearly in a distinct cluster, starting from $K=3$ (Yellow colour in Figure II-S5). Once again, no introgression with other species of the section *Armeniaca* was observed and no relationship between *P. brigantina* and other species of the section *Armeniaca* was detected (Figure II-S5). Similar results were obtained by Bortiri et al (2001) in which the single sample of *P. brigantina* did not group with the *Armeniaca* clade.

To further explore the classification of *P. brigantina* in the genus *Prunus*, we added in the study representatives of two diploid plum species, *i.e.* Japanese plum (*P. salicina*) and myrobolan (*P. cerasifera*) (third dataset of Table II-1), which also included representatives of wild/cultivated *P. armeniaca*, *P. sibirica*, *P. mandshurica* and *P. mume* species based on previous structure analyses (details in Tables II-1 and II-S1b). At $K=2$, *Armeniaca* and plum species grouped in one single cluster (blue in Figure II-S6) while *P. brigantina* separated (orange cluster in Figure II-S6). This result was supported by the highest value detected in delta K estimation (Figure II-S7). At higher K values, the different *Armeniaca* and plum species differentiated while at $K=10$, three distinct clusters of *P. brigantina* were evidenced (Figure II-S7). Those three *P. brigantina* clusters corresponded

to the ones depicted in Figure II-S3. However, we could not yet find any introgression from other *Prunus* species into *P. brigantina*.

We further explored the genetic differentiation and relationships among *Prunus* samples using an unweighted neighbour-joining tree in which relative bootstrap support values for relationships are displayed above the branches (Figure II-4). In the parsimony analysis, the delimitation of the apricot and plum species as distinct species from *P. brigantina* was well supported (100% bootstrap support), whilst representatives of *P. brigantina* remained closer to the diploid plum species (*P. salicina* and *P. cerasifera*). Analogous results were obtained from the phylogenetic relationship among Eurasian plum species, according to chloroplast DNA sequences (Reales et al., 2010). In this study, *P. brigantina* grouped closer with European *Prunus* species, such as *P. spinosa*, *P. insititia* and *P. domestica*, but separated clearly from *P. armeniaca* (apricot) and *P. salicina* (Japanese plum). The only diploid plum species that grouped close to *P. brigantina* was *P. ramburii* Boiss., a relict, wild species endemic in the southern Spanish mountains. The most widely distributed species in Europe is the tetraploid *P. spinosa* but this poses the question of the relationship between diploid and tetraploid *Prunus* species. Nevertheless, our study provided full evidence that *P. brigantina* does not belong to the section *Armeniaca* and that the discrepancy we observed in a former study, i.e. relatedness of *P. brigantina* and *P. mume*, was due to the low sample size (Liu et al, 2019).

Construction of a *P. brigantina ex-situ* core collection

P. brigantina is still found in a few alpine valleys, between France and northwest Italy. It grows above 1,400m of altitude as single isolated trees (except for the plateau of Nevache, see Table II-S1), in arid places such as shrub thickets. In France, it is confined to the three southeastern departments of Alpes-Maritimes, Alpes-de-Haute-Provence and Hautes-Alpes. Sustainability of *P. brigantina* ecosystems is threatened by habitat fragmentation. This poses the question of long-term conservation of the species since no germplasm accessions of *P. brigantina* are reported by EURISCO to be held in European genebanks. Large field collections of clonal and perennial crops are expensive to maintain which results in the need of the identification of a restricted number of representatives of *P. brigantina* population, for long-term conservation. Core collections are representative subsets of germplasm collections that are developed to improve the efficiency of germplasm evaluation while increasing the probability of finding genes of interest (Simon and Hannan, 1995). In the case of *P. brigantina*, our results from the COREFINDER program indicate that a smaller core collection may be sufficient to capture molecular diversity documented using the twenty-four

SSR loci. Taking advantage with the M strategy implemented in COREFINDER, we propose a core set of 32 individuals (~45% of the whole *P. brigantina* population) that captures 100% of the measured diversity (Figure II-5, Table II-S3). Further pairwise comparison in Mann-Whitney u test showed no significant ($p>0.05$) difference found on I , H_o , H_e and uH_e indexes between the *P. brigantina* entire population ($n=71$) and the core collection ($n=32$) (Tables II-S2, II-S4 and II-S5). It indicates that our mini core collection could be used as an ex-situ germplasm repository that encompasses *P. brigantina* intraspecific diversity. It will serve in the future for *P. brigantina* conservation as well as for stone fruit breeding programs benefiting from *P. brigantina* resilience characteristics, especially in a context of Mediterranean climate changes.

Conclusion

In this study, we evaluated a low level of genetic diversity of natural *P. brigantina* trees and indicated three genetic populations distributed in the Ecrins, Queyras and Mercantour national parks, respectively. We further successfully established a core collection of 32 individuals to represent the entire *P. brigantina* population above. In addition, Bayesian analyses did not support a close relationship between *P. brigantina* and the other *Armeniaca* species, and only revealed a weak kinship to wild diploid plum, within the section *Prunus*. Therefore, we here recommend correcting *P. brigantina* classification, in which it should no longer be considered as an *Armeniaca* species. However, further research is needed to clarify its relationship with other diploid and polyploid, European species of the subgenus *Prunophora*.

Acknowledgements

S.L. is a recipient of a Chinese Scholarship Council PhD grant. We would like to thanks to Tatiana Giraud for the critical reading of the manuscript. Molecular analysis was performed at the GenoToul Get-PlaGe platform at INRA Centre de Toulouse. The authors wish to acknowledge all the people who helped in collecting the samples, in particular, collaborators from Le Plantivore at Château Ville-vieille; Histoire de Confiture, Plampinet, Nevache; C. Gatineau from Cervières and the managers of the Queyras and Mercantour national parks. The official authorization for the survey and sampling of *P. brigantina* genetic resources is registered and accessible through the following link: <https://absch.cbd.int/database/ABSCH-IRCC-FR-246978>.

Data availability

The datasets generated during the current study, *i.e.* the SSR genotyping, are available at the INRA data portal (<https://data.inra.fr/>) where they can be freely retrieved under the link:

Reference

- Akagi, T., Hanada, T., Yaegaki, H., Gradziel, T. M., & Tao, R. (2016). Genome-wide view of genetic diversity reveals paths of selection and cultivar differentiation in peach domestication. *DNA Research*, 23(3), 271-282. doi:10.1093/dnares/dsw014.
- Alfred, R. (1986). Manual of cultivated trees and shrubs hardy in North America: exclusive of the subtropical and warmer temperate regions. *Biosystematics*.
- Bortiri, E., Oh, S.-H., Jiang, J., Baggett, S., Granger, A., Weeks, C., . . . Parfitt, D. E. (2001). Phylogeny and systematics of *Prunus* (Rosaceae) as determined by sequence analysis of ITS and the chloroplast trnL-trnF spacer DNA. *Systematic Botany*, 26(4), 797-808. doi:10.1043/0363-6445-26.4.797.
- Branca, F., & Donnini, D. (2011). *Prunus brigantina*. e.T172164A121228349. .
- Brown, A. H. D. (1989). Core collections: a practical approach to genetic resources management. *Genome*, 31(2), 818-824. doi:10.1139/g89-144.
- Cao, K., Zheng, Z., Wang, L., Liu, X., Zhu, G., Fang, W., . . . Wang, J. (2014). Comparative population genomics reveals the domestication history of the peach, *Prunus persica*, and human influences on perennial fruit crops. *Genome Biology*, 15(7), 415. doi:10.1186/s13059-014-0415-1.
- Cipriani, G., Spadotto, A., Jurman, I., Di Gaspero, G., Crespan, M., Meneghetti, S., . . . Testolin, R. (2010). The SSR-based molecular profile of 1005 grapevine (*Vitis vinifera* L.) accessions uncovers new synonymy and parentages, and reveals a large admixture amongst varieties of different geographic origin. *Theoretical and Applied Genetics*, 121(8), 1569-1585. doi:10.1007/s00122-010-1411-9.
- Couplan, F. (2009). *Le régal végétal: plantes sauvages comestibles: Sang de la Terre*.
- Cuttelod, A., García, N., & Malak, D. A. (2009). The Mediterranean: a biodiversity hotspot under threat. *Wildlife in a Changing World—an analysis of the 2008 IUCN Red List of Threatened Species* (pp. 89).
- de l'Environnement, A. M. (2003). Conservatoire Botanique National Méditerranéen de Porquerolles. *Plantes envahissantes de la région méditerranéenne*. Aix-en-Provence, Agence Régionale Pour l'Environnement Provence-Alpes-Côte d'Azur.
- Decroocq, S., Cornille, A., Tricon, D., Babayeva, S., Chague, A., Eyquard, J.-P., . . . Decroocq, V. (2016). New insights into the history of domesticated and wild apricots and its contribution to Plum pox virus resistance. *Molecular Ecology*, 25(19), 4712-4729. doi:10.1111/mec.13772.
- Devoghalaere, F., Doucen, T., Guitton, B., Keeling, J., Payne, W., Ling, T. J., . . . Dayatilake, G. (2012). A genomics approach to understanding the role of auxin in apple (*Malus x domestica*) fruit size control. *BMC Plant Biology*, 12(1), 7. doi:10.1186/1471-2229-12-7.
- Don, G. (1834). *A General System of Gardening and Botany: Containing a Complete Enumeration and Description of All Plants Hitherto Known with Their Generic and Specific Characters, Places of Growth, Time of Flowering, Mode of Culture and Their Uses in Medicine and Domestic Economy : Preceded by Introductions to the Linnaean and Natural Systems and a Glossary of the Terms Used ; Founded Upon Miller's Gardener's Dictionary and Arranged According to the Natural System ; in Four Volumes*: Rivington.
- Duan, N. B., Bai, Y., Sun, H. H., Wang, N., Ma, Y. M., Li, M. J., . . . Chen, X. S. (2017). Genome re-sequencing reveals the history of apple and supports a two-stage model for fruit enlargement. *Nature Communications*, 8. doi:10.1038/s41467-017-00336-7.
- Dupouy, J. (1959). Le prunier de Briançon ou Marmottier (*Prunus brigantiaca* Vill.) et les huiles de marmotte. *Bull. SAJA*, 141, 2.
- Earl, D. A., & Vonholdt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, 4(2), 359-361. doi:10.1007/s12686-011-9548-7.

- Escribano, P., Viruel, M. A., & Hormaza, J. I. (2008). Comparison of different methods to construct a core germplasm collection in woody perennial species with simple sequence repeat markers. A case study in cherimoya (*Annona cherimola*, Annonaceae), an underutilised subtropical fruit tree species. *Annals of Applied Biology*, 153(1), 25-32. doi:10.1111/j.1744-7348.2008.00232.x.
- 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. doi:10.1111/j.1365-294X.2005.02553.x.
- Falush, D., Stephens, M., & Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics*, 164(4), 1567-1587.
- Falush, D., Stephens, M., & Pritchard, J. K. (2007). Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular Ecology Notes*, 7(4), 574-578. doi:10.1111/j.1471-8286.2007.01758.x.
- Frankel, O. H. (1984). Plant genetic resources today: a critical appraisal. *Crop Genetic Resources: Conservation and Evaluation*, 241-256.
- Glaszmann, J. C., Kilian, B., Upadhyaya, H. D., & Varshney, R. K. (2010). Accessing genetic diversity for crop improvement. *Current Opinion in Plant Biology*, 13(2), 167-173. doi:10.1016/j.pbi.2010.01.004
- Govindaraj, M., Vetriventhan, M., & Srinivasan, M. (2015). Importance of genetic diversity assessment in crop plants and its recent advances: an overview of its analytical perspectives. *Genetics Research International*, 2015, 14. doi:10.1155/2015/431487.
- Hagen, L., Khadari, B., Lambert, P., & Audergon, J.-M. (2002). Genetic diversity in apricot revealed by AFLP markers: species and cultivar comparisons. *Theoretical and Applied Genetics*, 105(2-3), 298-305. doi:10.1007/s00122-002-0910-8.
- Hazzouri, K. M., Flowers, J. M., Visser, H. J., Khierallah, H. S., Rosas, U., Pham, G. M., . . . Masmoudi, K. (2015). Whole genome re-sequencing of date palms yields insights into diversification of a fruit tree crop. *Nature Communications*, 6, 8824. doi:10.1038/ncomms9824.
- Heled, J., & Drummond, A. J. (2009). Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution*, 27(3), 570-580.
- Hintum, T., L., T. J., Brown, A., Spillane, C., & Hodgkin, T. (2000). *Core collections of plant genetic resources* (Vol. 3): Bioversity International.
- Ingvarsson, P. K. (2005). Molecular population genetics of herbivore-induced protease inhibitor genes in European Aspen (*Populus tremula* L., Salicaceae). *Molecular Biology and Evolution*, 22(9), 1802-1812. doi:10.1093/molbev/msi171.
- Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, 23(14), 1801-1806. doi:10.1093/bioinformatics/btm233.
- Janick, J., & Moore, J. N. (1975). *Advances in fruit breeding* (Vol. 135): Purdue University Press West Lafayette.
- Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*, 27(21), 3070-3071. doi:10.1093/bioinformatics/btr521.
- Khan, M. A., Olsen, K. M., Sovero, V., Kushad, M. M., & Korban, S. S. (2014). Fruit quality traits have played critical roles in domestication of the apple. *Plant Genome*, 7(3). doi:10.3835/plantgenome2014.04.0018.
- LaBonte, N. R., Zhao, P., & Woeste, K. (2018). Signatures of selection in the genomes of Chinese chestnut (*Castanea mollissima* Blume): the roots of nut tree domestication. *Frontiers in Plant Science*, 9. doi:10.3389/fpls.2018.00810.
- Lavergne, S., Thuiller, W., Molina, J., & Debussche, M. (2005). Environmental and human factors influencing rare plant local occurrence, extinction and persistence: a 115-year study in the Mediterranean region. *Journal of Biogeography*, 32(5), 799-811. doi:10.1111/j.1365-2699.2005.01207.x.

CHAPTER II

- Levi, A., Simmons, A. M., Massey, L., Coffey, J., Wechter, W. P., Jarret, R. L., . . . Reddy, U. K. (2017). Genetic diversity in the desert watermelon *Citrullus colocynthis* and its relationship with *Citrullus* species as determined by high-frequency oligonucleotides-targeting active gene markers. *Journal of the American Society for Horticultural Science*, 142(1), 47-56. doi:10.21273/Jashs03834-16.
- Ligges, U., & Mächler, M. (2002). *Scatterplot3d-an R package for visualizing multivariate data*.
- Medail, F., & Quezel, P. (1999). Biodiversity hotspots in the Mediterranean basin: Setting global conservation priorities. *Conservation Biology*, 13(6), 1510-1513. doi:10.1046/j.1523-1739.1999.98467.x.
- Meirmans, P. G., & Van Tienderen, P. H. (2004). GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes*, 4(4), 792-794. doi:10.1111/j.1471-8286.2004.00770.x.
- Menon, M., Barnes, W. J., & Olson, M. S. (2015). Population genetics of freeze tolerance among natural populations of *Populus balsamifera* across the growing season. *New Phytologist*, 207(3), 710-722. doi:10.1111/nph.13381.
- Noble, V., Van Es, J., Michaud, H., & Garraud, L. (2015). Catalogue de la flore vasculaire de la région Provence-Alpes-Côte d'Azur - Version 2.3. Conservatoires botaniques nationaux alpin et méditerranéen.
- Noble, V., Van Es, J., Michaud, H., & Garraud, L. (2015). *Liste Rouge de la flore vasculaire de Provence-Alpes-Côte d'Azur*. Aix-en-Provence, France: DREAL
- Peakall, R., & Smouse, P. E. (2012). GenAIEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics*, 28(19), 2537-2539. doi:10.1093/bioinformatics/bts460.
- Perrier, X. (2006). DARwin software. <http://darwin.cirad.fr/darwin>.
- Pignatti, S. (1982). Flora d'Italia. Edagricole, Bologna, vol: I-II-III.
- Pignatti, S., Guarino, R., & La Rosa, M. (2017). *Flora d'Italia* (Vol. 1).
- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155(2), 945-959.
- Quézel, P. (1985). Definition of the Mediterranean region and the origin of its flora. *Geobotany*.
- Reales, A., Sargent, D. J., Tobutt, K. R., & Rivera, D. (2010). Phylogenetics of Eurasian plums, *Prunus* L. section *Prunus* (Rosaceae), according to coding and non-coding chloroplast DNA sequences. *Tree genetics & genomes*, 6(1), 37-45. doi:10.1007/s11295-009-0226-9.
- Rosenberg, N. A. (2004). DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes*, 4(1), 137-138. doi:10.1046/j.1471-8286.2003.00566.x.
- Sala, O. E., Stuart Chapin, F., Armesto, J. J., Berlow, E., Bloomfield, J., Dirzo, R., . . . Wall, D. H. (2000). Global biodiversity scenarios for the year 2100. *Science*, 287(5459), 1770-1774. doi:10.1126/science.287.5459.1770.
- Schoen, D. J., & Brown, A. H. (1993). Conservation of allelic richness in wild crop relatives is aided by assessment of genetic markers. *Proceedings of the National Academy of Sciences*, 90(22), 10623-10627. doi:10.1073/pnas.90.22.10623.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(3), 379-423. doi:10.1002/j.1538-7305.1948.tb01338.x.
- Sokolov, I. M. (2000). How does insecticidal control of grasshoppers affect non-target arthropods? *Grasshoppers and Grassland Health*, 73, 181-192. doi:10.1007/978-94-011-4337-0_13.
- Takeda, T., Shimada, T., Nomura, K., Ozaki, T., Haji, T., Yamaguchi, M., & Yoshida, M. (1998). *Classification of apricot varieties by RAPD analysis* (Vol. 67).
- Tison, J.-M., & de Foucault, B. (2014). *Flora gallica: flore de France*: Biotope Mèze.

- Tison, J.-M., Jauzein, P., & Michaud, H. (2014). *Flore de la France méditerranéenne continentale*: Naturalia publications Turriers.
- Toivainen, T., Pyhajarvi, T., Niittyvuopio, A., & Savolainen, O. (2014). A recent local sweep at the PHYA locus in the Northern European Spiterstulen population of *Arabidopsis lyrata*. *Molecular Ecology*, *23*(5), 1040-1052. doi:10.1111/mec.12682.
- van Hintum, T. J. L., Brown, A. H. D., Institute, I. P. G. R., Spillane, C., & Hodkin, T. (2000). *Core collections of plant genetic resources*: IPGRI.
- Vié, J. C., Hilton-Taylor, C., Stuart, S. N., Union, I.-T. W. C., & Commission, I. S. S. (2009). *Wildlife in a changing world: an analysis of the 2008 IUCN Red List of Threatened Species*: IUCN.
- Villars, D., Chovin, J. A., & Robert, J. (1789). *Histoire des plantes de Dauphiné: contenant une préface historique, un dictionnaire des termes de botanique, les classes, les familles, les genres, & les herborisations des environs de Grenoble, de la Grande Chartreuse, de Briançon, de Gap & de Montelimar*: Chez l'Auteur & Chez les Libraires.
- Wang, J., Ding, J. H., Tan, B. Y., Robinson, K. M., Michelson, I. H., Johansson, A., . . . Ingvarsson, P. K. (2018). A major locus controls local adaptation and adaptive life history variation in a perennial plant. *Genome Biology*, *19*. doi:10.1186/s13059-018-1444-y.
- Wiens, J. J., & Servedio, M. R. (1997). Accuracy of phylogenetic analysis including and excluding polymorphic characters. *Systematic Biology*, *46*(2), 332-345. doi:10.1093/sysbio/46.2.332.
- Wu, J., Wang, Y., Xu, J., Korban, S. S., Fei, Z., Tao, S., . . . Zhang, S. (2018). Diversification and independent domestication of Asian and European pears. *Genome Biology*, *19*(1), 77. doi:10.1186/s13059-018-1452-y.
- Yao, J. L., Xu, J., Cornille, A., Tomes, S., Karunairetnam, S., Luo, Z. W., . . . Gleave, A. P. (2015). A microRNA allele that emerged prior to apple domestication may underlie fruit size evolution. *Plant Journal*, *84*(2), 417-427. doi:10.1111/tbj.13021.
- Zhou, Y. F., Massonnet, M., Sanjak, J. S., Cantu, D., & Gaut, B. S. (2017). Evolutionary genomics of grape (*Vitis vinifera* ssp *vinifera*) domestication. *Proceedings of the National Academy of Sciences*, *114*(44), 11715-11720. doi:10.1073/pnas.1709257114.

Figures

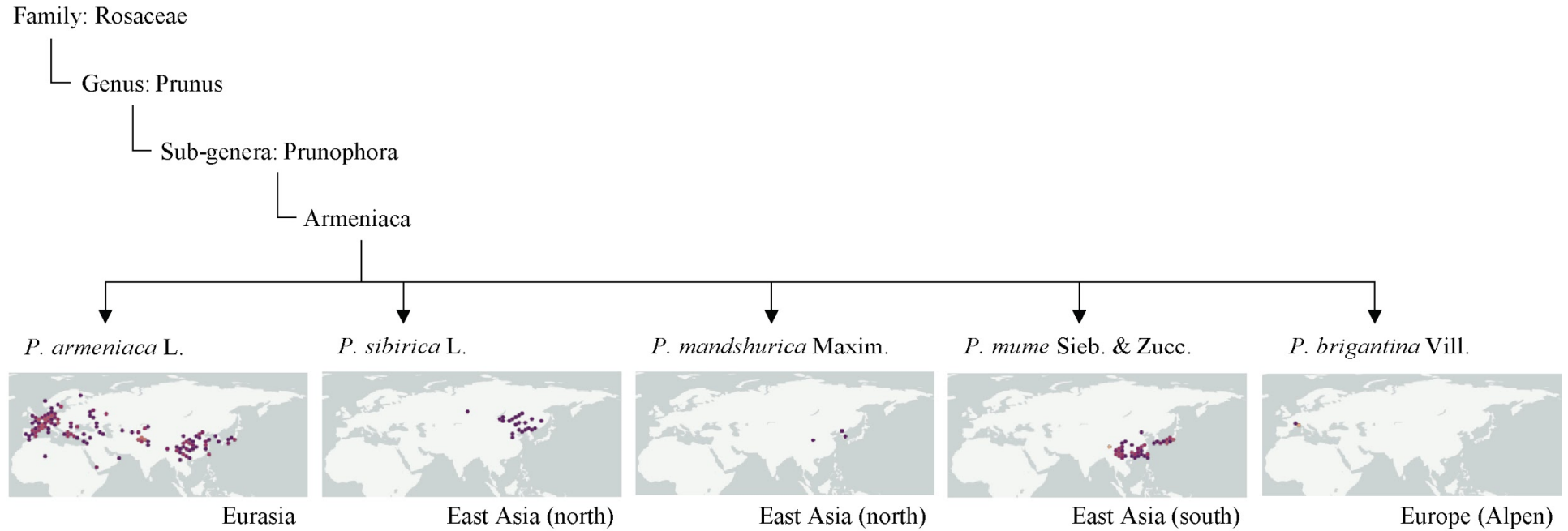


Figure II-1. Taxonomy and geographic distribution of the different species of section *Armeniaca*. Species classification is based on reports by Bailey (1916), Rehder (1940) and Ledbetter (2008). Data on species distribution data were retrieved from the Global Biodiversity Information Facility (GBIF) (<https://doi.org/10.15468/39omei>). Small triangles represent georeferenced species records from the year 1910 to 2017.

a.



b.

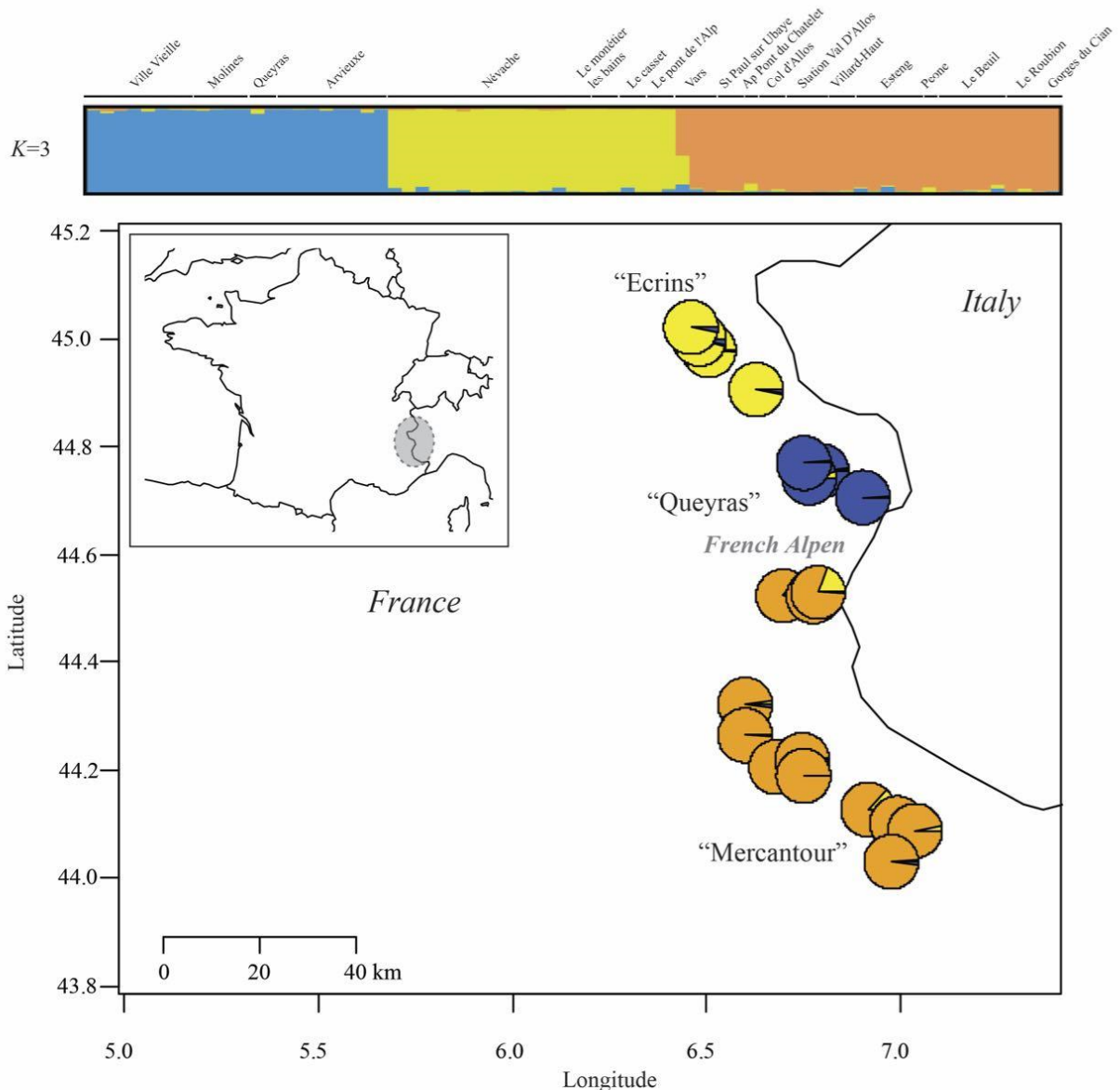


Figure II-2. *Prunus brigantina* morphological features, genetic clustering and spatial distribution over the French Alps. a. *P. brigantina* small tree in its natural habitat (Arviex) (left), aestival leaves (middle) and ripening fruits (right). b. *P. brigantina* three genetic clusters inferred from STRUCTURE analysis (Figure II-S3 at $K=3$) and their spatial distribution in French Alpen mountains. "Ecrins", "Queyras" and "Mercantour" refer to the three national parks in the southeast of France.

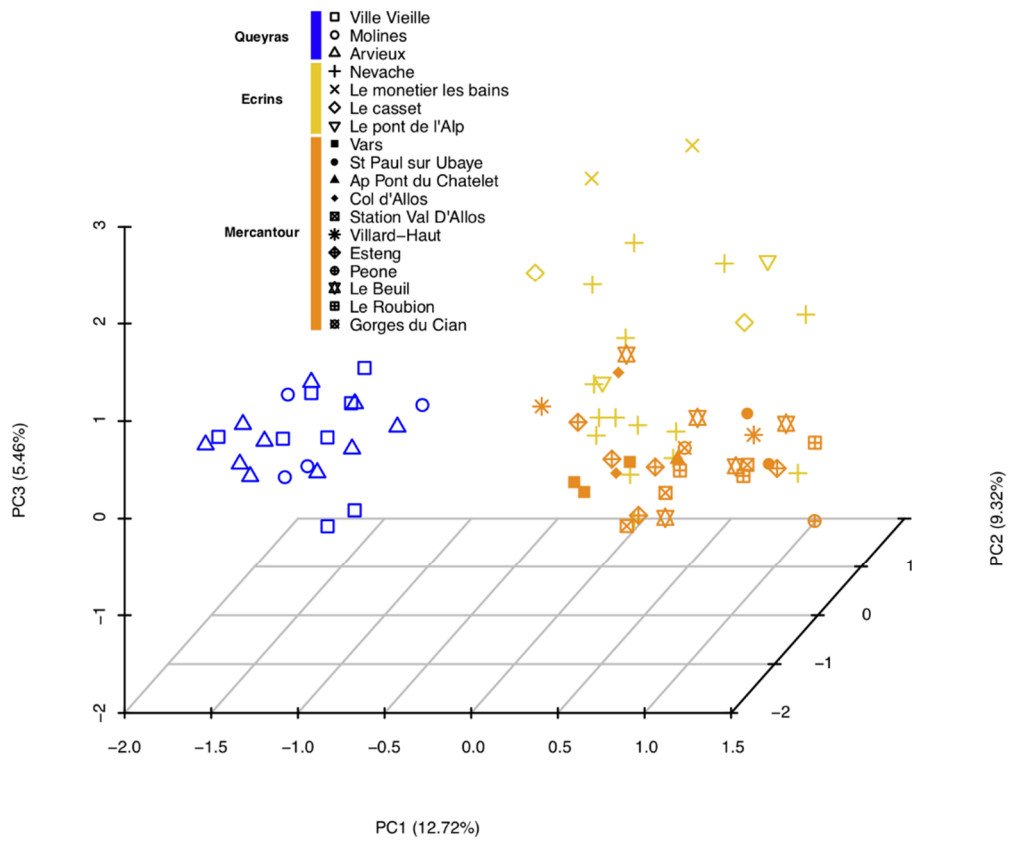


Figure II-3. Principal components analysis on *P. brigantina*. colour in Legend referred to genetic clusters inferred from structure analysis, according to barplots at K=3 in Figure II-S2.

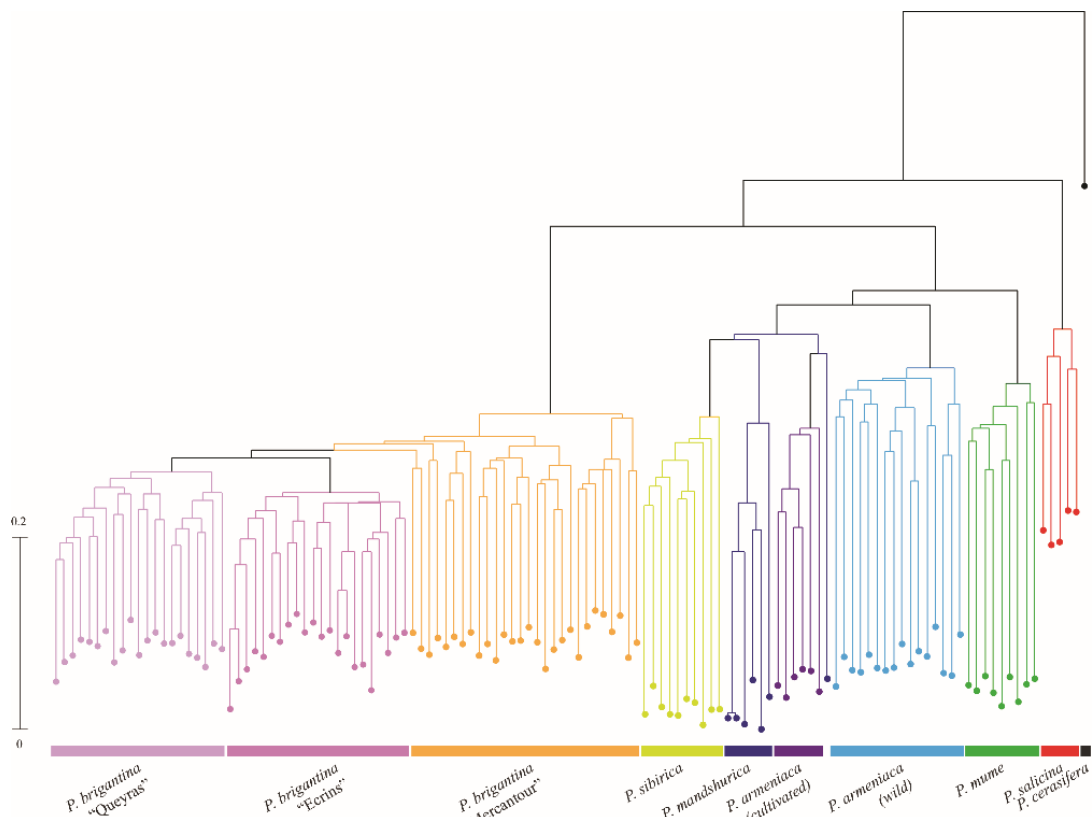


Figure II-4. Weighted neighbour-joining (NJ) tree of *P. brigantina* and other *Prunus* species. Bootstrap scores over 50 are depicted onto each branch. Each species is represented, below the node, by the same colour as the ones used in STRUCTURE (K=10, Figure II-S6).

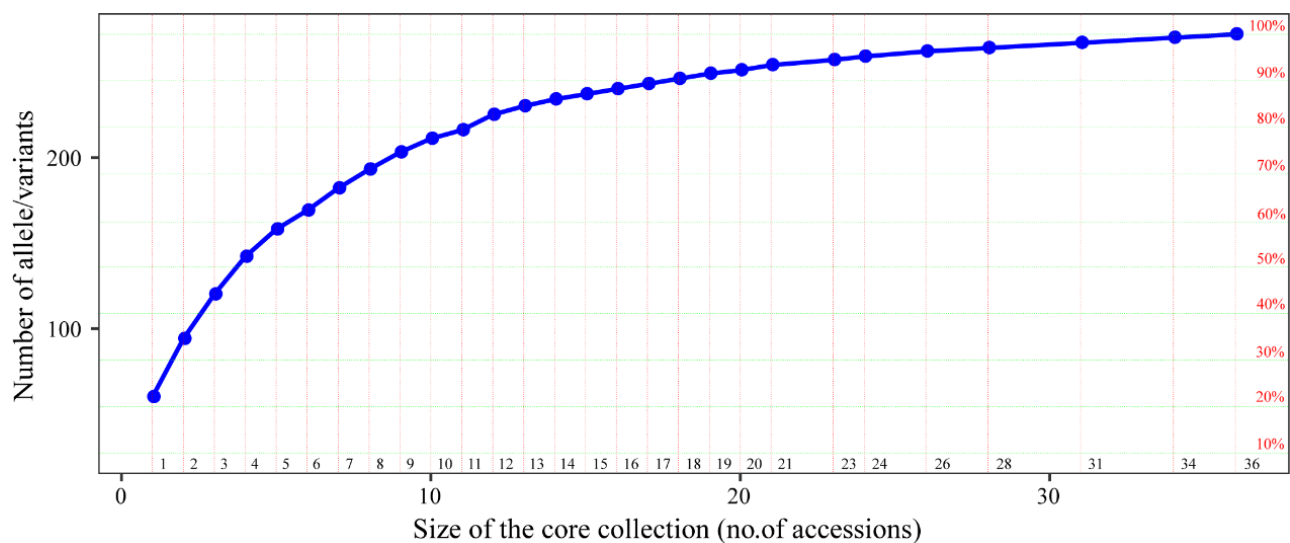


Figure II-5. The core collection of *P. brigantina* population based on the M strategy. The right Y-axis (in %) represents the degrees of genetic diversity for different core collection sizes. Details of accessions for each percentage rate is presented in Table II-S3.

Tables

Table II-1. Different datasets included *P. brigantina* and other apricot species in this study.

Datasets	Description	Number of accessions	Involved <i>Prunus</i> species (n)								
			<i>P. brigantina</i>	<i>P. armeniaca</i> (wild)	<i>P. armeniaca</i> (cultivated)	<i>P. sibirica</i>	<i>P. mume</i>	<i>P. mandshurica</i>	<i>P. salicina</i>	<i>P. cerasifera</i>	
1	<i>P. brigantina</i> and sub dataset of European cultivated <i>P. armeniaca</i>	160	73*		87						
2	<i>P. brigantina</i> and previous apricot collection accessions	648	73*	204	270	84	9	8			
3	Dataset of seven <i>Prunus</i> species representatives	125	71	16	6	10	9	7	5	1	

* indicate a *P. brigantina* dataset included 71 individuals sampled from Alpens and 2 samples from repository

Table II-2. Pairwise population *Jost's D* of *P. brigantina*.

Population	Queyras	Ecrins	Mercantour
Queyras	-	0.116	0.14
Ecrins		-	0.097
Mercantour			-

CHAPTER III

Distinct evolution and domestication processes in apricot (*Prunus armeniaca* L.) revealed by different patterns of selection

To be part of a more global article on apricot genome assembly

Foreword

In Chapter I, we provided insight into the complex domestication history of apricot, by using 25 ‘perfect’ SSR loci. This study allowed to distinguish the origin of domesticated European apricots from Chinese cultivated apricots. We demonstrated that populations of living wild, Central Asian *P. armeniaca* are descendants of the ancestral population and that they gave rise to domesticated apricots. We also showed that most modern cultivated apricots are heavily admixed with wild *P. armeniaca* or/and between European, Central Asian and Chinese domesticated apricots. However, it lacked the power to disentangle the origin of Chinese cultivated apricots. In the best scenarios, domesticated germplasm from China came either from Central Asian *P. armeniaca* or from a gene pool that arose from an ancient hybridization between Chinese Western *P. sibirica* and Central Asian *P. armeniaca*. Moreover, although European and Chinese apricots originate from two independent domestication events (Chapter I), ***the question remains if those two distinct domestication processes produced the same effect on the apricot crop genome.***

Plant domestication involves significant phenotypic changes driven by strong artificial selection and often results in new populations/species established by humans (Zeder, 2015). Indeed, the process of domestication accompanied by the selection of traits related to yield, morphology, fertility..., is believed to dramatically affect the frequency of alleles segregating among domesticated plants. Mutations conferring phenotypes favoured during domestication will be subject to a ‘selective sweep’ with a rapid increase in allele frequency by artificial selection. Therefore, to access a comprehensive analysis of genetic variations underlying domestication in both European and Chinese cultivated apricots, we focused in Chapter III on investigating distinct or similar patterns in genes under artificial selection. For this, we benefited from a substantial whole-genome resequencing data generated through three successive projects: the ANR CHEX “ABRIWG”, the “ABXING” G2P from Bordeaux University and the current “SWAG” France Génomique project. Those projects allowed (1) to assemble *de novo* the *P. armeniaca* genome, (2) to produce a high quality reference sequence, and (3) to sequence, by ILLUMINA at 15X depth, 126 European and US & Canadian cultivars, 27 Chinese local varieties and 75 wild Central Asian *P. armeniaca* accessions. This set of genomic data was used in Chapter III ***to identify genomic regions/genes under selection during the process of domestication of European and Chinese apricots.*** We also used population genomic data ***to unravel loci and candidate genes under selection as a signature of molecular adaptation among the wild, Central Asian apricots.***

Introduction on selective sweeps

Positive selection occurs when an allele is favoured by natural or artificial selection. The frequency of the favoured allele increases in the population, at high prevalence up to fixation, and due to hitchhiking the surrounding linked variation decreases, creating the so-called ***selective sweeps***. Indeed, when a beneficial allele and the neighbouring variants on the same haplotype reach high prevalence together, it produces a population-wide reduction in genetic diversity, summarized as a decrease in heterozygosity, polymorphism or variability (Pavlidis and Alachiotis, 2017). In consequence, a selective sweep is a reduction or elimination of variation among the nucleotides near a favoured allele in DNA (Figure III-1). Detecting traces of positive selection in genomes is achieved by searching for signatures introduced by selective sweeps.

Within the selective sweeps, we distinguish a complete selective sweep that originates from a new beneficial variant (also called hard selective sweep) from the soft selective sweep (Figure III-1). In the last case, positive selection acts on variation already segregating in the population (*i.e.* standing variation) or multiple beneficial alleles arise independently (*i.e.* the standing variant appears on multiple, distinct haplotype backgrounds as a result of recurrent mutation or migration).

Choosing an Adequate test for the identification of selective sweeps

A genomic region associated with the selective sweep of a favoured allele is a region of the genome that has lower than expected levels of nucleotide polymorphism in a population sample. This is true until recombination and mutation restore diversity to the population at the selected locus. On a micro-evolutionary scale, methods focusing on three measurements are commonly used: (1) linkage disequilibrium (LD); (2) site frequency spectrum (SFS); and (3) population differentiation-based tests (Table III-1). SFS-based methods (Tajima's D, Fu and Li's tests, Fay and Wu's H, CLR, XP-CLR, Pool-HMM...) rely on the assumption that selective sweeps affect the frequency of variants in a predictable manner, meaning an increased proportion of low- and high-frequency variants and a reduced proportion of intermediate-frequency variants. Most of the LD-based methods (LRH, iHS, XP-EHH, Rsb, H12, ω statistic, HapFLK...) focus on long homozygous regions with high frequencies of certain haplotypes generated by hard sweeps (Sabeti et al., 2002; Garud et al., 2015). Population differentiation-based tests (FDist, BayeScan, FLK...) assume that populations occur in different environments and thus encounter different selective regimes.

The decision to apply a specific method and test category to detect positive selection in a data set depends on several criteria. The specific time at which a selective sweep occurred is important, because the different characteristics analyzed in the tests show distinct detection rates for younger and older events (reviewed by Biswas & Akey, 2006; Hohenlohe et al., 2010). For example, the methods based on SFS comparing different estimates of θ show the highest statistical power when the frequency of the beneficial allele is approaching fixation in the analyzed population. Linkage disequilibrium-based methods have optimal detection rates in a range from low beneficial allele frequency up to close to fixation while tests comparing population differentiation have higher power close to fixation of the beneficial allele (Weigand et al., 2018).

Advantages and disadvantages of each test

For the SFS based methods (Tajima's D and Fu and Li's tests) a high power was found for high beneficial allele frequencies until sometime after fixation, with the highest power shortly before or at fixation (Fu, 1997; Zeng et al., 2006; Zhai et al., 2009; Ronen et al., 2013). In contrast, the power of the CLR statistic was highest shortly before but not ultimately at fixation (Ronen et al., 2013; Ma et al., 2015). Moreover, Tajima's D and Fu and Li's tests are used to detect not only signatures of selection but also demographic changes. Both Tajima's D and Fu and Li's tests were found to detect expansions similar to selection (Fu, 1997; Zeng et al., 2006; Ferrer-Admetlla et al., 2014; Wang et al., 2014), and are impacted by bottlenecks (Wang et al., 2014; Alachiotis & Pavlidis, 2016). However, the CLR statistic was found to be unaffected by expansions (Boitard et al., 2009) and was only slightly affected by weak bottlenecks (Pavlidis et al., 2010).

For LD based methods, the ω statistic was found to perform well for weak bottlenecks but is affected by strong ones (Alachiotis & Pavlidis, 2016). For iHS and XP-EHH, high migration rates led to a poor performance (Vatsiou et al., 2016). In contrast, the FPR of the ω statistic decreased with increasing migration, thus leading to lower differentiation between subpopulations, while the effect of migration on the power of the ω statistic was not tested (Jensen et al., 2007). HapFLK performed at high power in a wide range from intermediate to high beneficial allele frequencies, with a drop in statistical power close to fixation of the beneficial allele (Vatsiou et al., 2016). Simulations including refugia populations led to a high level of false positives ($> 15\%$) for BayeScan, whereas they had only minor effects on FDist and the FLK test (Lotterhos & Whitlock, 2014). HapFLK performs well in detecting hard selective sweeps, whereas its power decreases with an increased initial allele frequency and in the case of soft selective sweeps (Fariello et al., 2013; Vatsiou et al., 2016).

Methods used for selective sweep detection in this study

According to the specificities of our plant material and genomic data (genome-wide SNPs scanning), we selected a few summary statistics from three measurements commonly used in selective sweep detection. They are composed by three basic statistics like π , Tajima's D, FST and three well-implemented tools like CLR, ω -statistic, hapFLK and one newly developed method, μ statistic.

The CLR algorithm was substantially extended by Pavlidis et al. (2013; SweeD). SweeD can provide the option to employ a user-specified demographic model for the theoretical calculation of the expected neutral SFS following the theory of Živković & Stephan (2011), rather than being estimated empirically over the complete data set, if the demographic model is known with confidence. It can also alter the mathematical operations for the CLR test implementation to avoid numerical instability (floating-point underflows), allowing the analysis of datasets with thousands of sequences.

The ω -statistic computes the specific LD pattern of a sweep and accurately detects the targets of positive selection by using demographic parameters relevant to natural non-equilibrium population, such as the cosmopolitan population of *D. melanogaster* (Jensen et al. 2007). It allows for the analysis of large whole-genome data sets in OmegaPlus software that was released by Alachiotis (2012). It was indicated as preferred statistic to detect robust signatures of selective sweeps (Jacobs et al. 2016) and performed best if the beneficial allele is close to or at fixation (Sabeti et al. 2007; Fariello et al. 2013; Ronen et al. 2013; Ma et al. 2015).

The hapFLk statistic builds upon the original FLK statistic (Bonhomme et al. 2010). As FLK, it incorporates hierarchical structure of populations, but the test is extended to account for the haplotype structure in the sample. For this, it uses a multipoint linkage disequilibrium model (Scheet and Stephens 2006) that regroups individual chromosomes into local haplotype clusters. The principle is to exploit this clustering model to compute "haplotype frequencies," which are then used to measure differentiation between populations. It has been shown to be robust with respect to bottlenecks and migration. (<https://github.com/bcm-uga/SSMPG2017/blob/master/Presentations/hapflk/hapflk.org>).

The μ statistic serves as a measure of positive selection by assuming high values in regions where variation resembles the signatures that a sweep leaves in a genome (Alachiotis et al. 2018). It is a composite evaluation test that scores genomic regions by quantifying changes in the SFS, the level of LD, and the amount of genetic diversity along a chromosome. This detection is implemented in the RAiSD software to scan whole-genome

SNP data based on a composite evaluation scheme that captures multiple sweep signatures at once (<https://github.com/alachins/raisd>).

All the methods described above were developed to detect hard selective sweeps. The detection of soft selective sweeps is more challenging because they generally do not affect linked neutral polymorphism to the same extent. Moreover, positive selection on standing variation where the standing variant has time to recombine and associate with different haplotype backgrounds is notably more difficult to detect. In this case, SFS based methods have no predictive power owing to the fact that soft sweeps may involve an arbitrary number of distinct haplotypes but LD-based methods work albeit lower power.

The role of demography in selective sweep detection accuracy

High differentiation at a locus compared to other loci could be due to positive selection, although demographic processes could also explain such differentiation. Indeed, migration, expansions or bottlenecks (population contractions) can often create selection-like signals. One solution would be to compare locus-specific data to genome-wide data, as neutral demographic changes generate genome-wide patterns, whereas selection acts in a more targeted manner (Nielsen et al., 2005). In SFS-based sweep scans, this is implemented in a two-step computational approach, with an average, genome-wide SFS (background SFS) followed by a detection step for the genomic regions that fit the selection model but not the background SFS. However, this assumes a uniform behaviour of the SFS along the genome. This is not the case for demographic models such as bottlenecks which generate a great variance along the recombining chromosomes (Pavlidis et al., 2008). In consequence, several tools such as SweepFinder, SweepFinder2, SweeD and OmegaPlus which integrated this information and the use of these tools or, at minima, of the demographic model as the null model allow to alleviate (without completely overcoming) the problem of confounding the effect of demography with selection.

Material and Methods

(The schematic steps for detection of genomic regions under selection in our study were summarized in Figure III-2)

Sample preparation and sequencing

A total of 228 apricots (*Prunus armeniaca*) accessions were included in this study, of which 99 European and 27 North American apricots (mainly from the USA and Canada) coming from the French national repository (UGAFL-INRA, France) or from the Yalta botanical garden (Former Ukraine). It also included 75 wild apricot trees sampled in Central Asia (including Kazakhstan, Kyrgyzstan and Uzbekistan) as described in Decroocq et al (2016),

and 27 Chinese cultivars/landraces from the Chinese and French National Repositories (Liaoning Institute of Pomology and UGAFL-INRA, respectively). Details about the apricot accessions used in this study are given in the Supplementary Table III-S1. Genomic DNA was extracted from young leaves as described in Mariette et al. (2016). Libraries for paired-end sequencing were constructed with an insert size of over 300 bp according to the Illumina library preparation protocols. Libraries are then sequenced using the Illumina HiSeq 2000/2500 platforms to generate 100bp paired reads (GATC Inc. and the David H. Murdock Institute, Kannapolis, NC) and the HiSeq 3000 platform for 150bp paired reads (INRA-GeT-PlaGe) (Table III-S1). All accessions were sequenced to an estimated depth of 15X.

Raw read trimming, alignment, and filtering

Paired raw sequencing reads are trimmed using cutadapt (v1.2.1) (Martin, 2011) to remove the adapter sequences, and using NGS QC-toolkit (v2.3.3) (Patel and Jain, 2012) to remove the bases when average quality per base dropped below a score of 20. Reports of sequence quality control were generated for each apricot accession. Trimmed reads were mapped to the diploid *Prunus armeniaca* Marouch reference genome version 2 (Supplementary note 1) with modified parameters implemented in bwa (v0.7.17) (Li and Durbin, 2009) using the BWA-MEM algorithm. To account for the occurrence of PCR duplicates introduced during library construction, we used MarkDuplicates in software picard-tools (v2.9.2) (Picard Toolkit, 2019, Broad Institute, GitHub Repository. <http://broadinstitute.github.io/picard/>; Broad Institute) to remove reads with identical external coordinates and insert lengths.

SNP variant calling and filtering

We used HaplotypeCaller and GenotypeGVCFs implemented in GATK (v3.8) (McKenna et al., 2010) to get a set of preliminary SNP and genotype calling across all samples. Due to the absence of reference data for the GATK analysis steps of “Indel Realignment” and “Base Quality Score Recalibration (BQSR)”, we used a “self-training” strategy of performing three rounds of variants calling by GATK. Briefly, at first, we implemented an initial round of SNP calling on the original uncalibrated data. Then we used “Variant Filtration” function with default parameters in GATK to filter the SNPs. These SNPs with the highest confidence are then used as the database of known SNPs by feeding it as a variant call format (VCF) file to the base quality score recalibrator. This step is repeated in the second round calling analysis. Finally, we validated a real round of SNP calling with the recalibrated data. After three round variants calling, a final VCF file without obvious difference was obtained.

The preliminary SNPs variants from GATK were further filtered by a modified pipeline following the “SNP Filtering Tutorial” (Puritz et al., 2014) displayed on the website

(<http://ddocent.com/filtering/>) in five steps to reduce false positives: 1) we only keep bi-allelic SNPs used in this study; 2) we focus on eight well-assembled chromosomes and abandoned those SNPs from two additional scaffolds (Super-Scaffold_90 and Super-Scaffold_99); 3) we select the SNPs with a minor allele count of 3 (--mac = 3) and a minimum quality score of 30 (--minQ = 30) overall samples; 4) we remove SNPs with more than 85% missing genotypes across all samples (--max-missing 0.85); 5) we include the SNPs genotypes with a minor allele frequency over 0.5% (--maf = 0.005). Quality filter data in VCF files were further visualized in R package “vcfR” (Knaus et al., 2017).

Linkage disequilibrium (LD)

We quantified LD using the squared correlation coefficient (r^2) between pairs of SNPs over a 300 Kb physical distance as implemented in PLINK v1.9 (Chang et al., 2015). An average of 500,000 SNPs was randomly selected from each chromosome. The decayed physical distance between SNPs was identified as the distance at which the maximum r^2 dropped by half (averaged in short range of 10 bp) (Vos et al., 2017) (www.cog-genomics.org/plink/1.9/).

Identification of the apricot genetic groups and the analysis of their structuration

A total of 228 apricot accessions were classified into four groups based on the different sampling sites and germplasm origin, as follows: a wild Central Asian group ($n=75$) and three cultivated groups, one from Europe ($n=99$), one from North-America (United States of America and Canada, $n=27$), and the last from China ($n=27$) (Table III-S1). We distinguished the North-American from the European apricots because the first ones correspond to very recent cultivars issued from less than 100 years old modern breeding programs that used both European and Asian germplasm. They would thus represent a highly admixed group of cultivated apricots, from the first to second generations after hybridization between European and Asian material (from China or Central Asia). Abbreviations for these apricot groups are hereafter CA for Central Asia, EUR for Europe, N-A for North America, and CHN for China.

Population structure was inferred using the program “fastStructure (v1.0)” using two different SNP datasets. The first one includes all filtered SNPs from our initial data in which we removed SNPs with more than 15% of missing genotypes across all samples, a minimum quality score of 30 and a minor allele frequency (MAF) inferior to 0.5% (0.005). The second dataset corresponds to unlinked SNPs, in accordance with the Hardy-Weinberg’s equilibrium. To identify unlinked loci, we applied a threshold value of 0.04 which is well under the decayed r^2 as inferred in the above LD analysis using “bcftools (v1.6)”.

For both datasets, we predefined the number of genetic clusters K from 2 to 8, specified the random seed to 100, and kept all other parameters as default. We then used the “chooseK.py” command to infer the appropriate number of model components that explains structuration in our apricot dataset. The final admixture proportions inferred by *fastStructure* was visualized with POPHELPER v1.0.1 (<http://www.pophelper.com/>).

We further performed principal component analysis (PCA) using the smartPCA tool implemented in eigensoft program version 6.1.4 (<http://data.broadinstitute.org/alkesgroup/EIGENSOFT>) to detect outliers in each apricot group. We ran smartPCA using both the entire and reduced SNP datasets. Pairwise of SNPs in high LD was filtered following three steps: 1) First, we considered a window of 50 SNPs and calculated the LD between each pair of SNPs within the window; 2) second, we removed one of a pair of SNPs if their LD value (r^2) was greater than 0.5; 3) third, we shifted the window 5 SNPs forward and repeated the procedure. We predefined the maximum number of outliers to 5, under a sigma threshold of 6. Three eigenvectors were selected from the output results for each apricot group and were further visualized with the “scatterplot3d” R package (<https://CRAN.R-project.org/package=scatterplot3d>).

Genetic diversity and demographic history

Genome-wide genetic diversity was estimated by calculating pairwise nucleotide differences within populations, P_i (π) (Nei, 1987) and Tajima's D (Tajima, 1989). The genetic differences F_{ST} between wild and cultivated or between cultivated and cultivated were calculated in non-overlapping windows of 10 Kb in size with vcftools (v0.1.16) (Danecek et al., 2011). The demographic history and population size of the wild and cultivated, apricot samples were inferred with SMC++ (Terhorst et al. 2017). Divergence time was estimated between cultivated and wild apricots. We assumed a generation time of 5 years and a mutation rate of $4.46e^{-9}$ mutations per nucleotide per year as previously estimated in *Prunus sibirica* (Wang et al., 2017).

Identification of genomic signatures of selection

To identify the selective signals along the wild and cultivated apricot genomes, we implemented multiple methods of detection as follows:

- 1) OmegaPlus, a high-performance implementation of the ω statistic, to detect increased signals by LD (Alachiotis et al., 2012).

- 2) SweeD, which implements a composite likelihood ratio (CLR) test to detect complete selective sweeps using Site Frequency Spectrum (SFS) patterns of single-nucleotide polymorphisms (SNPs) (Nielsen et al., 2005).
- 3) hapFLK, which focuses on the differences in haplotype frequencies between populations. Based on the hierarchical structure of populations by FLK statistic (Bonhomme et al. 2010), the hapFLK test can be further calculated using multipoint linkage disequilibrium model (Fariello et al., 2013).

The nucleotide diversity (π also called Pi), Tajima's D and F_{ST} values as calculated in the above diversity analyses were also included in selective sweep detection. Besides, we also introduced one recently developed method: RAI_{SD} (Alachiotis et al., 2018). It uses a composite evaluation test of μ (μ)-statistic to score genomic regions under selection by quantifying changes in the SFS, in the level of LD and in the amount of genetic diversity along a chromosome. It also results in three statistics related to the pattern of polymorphism (μ -var), linkage disequilibrium (μ -ld), and site frequency spectrum (μ -sfs) (Alachiotis et al., 2018).

The 10 kb segments with the top 0.5% of maximum values were considered significant and were further compared between the different methods of detection of selective sweeps. Results were represented in circular graphs using the R package “circlize” (Gu, 2014).

Potential traits/genes associated with selective sweeps

For a better understanding of the genes and traits under artificial and natural selection in cultivated and wild samples, respectively, we determined the coding sequences (assembled and/or predicted) colocalising with selective sweeps using information from the *P. armeniaca* Marouch v2.0 genome annotation (<https://services.cbib.u-bordeaux.fr/apollo/annotator>). The function of each candidate gene was verified through BlastX as implemented on NCBI (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) and UniProt (<https://www.uniprot.org/>). Intervals of quantitative trait loci (QTLs) that were already mapped in crop *Prunus* species were retrieved from the SearchQTL function on the GDR database (<https://www.rosaceae.org/>).

Results

Data filtering

The preliminary outputs from GATK before data filtering is summarized in Figure III-3 and Figure III-S1. After data filtering, a total of 6,925,935 SNPs out of 12,432,085 remained

(55%). The number of SNPs was the highest in chromosome 1 (1,500,998), but the lowest for chromosome 5 (578,547), with an average density of 25-30 bp between consecutive SNPs (Table III-2).

Linkage disequilibrium in apricot

We estimated linkage disequilibrium in the four apricot groups as defined in the material and methods, *i.e.* cultivated European (EUR), North-American (N-A), Chinese (CHN) and wild Central Asian (CA) (Table III-3). Along the eight chromosomes, the minimum r^2 value ranged from 0.030 (CA) to 0.086 (N-A) with an average of 0.053; the maximum r^2 value indicated a mean of 0.258 ranging from 0.225 (CA) to 0.322 (N-A). Central Asian wild apricots had the lowest minimum r^2 ($p < 0.05$ by Wilcoxon Signed-Rank test), their maximum r^2 was similar to the one from Chinese cultivated apricots but significantly lower than for both European and North-American apricots. Within the cultivated groups, apricots from USA & Canada displayed a significantly higher LD level (both minimum and maximum) than for both European and Chinese groups.

Linkage disequilibrium decayed very quickly in apricot within a few hundred base pairs, over the eight chromosomes (Figures III-S2, S3 and S4). Table III-3 summarizes more properties of LD as a function of physical distance for each apricot group. A previous study found a similar range of LD decay in a set of cultivated European and North-American apricots (Mariette et al. 2016). In our case, knowing that LD is strongly affected by population structure, we chose to estimate the LD decay for each apricot group, separately. It is evident that the average distance over which LD decayed to ~50% of its maximum value differs between the Central Asian wild group (182.5 bp) and the other cultivated groups, especially for CHN (771.25 bp) (Table III-3). Moreover, LD analyses showed that apricot genomes have relatively rapid LD decays, within 1 kb distance, except for the N-A group (8.90 Kb, $p < 0.05$) (Figure III-4).

Inferred genetic admixture and sub-clusters within cultivated and wild apricot groups

A total of 228 apricots were assigned to four groups based on their origins. We investigated their genetic structuration using both the entire SNPs ($n=6,925,935$) data and unlinked SNPs (LD pruned, pairwise $r^2 < 0.05$; $n=442,794$). *FastStructure* runs including or excluding loci with significant linkage produced almost identical results (Figures III-S5 and III-5). The method identified an optimal partition in three genetic pools with a clear geographical and admixture pattern. The "chooseK" function estimated three model components explained on

our data. Nevertheless, the partition was clearer with unlinked SNP data (Figure III-5). We, therefore, considered these three clusters as the most relevant subdivision. In this study, most wild accessions from Central Asia (CA) formed a distinct population (light green at $K=4$ in Figure III-5). We detected two genetic clusters of cultivated apricots across Europe and China (light and dark blue, Figure III-5). We also found footprints of admixture with the light green wild cluster, especially for the Chinese cultivated group. The genetic subdivision among the N-A group confirmed a highly admixed genetic background between European (dark blue) and Chinese (light blue) apricots, with some degree of wild (light green) ancestry, which is not surprising knowing that all US and Canadian cultivars have a recently bred (~100 years) hybrid origin starting with Luther Burbank's work (Ledbetter, 2010).

We additionally used smartPCA (Principal Component Analysis) to investigate population structure among the European and Chinese cultivated and Central Asian wild apricots (Figure III-6). Apricot accessions from the N-A group were not retained at this stage because it clustered in between the Chinese, Central Asian and European groups. In view of the high degree of admixture of US and Canadian cultivars, which might influence the population structure of the other apricots, the population structure analysis for N-A apricots was conducted independently. Once again, we obtained a more discrete distribution of accessions in both cultivated and wild apricot groups using LD pruned SNPs than by using the entire SNP data. Based on the LD pruned data (Figure III-6 B), it was found that Central Asian wild apricots could be classified into two groups, the group on the right corresponding to the red (W1) cluster previously identified in SSR-based structure analysis (Liu et al, 2019, under revision, Chapter I of this manuscript). When combining the three apricot groups in the Principal Component analysis (Figure III-6 C); the three first axes explained 12.65%, 6.02% and 4.31% of the total genetic variance, respectively. PC1 separated wild Central Asian apricots from European and Chinese cultivars. PC2 and PC3 further divided the cultivated apricots in between Chinese and European origins. Additionally, when using the outlier function implemented in smartPCA, the European 'Zard' accession was found as wrongly assigned to the EUR group it was later eliminated from the dataset.

Genetic diversity of different apricots and their inferred subclusters

Population genetic parameters were estimated on Central Asian wild apricots and Chinese and European cultivars. Mean nucleotide diversity (π) in both Central Asian wild ($\pi=6.14e-3\pm 4.17e-4$) and Chinese cultivated apricots ($\pi=5.98e-3\pm 3.28e-4$) were significantly higher ($p<0.05$ by Wilcoxon Signed-Rank test) than in European cultivated apricots ($\pi=5.04e-3\pm 3.48e-4$) ($\pi_{CA}/\pi_{EUR}=1.21$, $\pi_{CHN}/\pi_{EUR}=1.18$) (Table III-S2). Besides the fact that π was

higher in Central Asian wild apricots, this group also had higher, positive values of Tajima's D (0.560 ∓ 0.068) while European and Chinese cultivated apricots also displayed positive D values of $0.234 (\mp 0.157)$ and $0.012 (\mp 0.058)$ respectively (Table III-S3).

Between-population F_{ST} estimates revealed relatively low genetic differentiation among apricot clusters, pairwise F_{ST} values ranging from 0.086 for EUR/CA down to 0.049 for CHN/CA (Table III-S4). However, F_{ST} estimates increased considerably when comparing Central Asian and European genotypes that were unequivocally assigned to a population with an assignment probability of $\geq 95\%$ ($F_{ST}=0.240$).

Demographic history of apricot

We inferred the demographic history of apricots using SMC++ which estimates changes in population size (N_e) (Terhorst et al., 2017). Historical climate periods of the most recent glaciations (12,500~110,000 ya -years ago-) and the last glacial maximum (LGM, ~21,500 ya) were also marked on the timeline (Figure III-7).

All three apricot groups are currently experiencing a decrease in population size, which is in concordance with positive Tajima's D values. However, while Chinese cultivated apricots experienced a continuous reduction of N_e since LGM (Figure III-7, red line), both European and Central Asian apricots encountered first expansion of their population sizes before decreasing steadily (Figure III-7, green and orange lines). Their ancestral N_e declined from over $4.37e^5$ around 1 Mya (million years ago) to almost $1.47e^4$ in the LGM period. There were in fact two obvious expansion peaks in the wild Central Asian and cultivated European apricots, one that occurred at the beginning of the glacial period and the second, in the 10 thousand year period that followed LGM. The rapid expansion in European N_e (up to a maximum of $3.6e^5$ about 5,000 years ago -5 Kya-) coincides with European apricot domestication as estimated by Liu et al (2019) (~3.98 Kya) followed by its dissemination westwards, from Central Asia to Europe. For Chinese cultivated apricots (red line in Figure III-7), their ancestral effective population size was estimated to reach $2.6e^5$ around 1.26 Mya, then underwent a trend of decline down to $N_e=2.9e^4$ until 67.86 Kya. In mid glacial period, SMC++ inferred a rapid increase of the Chinese population size (up to $1.8e^5$) followed by a rapid decline ahead of the LGM period. Finally, over the last 1,000 years, the effective population size of Chinese cultivated and Central Asian wild apricots grew slightly up to $5.29e^4$ and $2.48e^4$, respectively while European apricot N_e decreased down to $1.60e^4$.

Genomic signatures of selective sweeps within the apricot genome

Phenotypic traits that were favourably selected by humans or by local adaptation processes usually have low levels of variation and skewed allele frequency spectra, parameters that

have been successfully used in the current study to identify putative artificially and naturally selected genes in apricot. Our large SNP data set from both wild and cultivated apricot provides an opportunity to identify selected genes by comparing polymorphism levels in cultivated and wild apricots but also in between European and Chinese cultivars. Tables III-S5 to -S11 summarize statistics for genomic regions harbouring the strongest selection signals through multiple tests (described in the Material and Methods section), *i.e.* π , Tajima's D , CLR, Omega, F_{ST} , μ and hapFLK. All sweep analyses focused on sliding 10-Kb windows. Windows that scored in the top (or bottom, for π and Tajima's D) 0.5% were considered candidate sweep regions.

All detected signals from multiple tests as well as the density of SNPs were plotted against chromosomes for wild, Central Asian, and cultivated, Chinese and European groups (Figures III-8, -9 and -10, respectively).

Summary of selective sweeps detected through different tests

First, we implemented the following three tests: nucleotide diversity (Pi or π), Tajima's D and CLR which identify potentially selected regions by detecting skews in the site frequency spectrum (SFS) due to local reduction of the polymorphism level. A total of 58,552 signals were detected by nucleotide diversity (π) among cultivated and wild apricots (Table III-S5). The averaged π values were similar among Central Asian wild apricots ($\pi=3.542e^{-3}$), Chinese cultivated apricots ($\pi=3.310e^{-3}$) and European cultivated apricots ($\pi=3.085e^{-3}$). In Tajima's D test, we discovered a total of 58,552 positive (Tajima's D value >0) and negative (Tajima's D value <0) signals among wild and cultivated apricots (Table III-S6). Compared to cultivated apricots, wild apricots had significantly ($p<0.05$) higher positive Tajima's D values. Meanwhile, Tajima's D values in cultivated apricots from Europe was also significantly higher ($p<0.05$ by Wilcoxon Signed-Rank test) than that from China. We detected a total of 584 significant signals which included 38 (10 positives and 28 negatives) values in Chinese cultivated apricots, 393 (154 positives and 239 negatives) values in European cultivated apricots, and 152 (128 positives and 24 negatives) values in wild Central Asian apricots (Table III-S6). Within wild and cultivated apricots, CLR identified 293 10-Kb windows (Table III-S7).

In comparison with the three, Omega (ω) explores the selective signals by computing the regional LD pattern. In our study, Omega test resulted in a total of 57,990 signals from wild and cultivated apricots with mean values ranging from 1.975 (CHN) to 2.843 (CA) with an average of 2.514 (Table III-S8). The significant difference in Omega ($p<0.05$ by Wilcoxon Signed-Rank test) was found between Central Asian and Chinese apricots, as well as

between Chinese and European apricots. Overall, 289 regions were identified with a threshold of Omega value at 23.642 (top 0.5%).

Mu (μ) statistic was a recently introduced method that was developed to increase the accuracy of signal detection by integrating multiple measures such as local variance, LD and SFS. The 7,035,048 initial signals detected by *Mu* were reduced to 58,324 by averaging the values over 10 kb non-overlapping windows (Table III-S9). Wild apricots from Central Asia displayed significantly lower ($p < 0.05$ by Wilcoxon Signed-Rank test) *mu* values than for the cultivated apricots. At the 0.5% cutoff, *mu* identified a total of 291 10-Kb windows.

Based on population differentiation, both F_{ST} and hapFLK tests allow identifying regions of high divergence between wild and cultivated samples or between European and Chinese cultivars. Three pairwise combinations of apricot groups were thus investigated by computing F_{ST} , generating a total of 292 significant 10-Kb windows along the wild and cultivated apricot genomes (Table III-S10). A total of 6,925,951 signals-produced by hapFLK were further filtered by applying a p value < 0.05 significance threshold (from intermediate FLK test), resulting in 214,589 signals. Finally, by averaging hapFLK signals over 10-Kb windows, we identified a total of 105 regions under selection (Table III-S11).

We also applied barplot to visualise the distribution of signal values among apricot groups of Central Asian wild, European and Chinese cultivars. By comparing, we observe that different tests performed differently, depending on the apricot group. For example, while π and Tajima's D provided a similar distribution of the signals between the three apricot groups, CLR, omega and *mu* resulted in heterogeneous results (Figures III-S6, -S7 and -S8), depending on the apricot group and on the chromosome. However, such discrepancies were not correlated with the type of expected selective sweeps, hard or soft.

Convergent and divergent selective signatures among different apricot groups

Except for hapFLK, the other selective sweep tests such as π , CLR, Omega and Tajima' D were performed on the three apricot groups, independently, and on each chromosome, separately. Among the total of 1,753 significant signals obtained from the different tests (Table III-S12), some of them are unique to one of the three apricot groups but others are also shared between two or three groups (Table III-S12). This type of information is depicted in Venn diagrams in Figure III-11.

Based on the 0.5% cutoff and the 10-Kb sliding windows, the 1,753 significant signals obtained from the above tests were further integrated (Table III-S13). Some of the 1,753 windows were adjacent to one another or overlapped, which led to the designation of 1,459 distinct core regions as depicted in Figures III-12 and III-13. Figure III-12 shows that all

apricot chromosomes hold a significant amount of selective sweeps. In Figure III-13, we notice that whilst many selective sweeps are shared between European and Chinese cultivated apricots, many genomic regions are also specific to Europe or to China.

Putatively selected genes and pathways

Three strategies were used to investigate the genes or pathways underlying patterns of selection across the apricot genome. In a first approach, genomic intervals that were shown to be under selection with at least three different tests were fully annotated (Table III-4 A). For that, we used BlastX from NCBI and UniProt to perform a functional analysis based on the list of all genes in the regions showing signatures of a selective sweep. Note that as most genes have pleiotropic effects, selection may possibly act on other functional effects of genes than those highlighted in Table III-4. Second, an *a priori* approach was applied in which molecular markers linked to QTL of interest were mapped and checked for colocalization with selective sweeps (Table III-4 B). Finally, in the third approach, candidate genes involved in agronomically interesting traits (flowering induction, dormancy release, resistance to pathogens and fruit quality) were mapped and searched for colocalization with selective sweeps (Table III-4 C).

In the first approach, we remark an over-representation of genes associated with response to biotic (and to a lesser extent, abiotic) stresses: TMV resistance gene, NPR4 ankyrin repeat-containing protein, FLS2 LRR receptor kinase, cyclic nucleotide-gated ion channel 4, tyrosine-sulfated glycopeptide receptor 1, receptor-like protein EIX2, WAK2 wall-associated receptor kinase, RGA3 disease resistance protein (Table III-4 A). Another category of genes that are over-represented is the one related to the photoperiodic control of flowering (FAR1-Related Sequence 5 protein, Ubiquitin-like-specific protease ESD4, TOPLESS transcription factor, cycling DOF factor 2). Another striking result of selective sweeps mapping in apricots is the selection acting onto the (-)-alpha-pinene synthase-like gene but only in the cultivated European and Chinese groups (*SS_chr4_13000000*, Table III-4 A and Figure III-14). The region encompassing the (-)-alpha-pinene synthase gene was detected with μ , Tajima's *D* and CLR. This gene is involved in fruit flavour and in strawberry, *FaNES1* is exclusively present and highly expressed in the fruit of cultivated varieties (Aharoni et al., 2004).

In the second approach, we investigated how many of the genes with putative signatures of positive selection overlap with previously identified quantitative trait loci (QTL). For various agronomically important traits, we identified QTL candidate genes as those located within the QTL-intervals (<1 Mbp) on the genetic maps archived in the GDR

(Genome Database for Rosaceae) (Table III-4 B). Selective sweep analyses identified genes and regions that have been previously suspected to mediate agronomic changes. One example is that of the *PPVres* locus (linked to resistance to *Plum Pox Virus* in apricots) and colocalization with the MATH TRAF genes which are within a potential sweep region in Central Asian wild apricot genomes (*SS_chr1_7790000*) (Table III-4 B, Figure III-14). In a previous study, we showed a significantly higher frequency of the partially deleted MATH TRAF allele (ZP002) in natural populations of *P. armeniaca*, thus suspecting that selection was acting on this locus in natural stands (Decroocq et al, 2016). Another agronomically important loci under selection is the *Prunus* gametophytic self-incompatibility locus (*GSI*, also called *S* locus) located on chromosome 6 (*SS_chr6_23360000*, Table III-4 B), the positive selection being detected only for the Chinese cultivated group. We also observed a large proportion of overlap between selective sweeps (*SS_chr1_23856202* and *SS_chr7_8436653*) detected by the CLR method and two over three QTLs associated with bud breaking in apricot (qDBD.HAxRM-chr1 and qDBD.HAxRM-chr7, respectively) (Socquet-Juglard et al., 2012). Those selective sweeps are mapping over candidate genes linked to dormancy release and photoperiodic control of flowering while the third QTL (qDBD.HAxRM-ch4) encompasses selection candidate genes coding for LEUNIG and DOG1-like transcription factors, related to the control of flower meristem development and to the photoperiodic flowering initiation, respectively. Interestingly, the three selective sweeps overlapping the three qDBD.HAxRM QTLs were detected only in European cultivated apricots. The overlap between those three QTLs and selective sweeps associated with the flowering pathways might reflect that the underlying candidate genes play a crucial role in the flowering adaptation of European cultivars, and thus were under strong artificial selection during European apricot domestication and later dispersion and adaptation to European climates.

In the last approach, we sought to identify candidate genes associated with flowering initiation pathways, resistance to pathogens and fruit quality which would also map over genomic regions under selection either in the wild or cultivated apricots (Table III-4 C). While we did not detect selective signatures over the most common flowering hub genes (FLC, FT...), gene families related to the control of flowering time were nonetheless enriched in the candidate genes. It includes the *Apetala-1*, *Constans*, *Frigida-1*, *DOF*, *TOPLESS* and *TFL1*-like genes (Table III-4 C). Tajima's *D* also identified selection occurring over the *DAM6* locus on chromosome 1 (*SS_chr1_39090000*). *Dormancy-associated MADS-BOX (DAM)* genes encode transcription factors that were implicated in the control of dormancy transition initially in peach (Leida et al., 2012), then in pear (Niu et al., 2016), in apple (Wu et al., 2017),

in Japanese apricot (Zhao et al., 2018). Interestingly, in our apricot samples, the *DAM* locus is under strong positive selection in the wild Central Asian apricots but not in the Chinese and European cultivars. Concerning fruit traits, none of the candidate genes tested so far and related to fruit size (*CNR12* and *20*), sugar content (invertase, sucrose synthase, vacuolar pyrophosphatase), flesh colour (*CCD4*, *PpNAC1*), fruit skin colour (*MYB10.1* and *10.2*) colocalized with selective sweeps except for *cg44958.t1.p1* (*MYB*-like gene) and the *SS_chr3_8520000* genomic region that appear to be at proximity of the qFSC.HAxRM-ch3 QTL for fruit skin colour in apricot.

Although we were able to associate only a few genes related to fruit characteristics with selective sweeps, our result cannot suggest that fruit traits were not under selection during domestication. Indeed, we might simply not have tested the right candidate(s). To circumvent this problem, we retained among the genomic regions carrying selective sweeps differentiating apricot cultivars from wild populations, the ones that were shared between European and Chinese apricots, *i.e.* showing convergent genomic signatures of selection. From 1,753 sliding windows, 7.75% (136) are common between Chinese and European cultivated apricots (Table III-5). Out of these 136 regions, functional annotations are available for 48 loci, based on overlapping or close genes. Interestingly, 21 regions are associated among others to genes involved in response to pathogens (e.g., clusters of LRR-NBS resistance genes such as *SS_chr2_9390000*), 11 in response to abiotic stress (e.g., iron chlorosis, *SS_chr3_15560001*) while 11 other loci include at least one candidate gene involved in fruit quality traits such as sugar transport (*SS_chr2_16700001*), seed oil biosynthesis (*SS_chr1_36859000*), fruit softening at ripening (*SS_chr4_16570000*). We also identified a significant number of selective sweeps mapping over candidate genes linked to the control of flowering time through the photoperiodic pathway (e.g., *FAR1*, *ESD4* and *TOPLESS*) but also *VRN2*, a key regulator of the cold-mediated transition between vegetative and reproductive growth (*SS_chr5_12180000*) and *ALP1* that is needed for full reactivation of floral homeotic genes repressed by the Polycomb complex upon cold treatment (*SS_chr5_390000*).

To complete this work, we plan, in the near future and with the help of the bioinformaticians of CBIB (Centre de Bioinformatique de Bordeaux), to automatically annotate all 1,459 unique selective sweeps that were identified in this study. Functional categories will be assigned to underlying genes and we will then test for functional category enrichment using Fisher exact test. It is expected this way to unravel more biological processes that were under selection either during plant adaptation through climatic cycles in the wild Central Asian apricot or/and during Chinese and European apricot domestication.

Discussion

In this study, we provide a large genome variation data set for wild and cultivated apricots. Millions of SNPs in representative wild and cultivated apricot accessions provided an unprecedented opportunity to finely resolve the domestication history of cultivated apricot. Population structure and phylogenetic analyses based on this new set of genomic data support the hypothesis that European and Chinese apricots were independently domesticated, as previously suggested in Liu et al (2019, Chapter I of this manuscript) and Decroocq et al (2016). It also supports the hypothesis that both apricot cultivated groups encountered distinct demographic histories.

Indeed, in the current study, the combination of Bayesian clustering and SMC++ inference framework shed light on the past population history of apricot. Two aspects of our results, in particular, are worth re-emphasizing. First, we have shown that Central Asian natural populations retain the highest nucleotide diversity level but also that π values in cultivated apricots remain in a similar magnitude order than wild *P. armeniaca*. We estimated that Chinese cultivated apricots retained up to 97% of the nucleotide diversity of its wild Central Asian progenitors while European apricots retained 82%. This would indicate a limited loss of diversity and a weak bottleneck during apricot domestication, especially for Chinese cultivated apricots. In general, domestication bottlenecks seem to be less severe in perennials than in annuals (Miller and Gross, 2011). Perennial fruit crops would maintain an average of ~95% of the neutral variation found in wild populations (except for peach, for which only 34% was retained in Landraces and 25% in Western cultivars; Li et al, 2019) while annuals retain an average of ~60%. This poses the question of why Chinese cultivated apricots exhibit a lower loss of diversity than the European cultivars. And what mechanisms favour the retention of nucleotide variation within the Chinese apricot germplasm but not in the European one? The first is that cross-pollination is the dominant mating system in Chinese apricots which is no longer the case in European cultivars. While 51 to 58% of the modern and traditional European apricot cultivars are self-compatible (Burgos et al., 1997; Herrera et al., 2018), only 10% of the Chinese cultivated apricots are self-compatible (He et al., 2007). Another major difference between European and Chinese cultivated apricots is the past and ongoing gene flow between the domesticated crop and its wild relatives. Indeed, hybridization has often played a central role in the origin and diversification of perennials, leading to adaptation to new environments after dispersion (Gaut et al., 2015). European apricots originated from Central Asia and were later disseminated westwards to Europe where little to no wild-to-crop admixture occurred, except sporadically with wild plum (giving rise to the so-called black or purple apricot, Faust et al, 1998). In a recent study of

the only European *Armeniaca* wild related species, *P. brigantina*, we showed no signature of admixture between the apricot cultivated germplasm and its cross-compatible wild relative (Liu et al, Chapter II of this PhD manuscript). The situation is clearly different in China where at least three *Armeniaca* wild related species are sharing habitats with cultivated apricots, *i.e.* *P. sibirica* in the North, *P. mandshurica* in the NorthEast and *P. mume* in the SouthWest. Past hybridization (giving rise to the W3 *P. sibirica* group) and ongoing gene flow between *P. sibirica* and *P. armeniaca* were demonstrated by Liu et al (2019) (Chapter I of this PhD manuscript). As examples of documented wild-to-crop introgression in China, we can also cite the sweet kernel apricot (a hybrid between *P. sibirica* and *P. armeniaca* which is used for medicine purposes, Ai et al., 2011) and the Apricot Mei (hybrid between *P. mume* and *P. armeniaca*) (Zhang et al, 2018). Moreover, estimation of the pairwise F_{ST} showed that Chinese cultivated apricots were more closely related to Central Asian wild apricots than European cultivars were. This could also indicate recurrent gene flow between Chinese cultivated germplasm and Central Asian wild apricots, as previous results of Approximate Bayesian Computation (ABC) suggested (Liu et al, 2019; Chapter I of this PhD manuscript).

A second outstanding result of the apricot demographic history is the divergence between Chinese apricots on one side and Central Asian and European apricots, on the other side. From our SMC++ analysis, this happened early in the quaternary glaciation period, around 70 Kya. At this time, Chinese apricots encountered a severe reduction of population size followed by an expansion while a similar process occurred for the Central Asian and European apricots, not before the end of the Last Glaciation Moment. Microsatellite-based ABC results hypothesized in its best scenario, that Chinese apricots were arisen from Central Asian wild *P. armeniaca* or Chinese W3 *P. sibirica* ~93 to ~65 Kya (Liu et al, 2019, Chapter I of the manuscript) which is in accordance with the current study. Such differences in demographic history between Central Asian and Chinese apricots underscore the importance of glacial refugia and of postglacial migration patterns of both Chinese and Central Asian apricots. Indeed, niche modelling applied on other perennial species endemic in China showed that (i) multiple glacial refugia existed all along China, from West to East and from North to South, (ii) species persisted through the glaciation period in the mountains surrounding the Northern refugia (Hao et al., 2018; Zeng et al., 2015). This scenario stands in sharp contrast with the unidirectional long-distance south-to-north postglacial migration scenario suggested in the case of the Central Asian wild apricots (Liu et al., 2019, Chapter I of the manuscript). It would also limit the impact of glaciation period(s) on the Chinese apricot population size.

Genomic regions under natural selection provide extensive insights into the evolutionary history of wild apricot

In the current paper, we applied multiple statistics to detect genomic signatures of selection, in order to capture different signals left by positive selection along the wild and cultivated apricot genomes. Indeed, our genome-wide SNP data allow us not only to examine the adaptation of wild apricot to the local environment and climatic changes but also to investigate aspects of apricot domestication.

Since most Tajima's D values were positive in the Central Asian wild apricot genomes (thus indicating balancing selection) and natural apricot populations encountered only mild bottlenecks in their demographic history, both SFS- (e.g., Tajima's D) and LD-based (e.g., Omega statistics) methods should be able to alleviate the problem generated by the confounding effects of demography and detect selective sweeps accurately (Jensen et al., 2007; Pavlidis et al., 2017). Because balancing selection allows maintaining genetic variation within natural populations or species, it also plays a key role in adaptation (Wu et al., 2017). Therefore, we suppose that signals detected in the Central Asian wild apricot genomes, in particular, those detected by Tajima's D , might be linked to adaptive traits. Our preliminary analysis of the potential candidate genes underlying selective sweeps in Central Asian apricot genomes showed that those adaptive traits correspond mostly to pathogen resistance and synchronization of flowering (photoperiodic control of flower initiation) (Table III-4 A). We confirmed that, despite the absence of the viral pathogen, the *PPVres* locus is under selection and that accurate release of dormancy is a major trait in wild, Central Asian forests.

Selective sweeps and agronomically important genes

In plant evolutionary processes, domestication involves the genetic modification of wild species to create a new plant to meet human needs (Doebley et al., 2006). During this process, humans subjected common agronomic traits to artificial selection, thereby increasing the seed or fruit size, synchronization of growth and flowering, loss of seed dispersal, changes in plant architecture etc... (Harlan, 1992). Consequently, crop species have undergone extensive selection for these agronomically important traits, and genes impacted by artificial selection can be essential genetic factors in the domestication process. In the case of apricot, we detected selective sweeps over thousands of candidate genes that may have been artificially selected during the domestication of one or both of the two distinct cultivated groups. Unfortunately, because of lack of time, we were not able to unravel all biological pathways targeted by domestication but few trends were already observed: the

importance of fruit flavour, of resistance to pest and pathogen as well as the control of flowering time and synchronization of the vegetative and reproductive phases. One agronomically important trait for crop species is self-(in)compatibility since the self-compatible phenotype has been extensively selected during modern crop breeding. However, all wild *Armeniaca* species are, by nature, self-incompatible, following a gametophytic self-incompatibility (GSI) system. GSI occurs when pollen is rejected in the style or on the stigma if it possesses a matching allele with either of the ovule parent's S-alleles. This mechanism typically involves a single genetic locus that is highly polymorphic within populations and species. S-alleles are maintained by the strong negative frequency-dependent selection that essentially alleles when they become rare (Delph and Kelly, 2014). Therefore, we can hypothesize that this form of balancing selection where a rare allele has a selective advantage (Wright, 1939), maintains the high diversity observed in nature in the wild, Central Asian apricots. The situation is expected to be different in vegetatively propagated apricots such as for the Chinese and European cultivars, where the negative frequency-dependent selection is not acting anymore. However, while Chinese cultivars are predominantly self-incompatible, most Western European cultivars are self-compatible. Indeed, among the European cultivated germplasm, self-(in)compatibility is controlled by two independent loci, *S* (the above *GSI* locus, on chromosome 6) and *M* (*Modifier*-locus, distal part of chromosome 3) (Muñoz-Sanz et al., 2017). Over the *S* apricot locus, a knock-out mutation of the F-box open-reading frame, a 358 bp insertion in the *SFB* gene leading to self-compatibility, appeared a long time ago, along the apricot dissemination routes between Eastern Anatolia and Europe, well before the southern and northern European dissemination routes of apricot diverged (Halász et al., 2007). Independently and unlinked to the *S* locus, a mutation in *ParM-7* at the *M* locus also confers self-compatibility in European cultivated apricots (e.g., 'Canino' and 'Katy') (Muñoz-Sanz et al., 2017). The high frequency of self-compatible alleles within the European cultivated apricots would thus explain the absence of selective sweeps over the *S* and *M* loci, which is not the case in the Chinese cultivated group. We are not clear yet on how the self-incompatibility trait could benefit the Chinese population, but if so, the strong selection on self-incompatibility would disturb the fixation of positive sweep and reduce the number of significant signals in Chinese cultivated apricots. Indeed, our survey showed less significant selective sweeps found in Chinese cultivated apricots than in European cultivated apricots.

Conclusion

In our previous study, we showed that the apricot domestication history was mainly forced by migration (gene flow) events (Liu et al., 2019, Chapter I of the manuscript). Since

migration models can pose severe challenges to the current sweep detection methods (Alachiotis et al., 2018), we should, therefore, remain cautious when interpreting selective sweeps in apricot results. Nevertheless, the SNP data we generated will be useful as dense markers of genome variation for marker-assisted mapping of important apricot traits as well as for apricot breeding, and the candidate genes selected during domestication may be agronomically important. The data generated provide a valuable resource for apricot improvement; meanwhile, this study provides a genetic context for future research on the mechanisms and biological processes involved in adaptive evolution and domestication of apricots.

Acknowledgements

We thank A. Groppi, CBIB Université de Bordeaux, for providing us with the Marouch genome sequence and giving access to the apricot genome annotation browser. We also thank P. Civan, INRA GDEC Clermont-Ferrand, for helpful suggestions. All computations were performed at the Bordeaux Bioinformatics Center (CBiB). Computer resources were provided by the CBiB, Université de Bordeaux. We thank the GenoToul GeT-PlaGe platform in Toulouse for the long- and short-read sequencing and the CNRGV for the BIONANO optical map. We acknowledge funding from the ANR CHEX ABRIWG and Bordeaux University G2P and ATT initiatives (ABXING and SWAGMAN projects) which allowed to produce the *de novo* assembly of the apricot genome and ILLUMINA resequencing data as described in this paper.

Reference

- Aharoni, A., Giri, A. P., Verstappen, F. W., Berteaux, C. M., Sevenier, R., Sun, Z., . . . Bouwmeester, H. J. (2004). Gain and loss of fruit flavor compounds produced by wild and cultivated strawberry species. *The Plant Cell*, *16*(11), 3110-3131. doi:10.1105/tpc.104.023895
- Ai, P. F. (2011). Genetic diversity and relationships within sweet kernel apricot and related *Armeniaca* species based on sequence-related amplified polymorphism markers. *Biochemical systematics and ecology*, *39*(4-6), 694-699. doi:10.1016/j.bse.2011.05.026
- Alachiotis, N., & Pavlidis, P. (2016). Scalable linkage-disequilibrium-based selective sweep detection: a performance guide. *GigaScience*, *5*(1), 7. doi: 10.1186/s13742-016-0114-9
- Alachiotis, N., & Pavlidis, P. (2018). RAiSD detects positive selection based on multiple signatures of a selective sweep and SNP vectors. *Communications Biology*, *1*(1), 79. doi:10.1038/s42003-018-0085-8
- Alachiotis, N., Stamatakis, A., & Pavlidis, P. (2012). OmegaPlus: a scalable tool for rapid detection of selective sweeps in whole-genome datasets. *Bioinformatics*, *28*(17), 2274-2275. doi:10.1093/bioinformatics/bts419
- Balague, C., Lin, B., Alcon, C., Flottes, G., Malmstrom, S., Kohler, C., . . . Roby, D. (2003). HLM1, an essential signaling component in the hypersensitive response, is a member of the cyclic nucleotide-gated channel ion channel family. *The Plant Cell*, *15*(2), 365-379. doi:10.1105/tpc.006999
- Bao, Y., Song, W. M., Jin, Y. L., Jiang, C. M., Yang, Y., Li, B., . . . Zhang, H. X. (2014). Characterization of Arabidopsis Tubby-like proteins and redundant function of AtTLP3 and AtTLP9 in plant response to ABA and osmotic stress. *Plant molecular biology*, *86*(4-5), 471-483.
- Biswas, S., & Akey, J. M. (2006). Genomic insights into positive selection. *TRENDS in Genetics*, *22*(8), 437-446. doi: 10.1016/j.tig.2006.06.005
- Bonhomme, M., Chevalet, C., Servin, B., Boitard, S., Abdallah, J., Blott, S., & Sancristobal, M. (2010). Detecting selection in population trees: the Lewontin and Krakauer test extended. *Genetics*, *186*(1), 241-262. doi:10.1534/genetics.104.117275
- Bouwmeester, K., de Sain, M., Weide, R., Gouget, A., Klamer, S., Canut, H., & Govers, F. (2011). The lectin receptor kinase LecRK-I.9 is a novel Phytophthora resistance component and a potential host target for a RXLR effector. *PLoS Pathog*, *7*(3), e1001327. doi:10.1371/journal.ppat.1001327
- Boitard, S., Schlötterer, C., & Futschik, A. (2009). Detecting selective sweeps: a new approach based on hidden Markov models. *Genetics*, *181*(4), 1567-1578. doi:10.1534/genetics.108.100032
- Bowman, J. L., Alvarez, J., Weigel, D., Meyerowitz, E. M., & Smyth, D. R. (1993). Control of flower development in *Arabidopsis thaliana* by APETALA1 and interacting genes. *Development*, *119*(3), 721-743.
- Burgos, L., Egea, J., Guerriero, R., Viti, R., Monteleone, P., & Audergon, J. M. (1997). The self-compatibility trait of the main apricot cultivars and new selections from breeding programmes. *Journal of Horticultural Science*, *72*(1), 147-154. doi:10.1080/14620316.1997.11515501
- Carlsson, A. S., LaBrie, S. T., Kinney, A. J., Von Wettstein-Knowles, P., & Browse, J. (2002). A KAS2 cDNA complements the phenotypes of the Arabidopsis fab1 mutant that differs in a single residue bordering the substrate binding pocket. *The Plant Journal*, *29*(6), 761-770.
- Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, *4*(1). doi:10.1186/s13742-015-0047-8

CHAPTER III

- Chen, X. Y., & Kim, J. Y. (2009). Callose synthesis in higher plants. *Plant Signaling & Behavior*, 4(6), 489-492. doi:10.4161/psb.4.6.8359
- Cho, S. K., Ryu, M. Y., Kim, J. H., Hong, J. S., Oh, T. R., Kim, W. T., & Yang, S. W. (2017). RING E3 ligases: key regulatory elements are involved in abiotic stress responses in plants. *BMB reports*, 50(8), 393-400. doi:10.5483/bmbrep.2017.50.8.128
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., . . . Genomes Project Analysis, G. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156-2158. doi:10.1093/bioinformatics/btr330
- Daviere, J. M., & Achard, P. (2016). A pivotal role of DELLAs in regulating multiple hormone signals. *Mol Plant*, 9(1), 10-20. doi:10.1016/j.molp.2015.09.011
- De Lucia, F., Crevillen, P., Jones, A. M., Greb, T., & Dean, C. (2008). A PHD-polycomb repressive complex 2 triggers the epigenetic silencing of FLC during vernalization. *Proceedings of the National Academy of Sciences*, 105(44), 16831-16836.
- Decroocq, S., Chague, A., Lambert, P., Roch, G., Audergon, J. M., Geuna, F., . . . Decroocq, V. (2014). Selecting with markers linked to the PPVres major QTL is not sufficient to predict resistance to Plum Pox Virus (PPV) in apricot. *Tree Genetics & Genomes*, 10(5), 1161-1170. doi:10.1007/s11295-014-0750-0
- Decroocq, S., Cornille, A., Tricon, D., Babayeva, S., Chague, A., Eyquard, J. P., . . . Decroocq, V. (2016). New insights into the history of domesticated and wild apricots and its contribution to Plum pox virus resistance. *Molecular Ecology*, 25(19), 4712-4729. doi:10.1111/mec.13772
- Deeks, M. J., Calcutt, J. R., Ingle, E. K., Hawkins, T. J., Chapman, S., Richardson, A. C., . . . Smertenko, A. P. (2012). A superfamily of actin-binding proteins at the actin-membrane nexus of higher plants. *Current Biology*, 22(17), 1595-1600.
- DeGiorgio, M., Huber, C. D., Hubisz, M. J., Hellmann, I., & Nielsen, R. (2016). SweepFinder2: increased sensitivity, robustness and flexibility. *Bioinformatics*, 32(12), 1895-1897. doi:10.1093/bioinformatics/btw051
- Delph, L. F., & Kelly, J. K. (2014). On the importance of balancing selection in plants. *New Phytologist*, 201(1), 45-56. doi:10.1111/nph.12441
- Dinesh-Kumar, S., Tham, W.-H., & Baker, B. J. (2000). Structure–function analysis of the tobacco mosaic virus resistance gene N. *Proceedings of the National Academy of Sciences*, 97(26), 14789-14794.
- Dinesh-Kumar, S. P., & Baker, B. J. (2000). Alternatively spliced N resistance gene transcripts: their possible role in tobacco mosaic virus resistance. *Proceedings of the National Academy of Sciences*, 97(4), 1908-1913. doi:10.1073/pnas.020367497
- Doebley, J. F., Gaut, B. S., & Smith, B. D. (2006). The molecular genetics of crop domestication. *Cell*, 127(7), 1309-1321. doi:10.1016/j.cell.2006.12.006
- Duval, H., Hoerter, M., Polidori, J., Confolent, C., Masse, M., Moretti, A., . . . Esmenjaud, D. (2014). High-resolution mapping of the RMia gene for resistance to root-knot nematodes in peach. *Tree Genetics & Genomes*, 10(2), 297-306. doi:10.1007/s11295-013-0683-z
- Ellinger, D., & Voigt, C. A. (2014). Callose biosynthesis in Arabidopsis with a focus on pathogen response: what we have learned within the last decade. *Ann Bot*, 114(6), 1349-1358. doi:10.1093/aob/mcu120
- Fariello, M. I., Boitard, S., Naya, H., SanCristobal, M., & Servin, B. (2013). Detecting signatures of selection through haplotype differentiation among hierarchically structured populations. *Genetics*, 193(3), 929-941. doi:10.1534/genetics.112.147231

- Faust, M., Suranyi, D., & Nyujto, F. (1998). Origin and dissemination of apricot. *Horticultural Reviews -Westport then New York*, 22, 225-260.
- Ferrer-Admetlla, A., Liang, M., Korneliussen, T., & Nielsen, R. (2014). On detecting incomplete soft or hard selective sweeps using haplotype structure. *Molecular biology and evolution*, 31(5), 1275-1291. doi: 10.1093/molbev/msu077
- Feuillet, C., Schachermayr, G., & Keller, B. (1997). Molecular cloning of a new receptor-like kinase gene encoded at the Lr10 disease resistance locus of wheat. *The Plant Journal*, 11(1), 45-52.
- Fornara, F., Panigrahi, K. C., Gissot, L., Sauerbrunn, N., Ruhl, M., Jarillo, J. A., & Coupland, G. (2009). Arabidopsis DOF transcription factors act redundantly to reduce CONSTANS expression and are essential for a photoperiodic flowering response. *Dev Cell*, 17(1), 75-86. doi:10.1016/j.devcel.2009.06.015
- Fu, Y. X. (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, 147(2), 915-925.
- Fu, J. Y., Ren, F., Lu, X., Mao, H. J., Xu, M. M., Degenhardt, J., . . . Wang, Q. (2016). A tandem array of kaurene synthases in maize with roles in gibberellin and more specialized metabolism. *Plant physiology*, 170(2), 742-751. doi:10.1104/pp.15.01727
- Garud, N. R., Messer, P. W., Buzbas, E. O., & Petrov, D. A. (2015). Recent selective sweeps in North American *Drosophila melanogaster* show signatures of soft sweeps. *PLoS genetics*, 11(2), e1005004.
- Gaut, B. S., Díez, C. M., & Morrell, P. L. (2015). Genomics and the contrasting dynamics of annual and perennial domestication. *Trends in Genetics*, 31(12), 709-719. doi:10.1016/j.tig.2015.10.002
- Gigolashvili, T., & Kopriva, S. (2014). Transporters in plant sulfur metabolism. *Frontiers in plant science*, 5(442). doi:10.3389/fpls.2014.00442
- Gomez-Gomez, L., & Boller, T. (2000). FLS2: an LRR receptor-like kinase involved in the perception of the bacterial elicitor flagellin in Arabidopsis. *Mol Cell*, 5(6), 1003-1011. doi:10.1016/S1097-2765(00)80265-8
- Goralogia, G. S., Liu, T. K., Zhao, L., Panipinto, P. M., Groover, E. D., Bains, Y. S., & Imaizumi, T. (2017). CYCLING DOF FACTOR 1 represses transcription through the TOPLESS co-repressor to control photoperiodic flowering in Arabidopsis. *The Plant Journal*, 92(2), 244-262. doi:10.1111/tpj.13649
- Gu, Z., Gu, L., Eils, R., Schlesner, M., & Brors, B. (2014). circlize Implements and enhances circular visualization in R. *Bioinformatics*, 30(19), 2811-2812. doi:10.1093/bioinformatics/btu393
- Halasz, J., Pedryc, A., & Hegedus, A. (2007). Origin and dissemination of the pollen-part mutated SC haplotype which confers self-compatibility in apricot (*Prunus armeniaca*). *New Phytologist*, 176(4), 792-803. doi:10.1111/j.1469-8137.2007.02220.x
- Hao, Q., de Lafontaine, G., Guo, D., Gu, H., Hu, F. S., Han, Y., . . . Liu, H. (2018). The critical role of local refugia in postglacial colonization of Chinese pine: joint inferences from DNA analyses, pollen records, and species distribution modeling. *Ecography*, 41(4), 592-606. doi:10.1111/ecog.03096
- Harlan, J. R. (1992). Crops and Man. American Society of Agronomy. *Crop Science Society of America, Madison, Wisconsin*, 16(2), 63-262.
- Hassdenteufel, S., Johnson, N., Paton, A. W., Paton, J. C., High, S., & Zimmermann, R. (2018). Chaperone-mediated Sec61 channel gating during ER import of small precursor proteins overcomes Sec61 inhibitor-reinforced energy barrier. *Cell Rep*, 23(5), 1373-1386. doi:10.1016/j.celrep.2018.03.122

CHAPTER III

- He, T. M., Chen, X. S., Zhang, D. H., Xu, L., Liu, N., Gao, J. S., & Xu, Z. (2007). Frequency distribution of several biological characters in different apricot eco-geographical groups native to China. *Acta Horticulturae Sinica*, 34(1), 17.
- Heras, B., & Drøbak, B. K. (2002). Plants and the Environment. PARF-1: an *Arabidopsis thaliana* FYVE-domain protein displaying a novel eukaryotic domain structure and phosphoinositide affinity. *J Exp Bot*, 53(368).
- Herrera, S., Lora, J., Hormaza, J. I., Herrero, M., & Rodrigo, J. (2018). Optimizing production in the new generation of apricot cultivars: self-incompatibility, S-RNase allele identification, and incompatibility group assignment. *Frontiers in plant science*, 9, 527-527. doi:10.3389/fpls.2018.00527
- Hohenlohe, P. A., Phillips, P. C., & Cresko, W. A. (2010). Using population genomics to detect selection in natural populations: key concepts and methodological considerations. *International Journal of Plant Sciences*, 171(9), 1059-1071. doi: 10.1086/656306
- Huber, C. D., DeGiorgio, M., Hellmann, I., & Nielsen, R. (2016). Detecting recent selective sweeps while controlling for mutation rate and background selection. *Molecular Ecology*, 25(1), 142-156. doi:10.1111/mec.13351
- Jauvion, V., Elmayan, T., & Vaucheret, H. (2010). The conserved RNA trafficking proteins HPR1 and TEX1 are involved in the production of endogenous and exogenous small interfering RNA in Arabidopsis. *The Plant Cell*, 22(8), 2697-2709.
- Jensen, J. D., Thornton, K. R., Bustamante, C. D., & Aquadro, C. F. (2007). On the utility of linkage disequilibrium as a statistic for identifying targets of positive selection in nonequilibrium populations. *Genetics*, 176(4), 2371-2379. doi:10.1534/genetics.106.069450
- Jin, W. M., Wang, H., Li, M. F., Wang, J., Yang, Y., Zhang, X. M., . . . Zhang, K. C. (2016). The R2R3 MYB transcription factor PavMYB10.1 involves in anthocyanin biosynthesis and determines fruit skin colour in sweet cherry (*Prunus avium* L.). *Plant Biotechnology Journal*, 14(11), 2120-2133. doi:10.1111/pbi.12568
- Kim, J., Kim, J. H., Lyu, J. I., Woo, H. R., & Lim, P. O. (2017). New insights into the regulation of leaf senescence in Arabidopsis. *J Exp Bot*, 69(4), 787-799.
- Kim, Y., & Nielsen, R. (2004). Linkage disequilibrium as a signature of selective sweeps. *Genetics*, 167(3), 1513-1524. doi:10.1534/genetics.103.025387
- Knaus, B. J., & Grünwald, N. J. (2017). vcfr: a package to manipulate and visualize variant call format data in R. *Molecular Ecology Resources*, 17(1), 44-53. doi:10.1111/1755-0998.12549
- Knip, M., de Pater, S., & Hooykaas, P. J. (2012). The SLEEPER genes: a transposase-derived angiosperm-specific gene family. *BMC Plant Biology*, 12(1), 192.
- Ledbetter, C. A. (2010). Apricot breeding in North America: Current status and future prospects. *Acta horticulturae*, 862, 85-92. doi:10.17660/ActaHortic.2010.862.12
- Leida, C., Conesa, A., Llacer, G., Badenes, M. L., & Rios, G. (2012). Histone modifications and expression of DAM6 gene in peach are modulated during bud dormancy release in a cultivar-dependent manner. *New Phytologist*, 193(1), 67-80. doi:10.1111/j.1469-8137.2011.03863.x
- Leida, C., Terol, J., Martí, G., Agustí, M., Llácer, G., Badenes, M. L., & Ríos, G. (2010). Identification of genes associated with bud dormancy release in *Prunus persica* by suppression subtractive hybridization. *Tree Physiology*, 30(5), 655-666. doi:10.1093/treephys/tpq008
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14), 1754-1760. doi:10.1093/bioinformatics/btp324

- Li, J., Zhao-Hui, C., Batoux, M., Nekrasov, V., Roux, M., Chinchilla, D., . . . Jones, J. D. (2009). Specific ER quality control components required for biogenesis of the plant innate immune receptor EFR. *Proceedings of the National Academy of Sciences*, 106(37), 15973-15978.
- Li, Y., Cao, K., Zhu, G. R., Fang, W. C., Chen, C. W., Wang, X. W., . . . Wang, L. R. (2019). Genomic analyses of an extensive collection of wild and cultivated accessions provide new insights into peach breeding history. *Genome Biology*, 20(1), 36. doi:10.1186/s13059-019-1648-9
- Liang, S. C., Hartwig, B., Perera, P., Mora-García, S., de Leau, E., Thornton, H., . . . Goodrich, J. (2015). Kicking against the PRCs – A domesticated transposase antagonises silencing mediated by polycomb group proteins and is an accessory component of polycomb repressive complex 2. *PLoS Genet*, 11(12), e1005660. doi:10.1371/journal.pgen.1005660
- Lin, R., Ding, L., Casola, C., Ripoll, D. R., Feschotte, C., & Wang, H. (2007). Transposase-derived transcription factors regulate light signaling in Arabidopsis. *Science*, 318(5854), 1302-1305. doi:10.1126/science.1146281
- Liu, B., Butenko, M. A., Shi, C.-L., Bolivar, J. L., Winge, P., Stenvik, G.-E., . . . Aalen, R. B. (2013). NEVERSHED and INFLORESCENCE DEFICIENT IN ABSCISSION are differentially required for cell expansion and cell separation during floral organ abscission in *Arabidopsis thaliana*. *J Exp Bot*, 64(17), 5345-5357. doi:10.1093/jxb/ert232
- Lotterhos, K. E., & Whitlock, M. C. (2014). Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. *Molecular Ecology*, 23(9), 2178-2192. doi: 10.1111/mec.12725
- Lv, F. N., Wang, H. H., Wang, X. Y., Han, L. B., Ma, Y. P., Wang, S., . . . Guo, W. Z. (2015). GhCFE1A, a dynamic linker between the ER network and actin cytoskeleton, plays an important role in cotton fibre cell initiation and elongation. *J Exp Bot*, 66(7), 1877-1889. doi:10.1093/jxb/eru530
- Mackey, D., Belkhadir, Y., Alonso, J. M., Ecker, J. R., & Dangl, J. L. (2003). Arabidopsis RIN4 is a target of the type III virulence effector AvrRpt2 and modulates RPS2-mediated resistance. *Cell*, 112(3), 379-389. doi:10.1016/s0092-8674(03)00040-0
- Marathe, R., Anandalakshmi, R., Liu, Y., & Dinesh-Kumar, S. P. (2002). The tobacco mosaic virus resistance gene, N. *Molecular Plant Pathology*, 3(3), 167-172. doi:10.1046/j.1364-3703.2002.00110.x
- Mariette, S., Wong Jun Tai, F., Roch, G., Barre, A., Chague, A., Decroocq, S., . . . Decroocq, V. (2016). Genome-wide association links candidate genes to resistance to Plum Pox Virus in apricot (*Prunus armeniaca*). *New Phytologist*, 209(2), 773-784. doi:10.1111/nph.13627
- Marín-Rodríguez, M. C., Orchard, J., & Seymour, G. B. (2002). Pectate lyases, cell wall degradation and fruit softening. *J Exp Bot*, 53(377), 2115-2119. doi:10.1093/jxb/erf089
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal*, 17(1), 10-12. doi:10.14806/ej.17.1.200
- Matsumoto, D., & Tao, R. (2016). Distinct self-recognition in the *Prunus* S-RNase-based gametophytic self-incompatibility system. *The Horticulture Journal, advpub*. doi:10.2503/hortj.MI-IR06
- Ma, Y., Ding, X., Qanbari, S., Weigend, S., Zhang, Q., & Simianer, H. (2015). Properties of different selection signature statistics and a new strategy for combining them. *Heredity*, 115(5), 426. doi: 10.1038/hdy.2015.42
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., . . . DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9), 1297-1303. doi:10.1101/gr.107524.110

CHAPTER III

- Michaels, S. D., Bezerra, I. C., & Amasino, R. M. (2004). FRIGIDA related genes are required for the winter-annual habit in Arabidopsis. *Proceedings of the National Academy of Sciences*, *101*(9), 3281-3285. doi:10.1073/pnas.0306778101
- Miller, A. J., & Gross, B. L. (2011). From forest to field: perennial fruit crop domestication. *Am J Bot*, *98*(9), 1389-1414. doi:10.3732/ajb.1000522
- Miller, J. C., Chezem, W. R., & Clay, N. K. (2016). Ternary WD40 repeat-containing protein complexes: evolution, composition and roles in plant immunity. *Frontiers in plant science*, *6*, 1108.
- Mohamed, R., Wang, C.-T., Ma, C., Shevchenko, O., Dye, S. J., Puzey, J. R., . . . Brunner, A. M. (2010). Populus CEN/TFL1 regulates first onset of flowering, axillary meristem identity and dormancy release in Populus. *The Plant Journal*, *62*(4), 674-688. doi:10.1111/j.1365-313X.2010.04185.x
- Mosher, S., Seybold, H., Rodriguez, P., Stahl, M., Davies, K. A., Dayaratne, S., . . . Kemmerling, B. (2013). The tyrosine-sulfated peptide receptors PSKR1 and PSY1R modify the immunity of Arabidopsis to biotrophic and necrotrophic pathogens in an antagonistic manner. *The Plant Journal*, *73*(3), 469-482. doi:10.1111/tpj.12050
- Moshkanbaryans, L., Xue, J., Wark, J. R., Robinson, P. J., & Graham, M. E. (2016). A novel sequence in AP180 and CALM promotes efficient clathrin binding and assembly. *PLOS ONE*, *11*(8), e0162050. doi:10.1371/journal.pone.0162050
- Mou, S. L., Zhang, X. X., Han, Z. F., Wang, J. W., Gong, X. Q., & Chai, J. J. (2017). CLE42 binding induces PXL2 interaction with SERK2. *Protein & Cell*, *8*(8), 612-617. doi:10.1007/s13238-017-0435-1
- Muñoz-Sanz, J. V., Zuriaga, E., López, I., Badenes, M. L., & Romero, C. (2017). Self-(in)compatibility in apricot germplasm is controlled by two major loci, S and M. *BMC Plant Biology*, *17*(1), 82. doi:10.1186/s12870-017-1027-1
- Murtas, G., Reeves, P. H., Fu, Y. F., Bancroft, I., Dean, C., & Coupland, G. (2003). A nuclear protease required for flowering-time regulation in Arabidopsis reduces the abundance of SMALL UBIQUITIN-RELATED MODIFIER conjugates. *The Plant Cell*, *15*(10), 2308-2319. doi:10.1105/tpc.015487
- Nagatoshi, M., Terasaka, K., Nagatsu, A., & Mizukami, H. (2011). Iridoid-specific glucosyltransferase from *Gardenia jasminoides*. *J Biol Chem*, *286*(37), 32866-32874. doi:10.1074/jbc.M111.242586
- Nei, M., & Li, W. H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences*, *76*(10), 5269-5273. doi:10.1073/pnas.76.10.5269
- Nielsen, R., Williamson, S., Kim, Y., Hubisz, M. J., Clark, A. G., & Bustamante, C. (2005). Genomic scans for selective sweeps using SNP data. *Genome research*, *15*(11), 1566-1575. doi:10.1101/gr.4252305
- Niu, Q., Li, J., Cai, D., Qian, M., Jia, H., Bai, S., . . . Zheng, X. (2016). Dormancy-associated MADS-box genes and microRNAs jointly control dormancy transition in pear (*Pyrus pyrifolia* white pear group) flower bud. *J Exp Bot*, *67*(1), 239-257. doi:10.1093/jxb/erv454
- Pastuglia, M., Swarup, R., Rocher, A., Saindrenan, P., Roby, D., Dumas, C., & Cock, J. M. (2002). Comparison of the expression patterns of two small gene families of S gene family receptor kinase genes during the defence response in *Brassica oleracea* and *Arabidopsis thaliana*. *Gene*, *282*(1-2), 215-225.
- Patel, R. K., & Jain, M. (2012). NGS QC toolkit: a toolkit for quality control of next generation sequencing data. *PLOS ONE*, *7*(2), e30619. doi:10.1371/journal.pone.0030619
- Pavlidis, P., & Alachiotis, N. (2017). A survey of methods and tools to detect recent and strong positive selection. *Journal of biological research (Thessalonike, Greece)*, *24*, 7. Retrieved from doi:10.1186/s40709-017-0064-0

- Pavlidis, P., Hutter, S., & Stephan, W. (2008). A population genomic approach to map recent positive selection in model species. *Molecular Ecology*, 17(16), 3585-3598. doi:10.1111/j.1365-294X.2008.03852.x
- Pavlidis, P., Jensen, J. D., & Stephan, W. (2010). Searching for footprints of positive selection in whole-genome SNP data from nonequilibrium populations. *Genetics*, 185(3), 907-922. doi:10.1534/genetics.110.116459
- Pavlidis, P., Živkovic, D., Stamatakis, A., & Alachiotis, N. (2013). SweeD: likelihood-based detection of selective sweeps in thousands of genomes. *Molecular biology and evolution*, 30(9), 2224-2234. doi:10.1093/molbev/mst112
- Puritz, J. B., Hollenbeck, C. M., & Gold, J. R. (2014). dDocent: a RADseq, variant-calling pipeline designed for population genomics of non-model organisms. *PeerJ*, 2, e431. doi:10.7717/peerj.431
- Rate, D. N., Cuenca, J. V., Bowman, G. R., Guttman, D. S., & Greenberg, J. T. (1999). The gain-of-function *Arabidopsis* *acd6* mutant reveals novel regulation and function of the salicylic acid signaling pathway in controlling cell death, defenses, and cell growth. *The Plant Cell*, 11(9), 1695-1708.
- Ron, M., & Avni, A. (2004). The receptor for the fungal elicitor ethylene-inducing xylanase is a member of a resistance-like gene family in tomato. *The Plant Cell*, 16(6), 1604-1615. doi:10.1105/tpc.022475
- Ronen, R., Udpa, N., Halperin, E., & Bafna, V. (2013). Learning natural selection from the site frequency spectrum. *Genetics*, 195(1), 181-193. doi: 10.1534/genetics.113.152587
- Roycewicz, P. S., & Malamy, J. E. (2014). Cell wall properties play an important role in the emergence of lateral root primordia from the parent root. *J Exp Bot*, 65(8), 2057-2069.
- Sabeti, P. C., Reich, D. E., Higgins, J. M., Levine, H. Z., Richter, D. J., Schaffner, S. F., ... & Ackerman, H. C. (2002). Detecting recent positive selection in the human genome from haplotype structure. *Nature*, 419(6909), 832. doi.org/10.1038/nature01140
- Schnurr, J. A., Shockey, J. M., & de Boer, G.-J. (2002). Fatty acid export from the chloroplast. Molecular characterization of a major plastidial acyl-coenzyme A synthetase from *Arabidopsis*. *Plant physiology*, 129(4), 1700-1709.
- Sebastian, J., Ravi, M., Andreuzza, S., Panoli, A. P., Marimuthu, M. P., & Siddiqi, I. (2009). The plant adherin AtSCC2 is required for embryogenesis and sister-chromatid cohesion during meiosis in *Arabidopsis*. *The Plant Journal*, 59(1), 1-13.
- Shen, Y., & Diener, A. C. (2013). *Arabidopsis thaliana* resistance to *fusarium oxysporum* 2 implicates tyrosine-sulfated peptide signaling in susceptibility and resistance to root infection. *PLoS Genet*, 9(5), e1003525. doi:10.1371/journal.pgen.1003525
- Shim, J. S., Kubota, A., & Imaizumi, T. (2017). Circadian clock and photoperiodic flowering in *Arabidopsis*: CONSTANS is a hub for signal integration. *Plant physiology*, 173(1), 5-15. doi:10.1104/pp.16.01327
- Shriver, M. D., Kennedy, G. C., Parra, E. J., Lawson, H. A., Sonpar, V., Huang, J., . . . Jones, K. W. (2004). The genomic distribution of population substructure in four populations using 8,525 autosomal SNPs. *Hum Genomics*, 1(4), 274-286. doi:10.1186/1479-7364-1-4-274
- Sitaraman, J., Bui, M., & Liu, Z. C. (2008). LEUNIG_HOMOLOG and LEUNIG perform partially redundant functions during *Arabidopsis* embryo and floral development. *Plant physiology*, 147(2), 672-681. doi:10.1104/pp.108.115923
- Smith, J. M., & Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetics Research*, 23(1), 23-35. doi:10.1017/S0016672300014634

CHAPTER III

- Socquet-Juglard, D., Christen, D., Devènes, G., Gessler, C., Duffy, B., & Patocchi, A. (2013). Mapping architectural, phenological, and fruit quality QTLs in apricot. *Plant Molecular Biology Reporter*, 31, 387-397. doi:10.1007/s11105-012-0511-x
- Sunkar, R., Kaplan, B., Bouche, N., Arazi, T., Dolev, D., Talke, I. N., . . . Fromm, H. (2000). Expression of a truncated tobacco NtCBP4 channel in transgenic plants and disruption of the homologous Arabidopsis CNGC1 gene confer Pb²⁺ tolerance. *The Plant Journal*, 24(4), 533-542. doi:10.1111/j.1365-313X.2000.00901.x
- Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 123(3), 585-595.
- Takagi, J., Renna, L., Takahashi, H., Koumoto, Y., Tamura, K., Stefano, G., . . . Shimada, T. (2013). MAIGO5 functions in protein export from Golgi-associated endoplasmic reticulum exit sites in Arabidopsis. *The Plant Cell*, 25(11), 4658-4675.
- Takeda, A., Iwasaki, S., Watanabe, T., Utsumi, M., & Watanabe, Y. (2008). The mechanism selecting the guide strand from small RNA duplexes is different among argonaute proteins. *Plant and Cell Physiology*, 49(4), 493-500.
- Terhorst, J., Kamm, J. A., & Song, Y. S. (2017). Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nat Genet*, 49(2), 303-309. doi:10.1038/ng.3748
- Tran, L.-S. P., Nakashima, K., Sakuma, Y., Simpson, S. D., Fujita, Y., Maruyama, K., . . . Yamaguchi-Shinozaki, K. (2004). Isolation and functional analysis of Arabidopsis stress-inducible NAC transcription factors that bind to a drought-responsive cis Element in the early responsive to dehydration stress 1 Promoter. *The Plant Cell*, 16(9), 2481-2498. doi:10.1105/tpc.104.022699
- Vatsiou, A. I., Bazin, E., & Gaggiotti, O. E. (2016). Detection of selective sweeps in structured populations: a comparison of recent methods. *Molecular Ecology*, 25(1), 89-103. doi: 10.1111/mec.13360
- Verica, J. A., & He, Z. H. (2002). The cell wall-associated kinase (WAK) and WAK-like kinase gene family. *Plant physiology*, 129(2), 455-459.
- Vo, K. T. X., Kim, C.-Y., Chandran, A. K. N., Jung, K.-H., An, G., & Jeon, J.-S. (2015). Molecular insights into the function of ankyrin proteins in plants. *Journal of plant biology*, 58(5), 271-284.
- Vorwerk, S., Schiff, C., Santamaria, M., Koh, S., Nishimura, M., Vogel, J., . . . Somerville, S. (2007). EDR2 negatively regulates salicylic acid-based defenses and cell death during powdery mildew infections of Arabidopsis thaliana. *BMC Plant Biology*, 7(1), 35.
- Vos, P. G., Paulo, M. J., Voorrips, R. E., Visser, R. G. F., van Eck, H. J., & van Eeuwijk, F. A. (2017). Evaluation of LD decay and various LD-decay estimators in simulated and SNP-array data of tetraploid potato. *Theoretical and Applied Genetics*, 130(1), 123-135. doi:10.1007/s00122-016-2798-8
- Wagner, T. A., & Kohorn, B. D. (2001). Wall-associated kinases are expressed throughout plant development and are required for cell expansion. *The Plant Cell*, 13(2), 303-318. doi:10.1105/tpc.13.2.303
- Wang, M., Allefs, S., van den Berg, R. G., Vleeshouwers, V. G. A. A., van der Vossen, E. A. G., & Vosman, B. (2008). Allele mining in Solanum: conserved homologues of Rpi-blb1 are identified in Solanum stoloniferum. *Theoretical and Applied Genetics*, 116(7), 933-943. doi:10.1007/s00122-008-0725-3
- Wang, M., Huang, X., Li, R., Xu, H., Jin, L., & He, Y. (2014). Detecting recent positive selection with high accuracy and reliability by conditional coalescent tree. *Molecular Biology and Evolution*, 31(11), 3068-3080. doi: 10.1093/molbev/msu244

- Wang, T., Liang, L., Xue, Y., Jia, P. F., Chen, W., Zhang, M. X., . . . Yang, W. C. (2016). A receptor heteromer mediates the male perception of female attractants in plants. *Nature*, *531*(7593), 241-244. doi:10.1038/nature16975
- Wang, Y., Cordewener, J. H., America, A. H., Shan, W., Bouwmeester, K., & Govers, F. (2015). Arabidopsis lectin receptor kinases LecRK-IX. 1 and LecRK-IX. 2 are functional analogs in regulating Phytophthora resistance and plant cell death. *Molecular Plant-Microbe Interactions*, *28*(9), 1032-1048. doi:10.1094/MPMI-02-15-0025-R
- Wang, Z., Zeng, Y. F., Zhang, Z. D., Sheng, S. B., Tian, J., Wu, R. L., & Pang, X. M. (2017). Phylogeography study of the siberian apricot (*Prunus sibirica* L.) in northern China assessed by chloroplast microsatellite and DNA makers. *Frontiers in plant science*, *8*, 1989-1989. doi:10.3389/fpls.2017.01989
- Wei, X. Y., Liu, F. L., Chen, C., Ma, F. W., & Li, M. J. (2014). The *Malus domestica* sugar transporter gene family: identifications based on genome and expression profiling related to the accumulation of fruit sugars. *Frontiers in plant science*, *5*, 569.
- Weigand, H., & Leese, F. (2018). Detecting signatures of positive selection in non-model species using genomic data. *Zoological Journal of the Linnean Society*, *184*(2), 528-583. doi:10.1093/zoolinnean/zly007
- Wright, S. (1939). The distribution of self-sterility alleles in populations. *Genetics*, *24*(4), 538-552.
- Wu, H. L., Li, L. H., Du, J., Yuan, Y. X., Cheng, X. D., & Ling, H. Q. (2005). Molecular and biochemical characterization of the Fe (III) chelate reductase gene family in *Arabidopsis thaliana*. *Plant and Cell Physiology*, *46*(9), 1505-1514.
- Wu, Q., Han, T. S., Chen, X., Chen, J. F., Zou, Y. P., Li, Z. W., . . . Guo, Y. L. (2017). Long-term balancing selection contributes to adaptation in Arabidopsis and its relatives. *Genome Biology*, *18*(1), 217. doi:10.1186/s13059-017-1342-8
- Wu, R., Tomes, S., Karunairetnam, S., Tustin, S. D., Hellens, R. P., Allan, A. C., . . . Varkonyi-Gasic, E. (2017). SVP-like MADS box genes control dormancy and budbreak in apple. *Frontiers in plant science*, *8*, 477. doi:10.3389/fpls.2017.00477
- Xu, R. R., Qi, S. D., Lu, L. T., Chen, C. T., Wu, C. A., & Zheng, C. C. (2011). A DExD/H box RNA helicase is important for K⁺ deprivation responses and tolerance in *Arabidopsis thaliana*. *The FEBS journal*, *278*(13), 2296-2306.
- Yi, X., Liang, Y., Huerta-Sanchez, E., Jin, X., Cuo, Z. X., Pool, J. E., . . . Wang, J. (2010). Sequencing of 50 human exomes reveals adaptation to high altitude. *Science*, *329*(5987), 75-78. doi:10.1126/science.1190371
- Yuan, Y., Zhong, S., Li, Q., Zhu, Z., Lou, Y., Wang, L., . . . He, Z. (2007). Functional analysis of rice NPR1-like genes reveals that OsNPR1/NH1 is the rice orthologue conferring disease resistance with enhanced herbivore susceptibility. *Plant Biotechnology Journal*, *5*(2), 313-324. doi:10.1111/j.1467-7652.2007.00243.x
- Zaveska Drabkova, L., & Honys, D. (2017). Evolutionary history of callose synthases in terrestrial plants with emphasis on proteins involved in male gametophyte development. *PLOS ONE*, *12*(11), e0187331. doi:10.1371/journal.pone.0187331
- Zeder, M. A. (2015). Core questions in domestication research. *Proceedings of the National Academy of Sciences*, *112*(11), 3191-3198. doi:10.1073/pnas.1501711112
- Zeng, K., Fu, Y. X., Shi, S., & Wu, C. I. (2006). Statistical tests for detecting positive selection by utilizing high-frequency variants. *Genetics*, *174*(3), 1431-1439. doi:10.1534/genetics.106.061432

CHAPTER III

- Zeng, Y. F., Wang, W. T., Liao, W. J., Wang, H. F., & Zhang, D. Y. (2015). Multiple glacial refugia for cool-temperate deciduous trees in northern East Asia: the Mongolian oak as a case study. *Molecular Ecology*, 24(22), 5676-5691. doi:10.1111/mec.13408
- Zhai, W., Nielsen, R., & Slatkin, M. (2008). An investigation of the statistical power of neutrality tests based on comparative and population genetic data. *Molecular Biology and Evolution*, 26(2), 273-283. doi:10.1093/molbev/msn231
- Zhang, Q., Zhang, H., Sun, L., Fan, G., Ye, M., Jiang, L., . . . Cheng, T. (2018). The genetic architecture of floral traits in the woody plant *Prunus mume*. *Nat Commun*, 9(1), 1702. doi:10.1038/s41467-018-04093-z
- Zhao, K., Zhou, Y., Ahmad, S., Xu, Z., Li, Y., Yang, W., . . . Zhang, Q. (2018). Comprehensive cloning of *Prunus mume* dormancy associated MADS-Box genes and their response in flower bud development and dormancy. *Frontiers in plant science*, 9, 17. doi:10.3389/fpls.2018.00017
- Zhao, Z. C., Hu, G. B., Hu, F. C., Wang, H. C., Yang, Z. Y., & Lai, B. (2012). The UDP glucose: flavonoid-3-O-glucosyltransferase (UFGT) gene regulates anthocyanin biosynthesis in litchi (*Litchi chinensis* Sonn.) during fruit coloration. *Molecular Biology Reports*, 39(6), 6409-6415. doi:10.1007/s11033-011-1303-3
- Zhu, Z., Xu, F., Zhang, Y., Cheng, Y. T., Wiermer, M., Li, X., & Zhang, Y. (2010). Arabidopsis resistance protein SNC1 activates immune responses through association with a transcriptional corepressor. *Proceedings of the National Academy of Sciences*, 107(31), 13960-13965. doi:10.1073/pnas.1002828107
- Živković, D., & Stephan, W. (2011). Analytical results on the neutral non-equilibrium allele frequency spectrum based on diffusion theory. *Theoretical Population Biology*, 79(4), 184-191. doi:10.1016/j.tpb.2011.03.003
- Zuriaga, E., Soriano, J. M., Zhebentyayeva, T., Romero, C., Dardick, C., Cañizares, J., & Badenes, M. L. (2013). Genomic analysis reveals MATH gene(s) as candidate(s) for Plum pox virus (PPV) resistance in apricot (*Prunus armeniaca* L.). *Molecular Plant Pathology*, 14(7), 663-677. doi:10.1111/mpp.12037

Figures

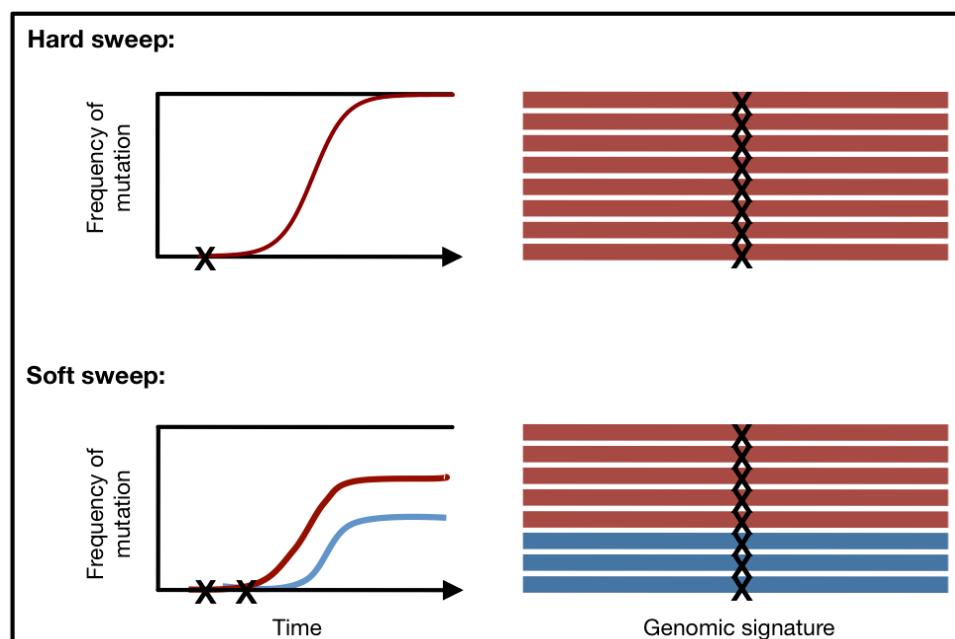


Figure III-1. Genomic signatures of hard and soft sweeps arising from one versus multiple mutations. Legend: Mutations are indicated by an X. In a hard sweep, only one haplotype will be present at high frequency, whereas in a soft sweep multiple haplotypes will be present at high frequencies (two haplotypes are present at high frequency for the depicted soft sweep). Illustration conceived by Philipp Messer. <https://garud.eeb.ucla.edu/selection-scan-scripts/>.

Workflow for detection of genomic regions under selection in this study

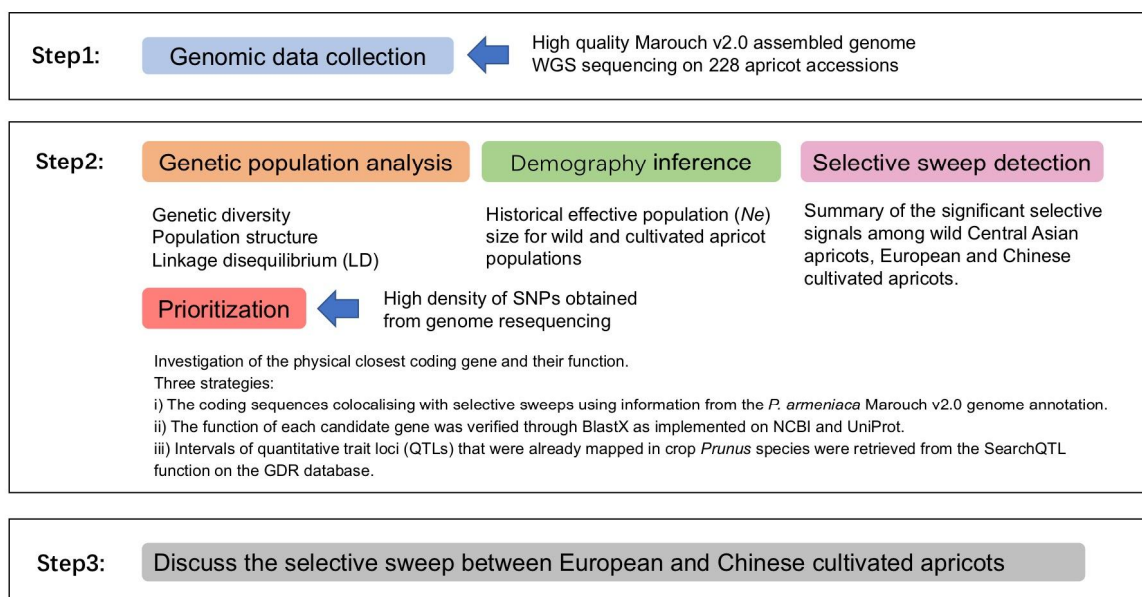


Figure III-2. Workflow for the detection of genomic regions under selection in wild and cultivated apricot genomes.

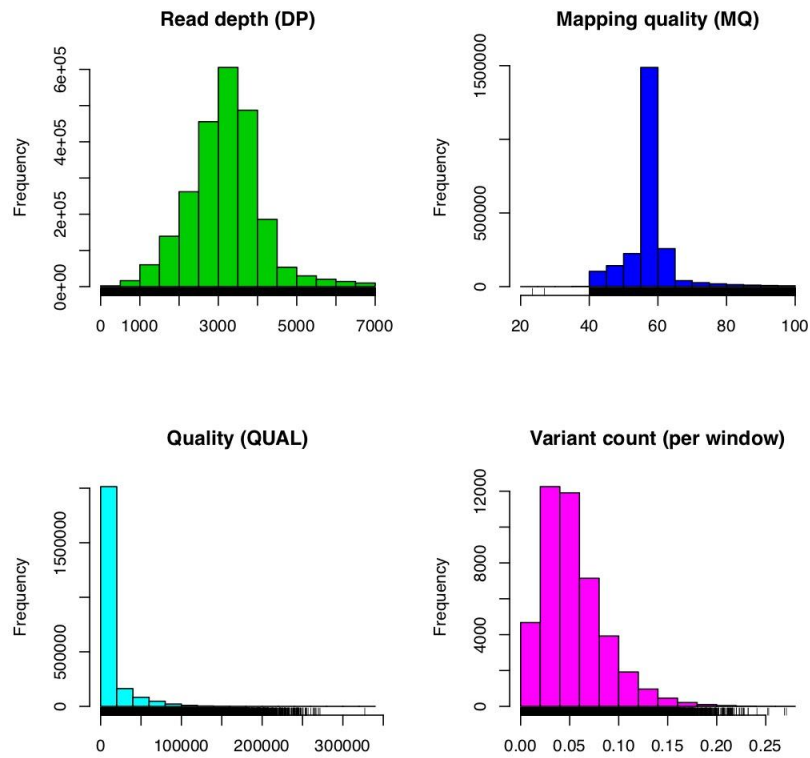


Figure III-3. Quality parameters of the filtered SNPs dataset used in this study. Legend: x-axes in DP, MQ and QUAL plot include both scores and their distribution. x-axes in “Variant count” plot indicates the variant incidence.

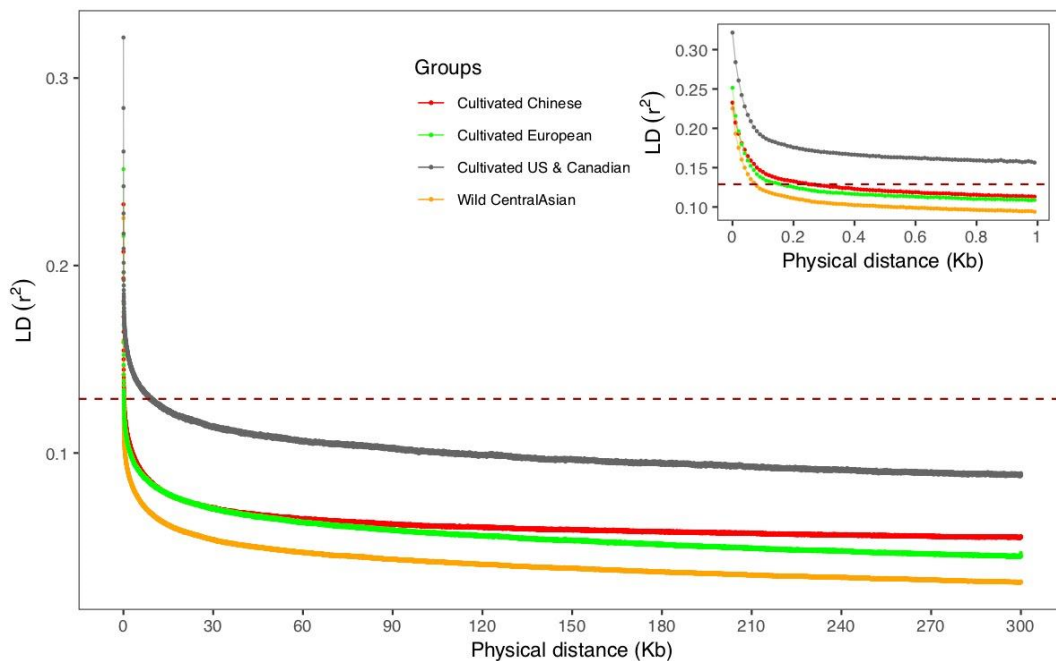


Figure III-4. Linkage disequilibrium decay as measured by the squared correlation coefficients (r^2) between all pairs of SNPs. Legend: The inner plot displays a higher resolution of LD in pairwise distances of < 1Kb. The red dotted lines in both the inner plot and the main chart refer to the 0.128 threshold.

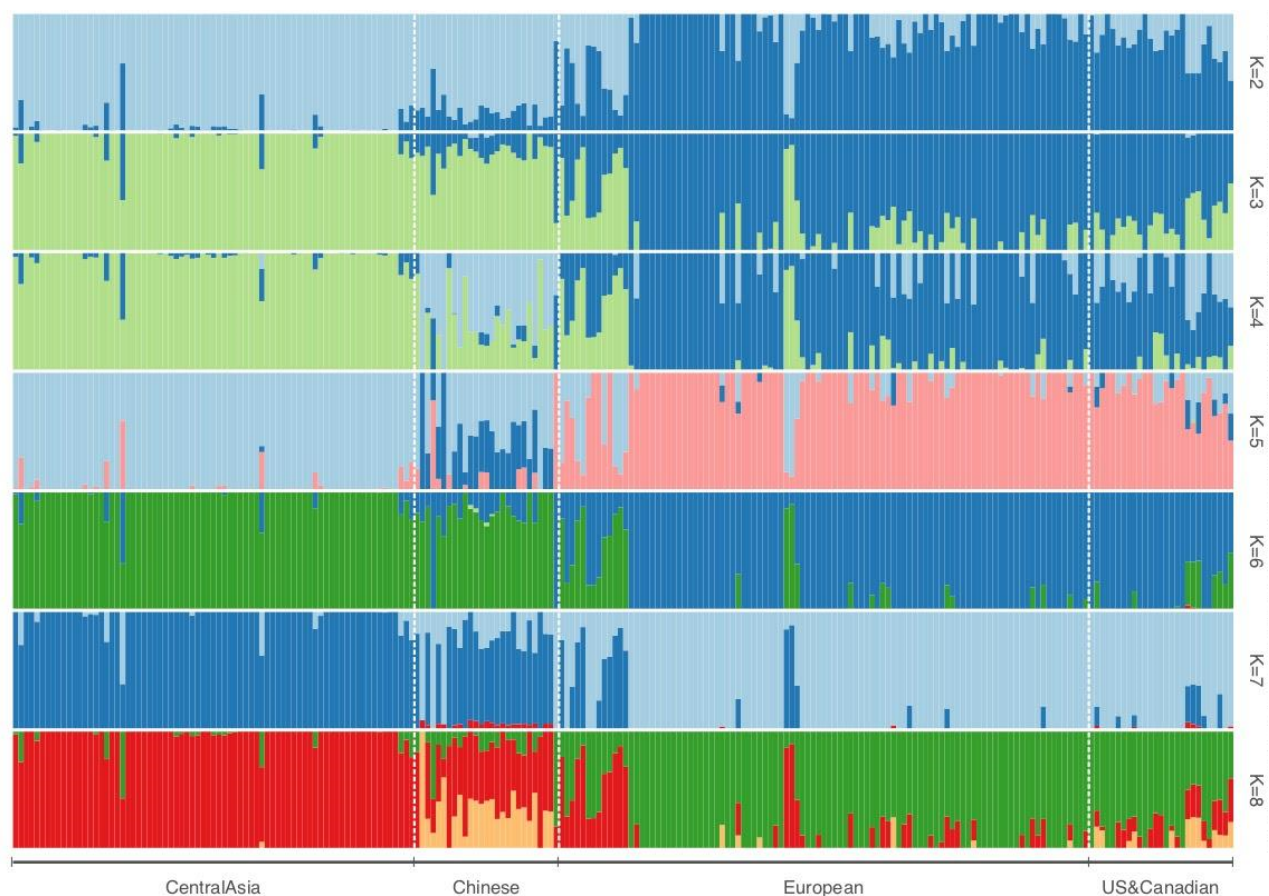


Figure III-5. Structure plots for Central Asian, Chinese, European and North-American apricots (from left to right) based on the unlinked SNPs dataset. Abbreviations under the structure bar plot indicate apricot groups as described in the Material and Methods.

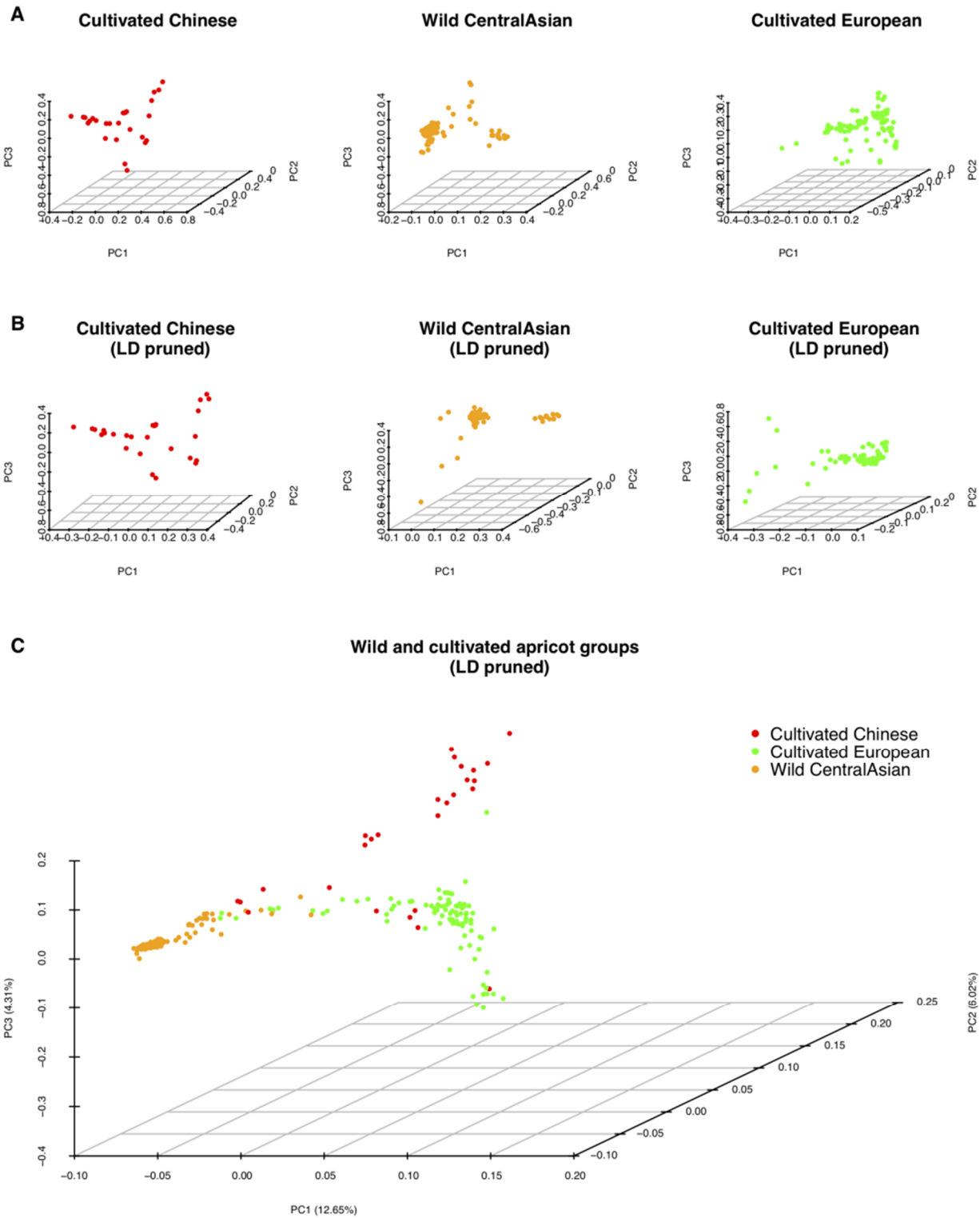


Figure III-6. Principal components of SNP variation obtained by using SmartPCA. Samples from the Central Asian wild (orange), Chinese (red) and European (green) apricot groups are shown. A) Principal components obtained from the whole SNP dataset, B) from LD pruned SNPs and C) a combination of the three apricot groups. Each dot represents one individual. The fraction of population variance explained by each principal component is in parentheses.

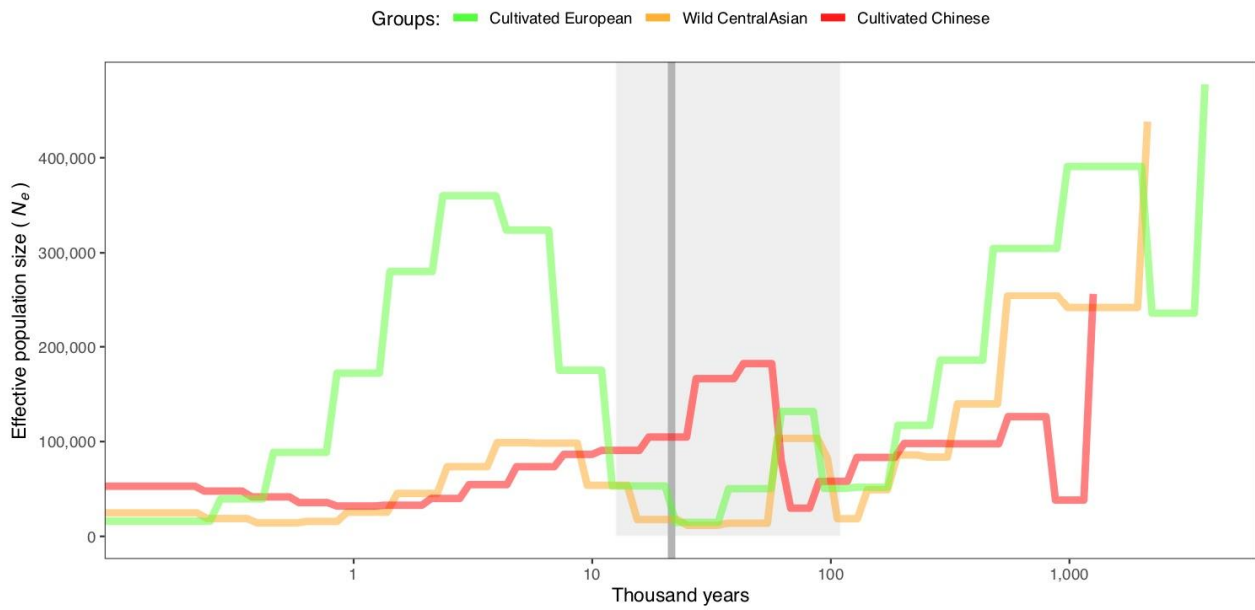


Figure III-7. SMC++-inferred demographic history of wild and cultivated apricots. Divergence time and past effective population size (N_e) changes inferred by the SMC++ analyses, based on unphased genotypes. The orange line represents the Central Asian wild apricots, the green line, the European cultivars and the red line, the Chinese cultivated apricots. The light grey vertical interval indicates the Quaternary glacial period (from 110,000 ya (years ago) to 12,500 ya); and the dark grey refers to the most recent maximum, the Last Glacial Maximum (LGM), that occurred about 20,000 ya. Generation time for both wild and cultivated apricots was set at five years, the mutation rate was $4.46e^9$ (Wang et al., 2017).

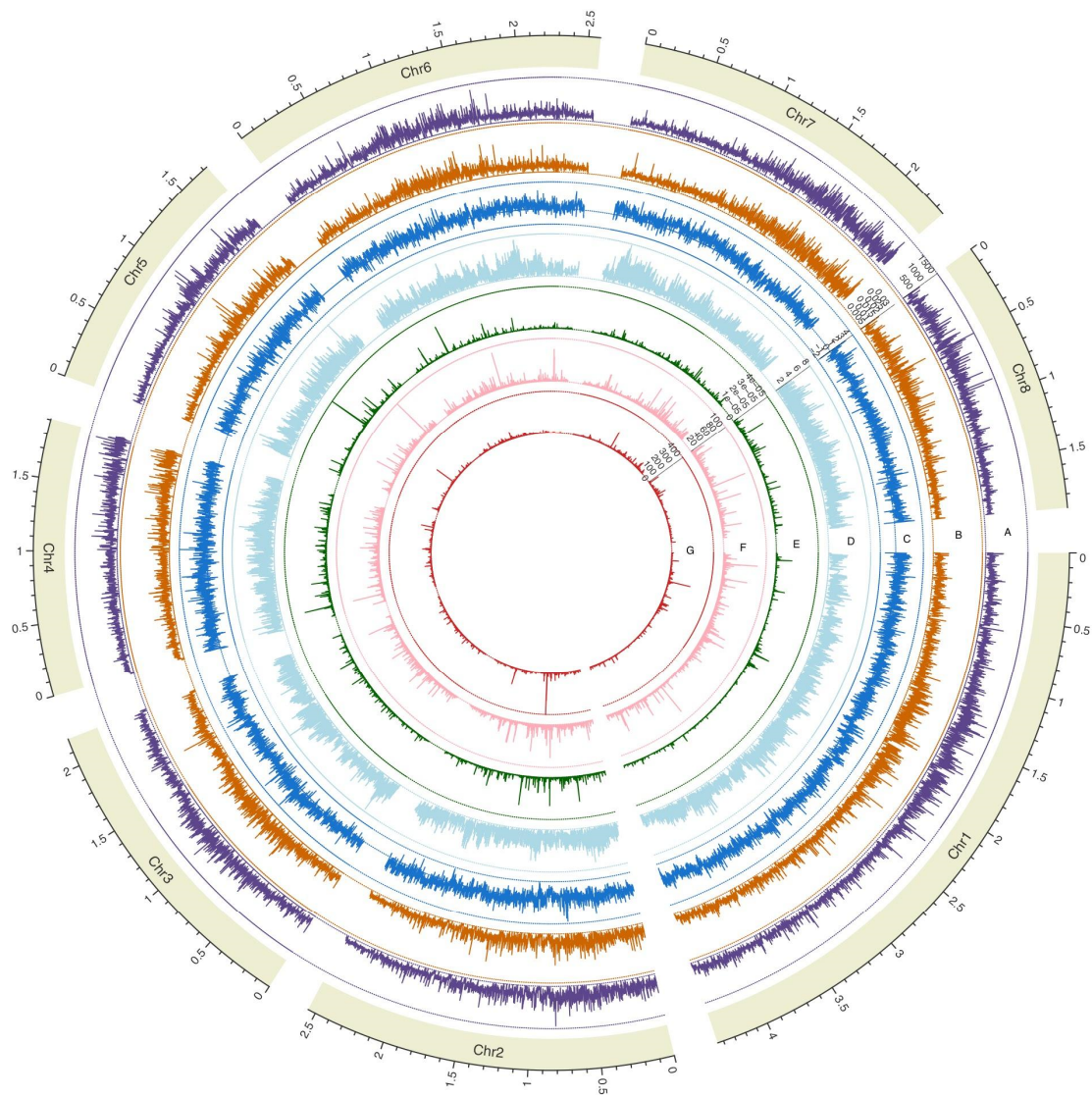


Figure III-8. Circos plot of the global distribution of SNP variants and signature of selective sweeps along the eight chromosomes of Central Asian wild apricots. Circles, from outside to inside, illustrate **A:** SNP density (purple); **B:** nucleotide diversity (π , brown); **C:** Tajima's D (blue); **D:** hapFLK (light blue); **E:** Mu (μ , green); **F:** Omega (pink); **G:** CLR values (red).

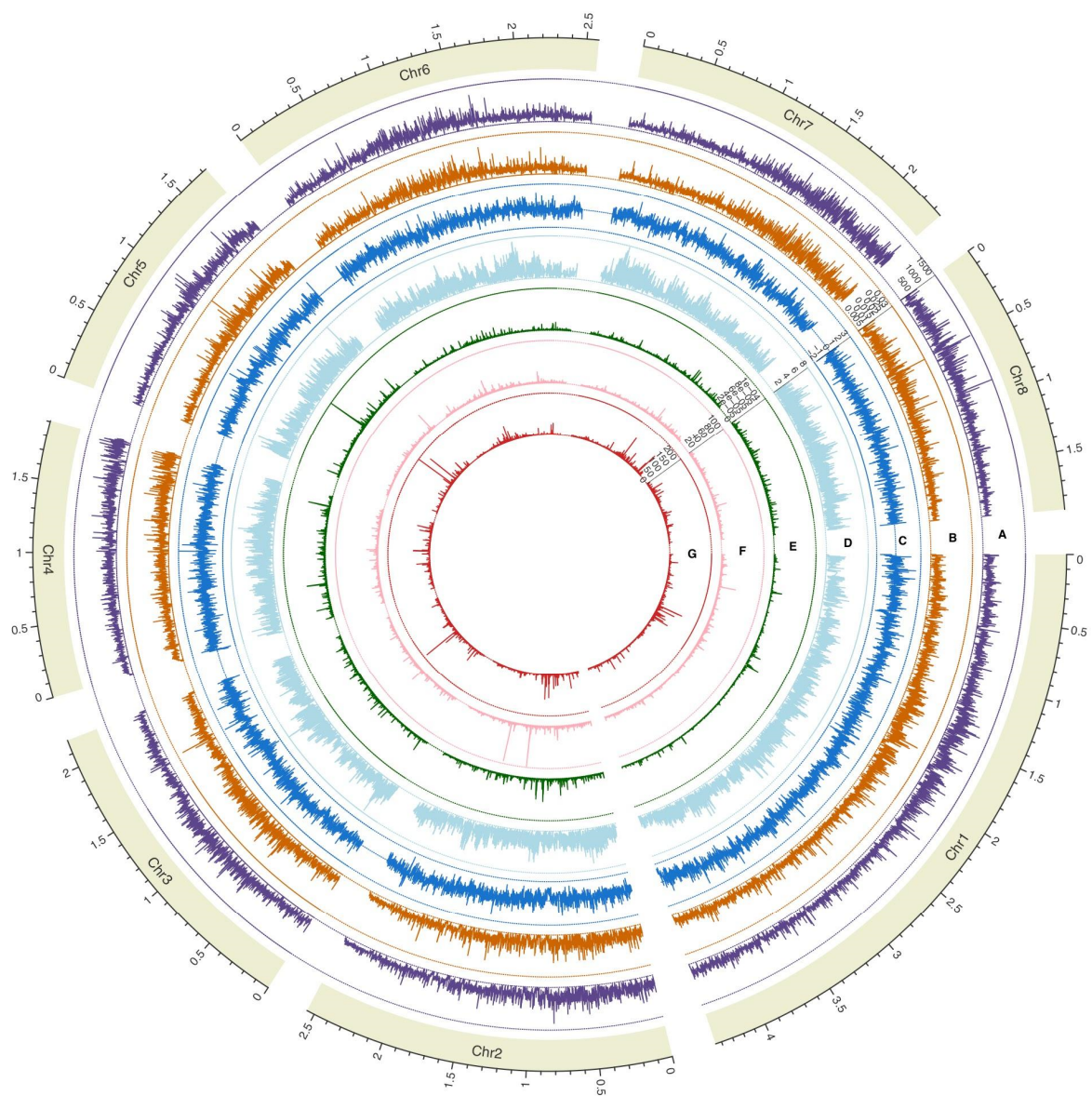


Figure III-9. Circos plot of the global distribution of SNP variants and signature of selective sweeps along the eight chromosomes of Chinese cultivated apricots. The circles in outside-in order were described in Figure III-8.

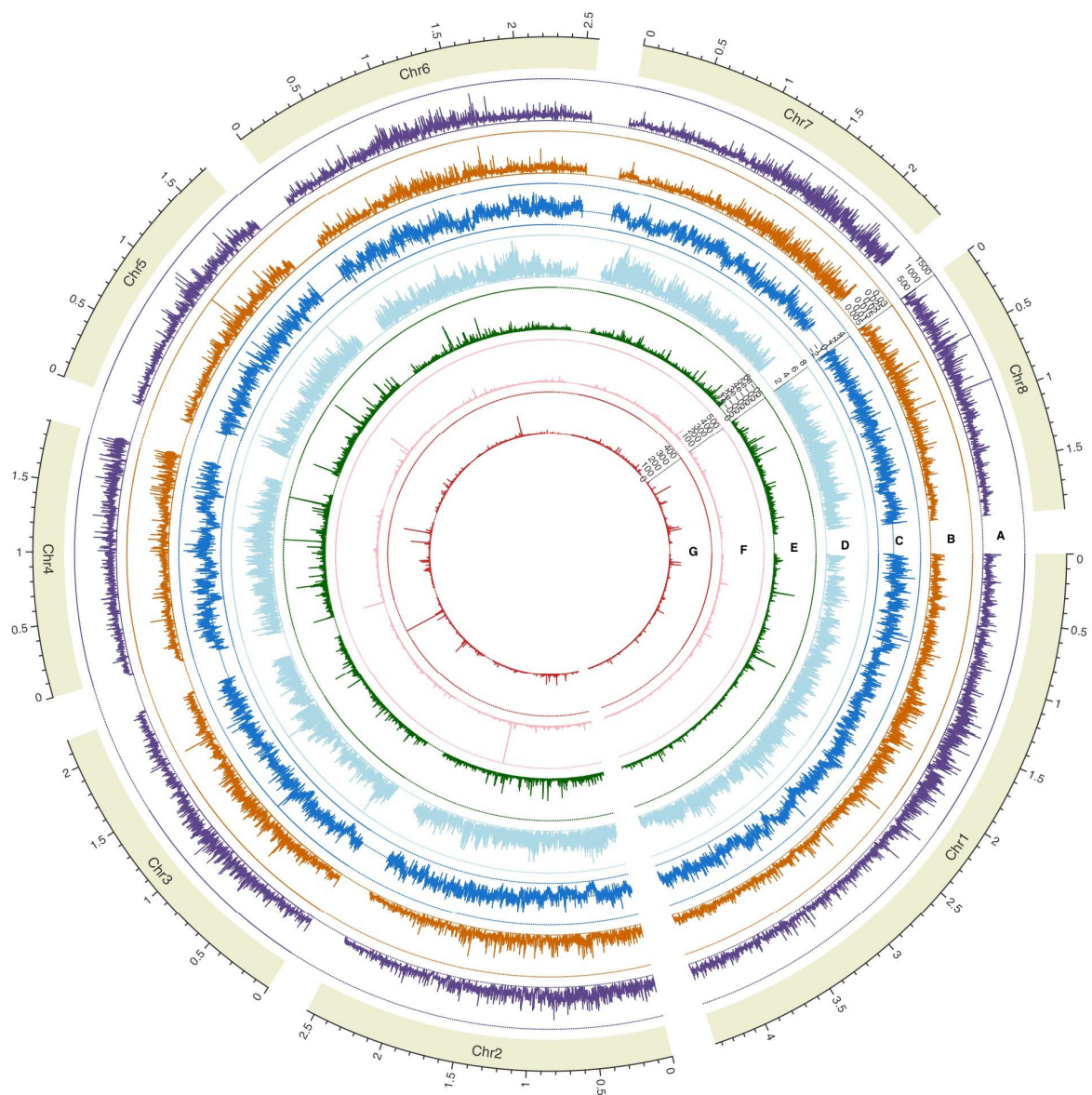


Figure III-10. Circos plot of the global distribution of SNP variants and signature of selective sweeps along the eight chromosomes of European cultivated apricots. The circles in outside-in order were described in Figure III-8.

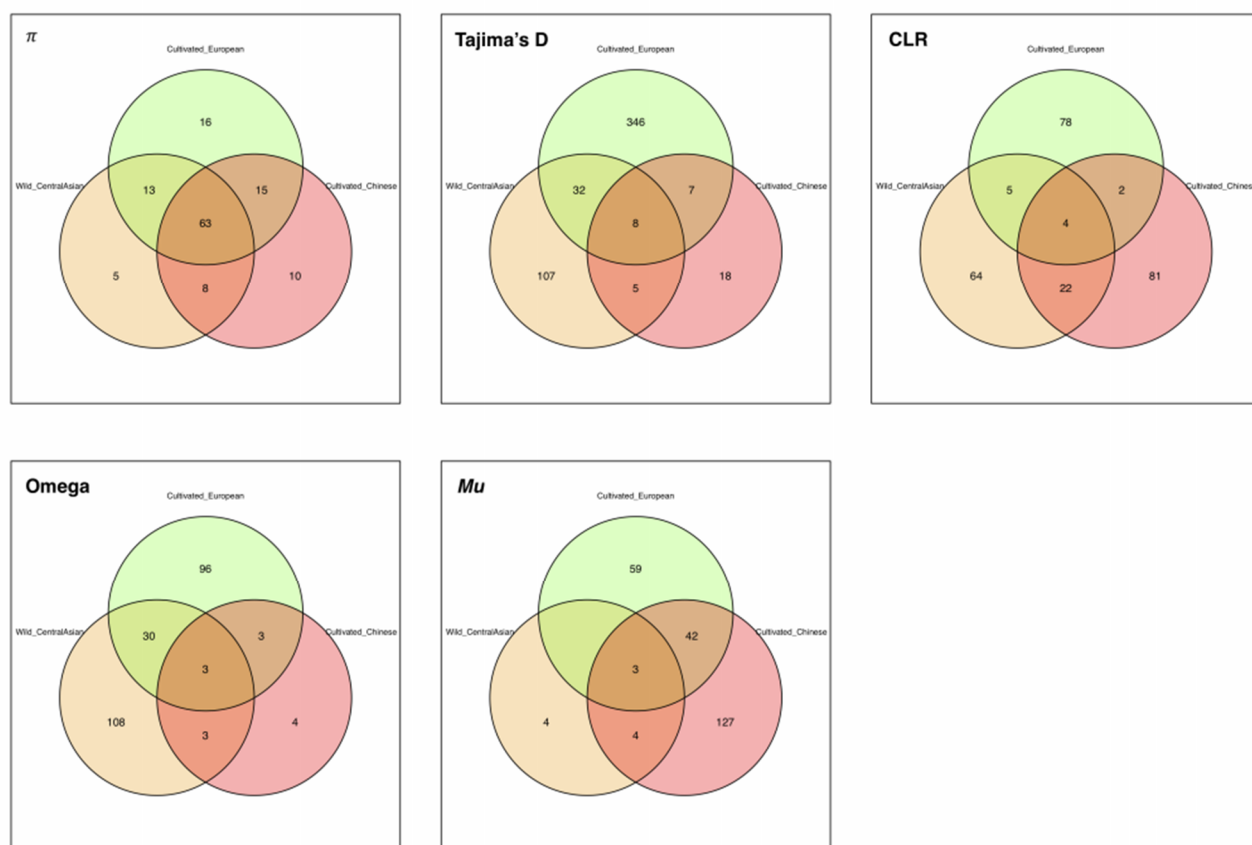


Figure III-11. Summary of the shared signals between the wild Central Asian and the cultivated Chinese and European apricots. Numbers in Venn diagrams indicate the number of unique or shared signals for each pair of apricot groups.

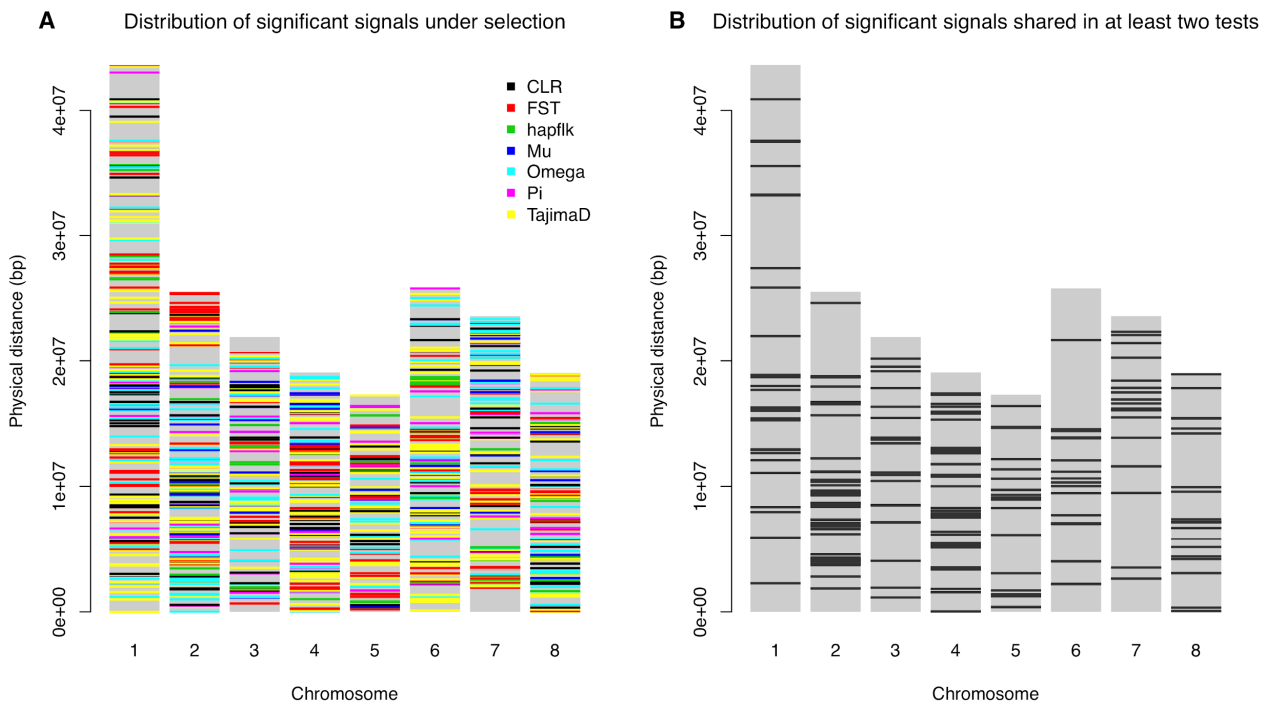


Figure III-12. Distribution of the most significant selection signals along the eight apricot chromosomes. A) Map location of the 1,753 signals distributed along the eight chromosomes; B) Map location of the genomic regions under selection which was identified by at least two tests.

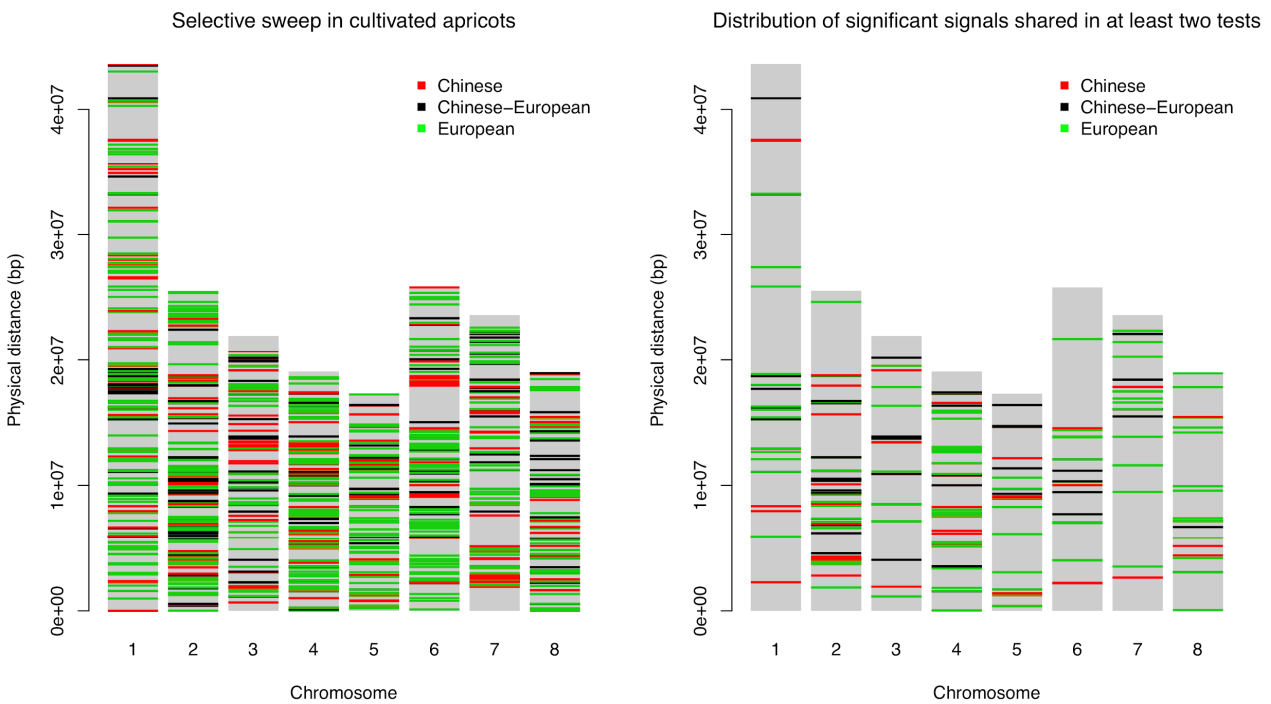


Figure III-13. Distribution of significant signals of selection in cultivated apricot genomes.

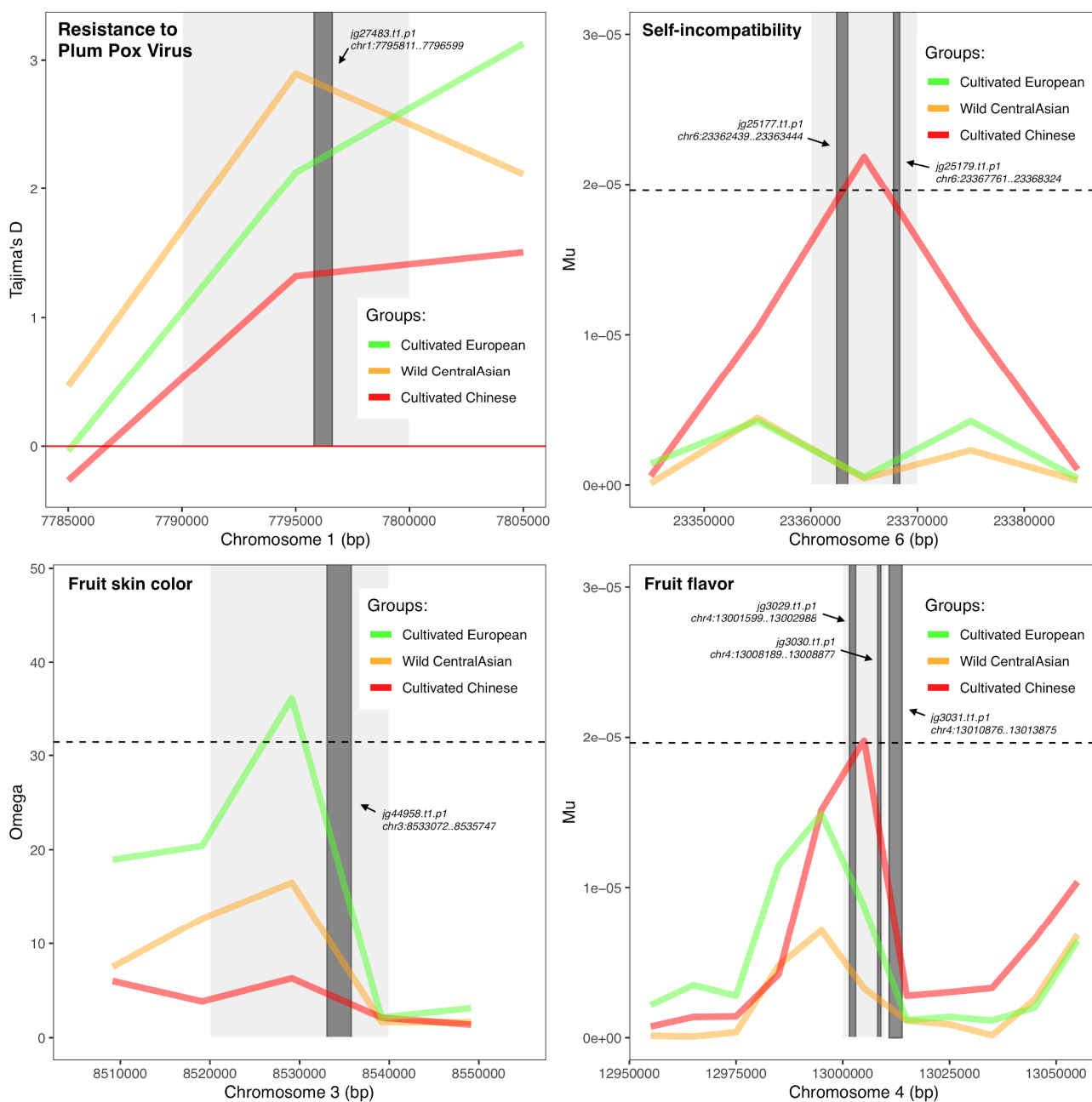


Figure III-14. Schematic diagrams of four selective sweeps and their closest relevant candidate genes. The vertical intervals in light grey indicate the position of the genomic region under selection while the dark grey intervals represent the position of the closest open reading frame(s). Gene IDs and genomic positions were retrieved from Marouch v2.0 genome browser. The red horizontal line was set at 0; the black dotted line represents the significance threshold for each test.

Tables

Table III-1. Three types of selective sweep detection methods.

Methods	Rationale	Schematic representation	Representative tests	References
Population differentiation based	Differences in allele frequencies, reflecting the population-specific action of selection, cause Wright's fixation index (F_{ST}) between two populations to increase, up to ~1 (fixation).		F_{ST} , hapFLK*	Shriver et al. (2004)
Local reduction of the polymorphism level, and frequency-based	When a favored allele and nearby genetic region sweep toward fixation, they shift the distribution of alleles in the population. The sweep causes a population-wide reduction of diversity and when new mutations appear they are generally rare (alleles that segregate at low frequency).		Nucleotide diversity (π), Tajima's D, Fay & Wu's H, CLR	Smith et al. (1974)
Linkage disequilibrium (LD) based	A selective sweep causes extended haplotype homozygosity (EHH) or LD decay which are both measure of LD, to rise across the haplotype that contains the selected allele. LD typically increases temporarily in regions under selection.		Ω (ω), hapFLK*	Kim and Nielsen (2004)

Footnote: * indicates the hapFLK method based on both population differentiation and the linkage disequilibrium.

Table III-2. Information and number of polymorphic loci identified by GATK and after quality filtering

Chromosome (Contig)	Length (bp)	Total raw SNPs	Total filtered SNPs	Av. distance per SNP (bp)	Filter/Raw (%)
Chr1	43,620,705	2,656,869	1,500,998	29	56.49%
Chr2	25,545,686	1,583,375	869,507	29	54.91%
Chr3	21,960,393	1,395,455	792,592	28	56.80%
Chr4	19,069,523	1,141,817	641,844	30	56.21%
Chr5	17,297,243	1,046,105	578,547	30	55.30%
Chr6	25,811,519	1,604,325	895,805	29	55.84%
Chr7	23,616,021	1,638,826	944,438	25	57.63%
Chr8	18,936,239	1,227,120	702,204	27	57.22%
Super-Scaffold_90	1,032,394	86,897			
Super-Scaffold_99	626,466	51,296			
Total	197,516,189	12,432,085	6,925,935		55.71%

Table III-3. Description of apricot linkage disequilibrium and their decays for each apricot chromosome.

Group	Chromosome	Mean r^2				LD decay	
		Min. value	Max. value	Median. value	sd	Threshold (1/2 max r^2)	Decay Distance (bp)
Wild CentralAsian	Chr1	0.032	0.221	0.025	0.010	0.111	185
	Chr2	0.032	0.234	0.026	0.010	0.117	175
	Chr3	0.033	0.224	0.027	0.011	0.112	205
	Chr4	0.031	0.243	0.030	0.011	0.122	205
	Chr5	0.031	0.228	0.027	0.011	0.114	225
	Chr6	0.029	0.227	0.027	0.011	0.114	175
	Chr7	0.028	0.207	0.023	0.010	0.104	135
	Chr8	0.028	0.217	0.026	0.010	0.109	155
	Mean	0.030	0.225	0.026	-	0.113	182.50
Cultivated Chinese	Chr1	0.055	0.231	0.009	0.042	0.115	805
	Chr2	0.055	0.240	0.009	0.044	0.120	695
	Chr3	0.055	0.237	0.010	0.044	0.119	925
	Chr4	0.052	0.241	0.009	0.046	0.121	475
	Chr5	0.053	0.227	0.009	0.041	0.114	825
	Chr6	0.056	0.238	0.010	0.045	0.119	945
	Chr7	0.054	0.219	0.009	0.039	0.110	1035
	Chr8	0.051	0.227	0.008	0.040	0.114	465
	Mean	0.054	0.233	0.009	-	0.116	771.25
Cultivated European	Chr1	0.044	0.245	0.011	0.024	0.122	205
	Chr2	0.041	0.251	0.011	0.025	0.126	205
	Chr3	0.043	0.244	0.011	0.026	0.122	225
	Chr4	0.051	0.273	0.012	0.032	0.137	255
	Chr5	0.043	0.251	0.012	0.026	0.125	245
	Chr6	0.043	0.259	0.012	0.028	0.130	165
	Chr7	0.042	0.233	0.010	0.021	0.117	155
	Chr8	0.042	0.255	0.011	0.025	0.127	185
	Mean	0.044	0.251	0.011	-	0.126	205.00
Cultivated US & Canadian	Chr1	0.091	0.325	0.012	0.096	0.162	625
	Chr2	0.080	0.327	0.012	0.091	0.163	545
	Chr3	0.086	0.309	0.011	0.087	0.155	1005
	Chr4	0.093	0.332	0.011	0.103	0.166	775
	Chr5	0.082	0.318	0.013	0.091	0.159	725
	Chr6	0.090	0.335	0.013	0.107	0.168	645
	Chr7	0.084	0.302	0.010	0.079	0.151	645
	Chr8	0.083	0.325	0.012	0.092	0.163	465
	Mean	0.086	0.322	0.012	-	0.161	678.75
Average	0.053	0.258	0.015		0.129		

Table III-4A. Summary of overlap between selective sweeps and annotated candidate genes or quantitative trait loci.

Selective sweep (SS) name	Detected with	Chr ¹	10Kb sliding window		SS detected in			Most interesting candidate gene(s) within the 10Kb sliding window ²	Position in Marouch v2.0	Annotated gene	NCBI BlastX ³	Function/Association	Reference
			Start position	End position	E U R	C H N	C A						
A) Selective sweep detected through at least 3 tests													
SS_chr1_8330000	TajimaD	1	8330000	8340000	█			ig27598.t1.p1	chr1:8334400..8338462	Claithrin assembly protein At5g35200 isoform X2 [<i>P. persica</i>]	1E-46 (91.91)	SNARE binding, clathrin coat assembly	Moshkanbaryans et al. (2016)
SS_chr1_8340000	TajimaD	1	8340000	8350000	█			ig27600.t1.p1	chr1:8341517..8344689	Putative disease resistance protein RGA3 [<i>P. persica</i>]	0E-0 (62.65)	Blight resistance protein that triggers a defense system which restricts the <i>Pseudomonas</i> growth.	Wang et al. (2008)
SS_chr1_8340001	FST	1	8340001	8350001	█	█		ig27602.t1.p1	chr1:8352992..8356710	Protein Sec61 subunit alpha [<i>Prunus persica</i>]	0.001 (50)	Transport of signal peptide-containing precursor polypeptides across endoplasmic reticulum (ER)	Hassdenteufel et al. (2018)
SS_chr1_8342654	CLR	1	8342654	8352654	█								
SS_chr2_6800000	Mu	2	6800000	6810000	█			ig38135.t1.p1	chr2:6800382..6800867	Protein FAR1-RELATED SEQUENCE 5-like [<i>Prunus avium</i>]	4E-68 (88.71)	Control of photoperiodic flowering	Lin et al. (2007)
SS_chr2_6800000	TajimaD	2	6800000	6810000	█			ig38137.t1.p1	chr2:6806140..6806396	Ubiquitin-like-specific protease ESD4		Control of flowering time by the SUMO pathway	Murtas et al. (2003)
SS_chr2_6804458	CLR	2	6804458	6814458	█								
SS_chr4_5200000	TajimaD	4	5200000	5210000	█			ig1295.t1.p1	chr4:5198857..5201776	WAK2 wall-associated receptor kinase 2 [<i>P. persica</i>]	2E-160 (71.23)		
SS_chr4_5200000	Mu	4	5200000	5210000	█			ig1296.t1.p1	chr4:5204317..5205609		2E-83 (51.79)	Binding to pectin that controls cell expansion, morphogenesis and development. Stress response or fruit/leaf cell expansion	Wagner et al. (2001)
SS_chr4_5200001	FST	4	5200001	5210001	█					35 copies of WAK2 from chr4:5040000..5455000			
SS_chr4_5300000	Mu	4	5300000	5310000	█			ig1316.t1.p1	chr4:5301389..5306530	Cyclic nucleotide-gated ion channel 1	2E-122 (79.40)	Involved in calcium signal transduction and tolerance to heavy metal	Sunkar et al. (2000)
SS_chr4_5300000	TajimaD	4	5300000	5310000	█			ig1317.t1.p1	chr4:5312406..5314676	WAK2 wall-associated receptor kinase 2 [<i>P. persica</i>]	0E-0 (85.47)	Binding to pectin that controls cell expansion	Wagner et al. (2001)
SS_chr4_5300001	FST	4	5300001	5310001	█								
SS_chr4_5320+A19:N21000	Mu	4	5320000	5330000	█			ig1321.t1.p1	chr4:5321390..5323822		7E-111 (62.41)	Binding to pectin that controls cell expansion, morphogenesis and development. Stress response or fruit/leaf cell expansion	Wagner et al. (2001)
SS_chr4_5320000	TajimaD	4	5320000	5330000	█					WAK2 wall-associated receptor kinase 2 [<i>P. persica</i>]			
SS_chr4_5320001	FST	4	5320001	5330001	█			ig1324.t1.p1	chr4:5337570..5340960		0E-0 (98.01)		
SS_chr4_8040000	TajimaD	4	8040000	8050000	█			ig1916.t1.p1	chr4:8038956..8043841	LRR receptor serine/threonine-protein kinase FLS2 [<i>P. avium</i>]	0E-0 (77.32)	FLS2 is involved in innate immunity.	Gomez-Gomez et al. (2000)
SS_chr4_8048254	Omega	4	8048254	8058254	█			ig1917.t1.p1	chr4:8042555..8045398	Receptor-like protein isoform X2 EIX2 [<i>P. avium</i>]	0E-0 (70.19)	Confers resistance to the fungal pathogen <i>T. viridie</i> through recognition of the EIX elicitor protein.	Ron et al. (2004)
SS_chr4_8048325	CLR	4	8048325	8058325	█			ig1918.t1.p1	chr4:8058081..8060929	LRR receptor serine/threonine-protein kinase FLS2 [<i>P. avium</i>]	0E-0 (80.20)	FLS2 is involved in innate immunity.	Gomez-Gomez et al. (2000)
SS_chr4_8050000	TajimaD	4	8050000	8060000	█								
SS_chr4_10790000	TajimaD	4	10790000	10800000	█								
SS_chr4_10790001	FST	4	10790001	10800001	█			ig2515.t1.p1	chr4:10790764..10792623	Tyrosine-sulfated glycopeptide receptor 1 [<i>P. persica</i>]	2E-97 (74.06)	Involved in plant immunity, with antagonistic effects on bacterial and fungal resistances	Mosher et al. (2013)
SS_chr4_10790994	Omega	4	10790994	10800994	█			ig2524.t1.p1 (<7Kb)	chr4:10807954..10808574				
SS_chr4_12810+A29:M31000	TajimaD	4	12810000	12820000	█								
SS_chr4_12810000	haplik	4	12810000	12820000	█			ig2991.t1.p1	chr4:12808707..12810301	Cyclic nucleotide-gated ion channel 4 [<i>Quercus suber</i>]	7E-82 (94.77)	Component of the signaling pathways leading to hypersensitive response (HR)	Balague et al. (2003)
SS_chr4_12810001	FST	4	12810001	12820001	█								
SS_chr4_13000000	Mu	4	13000000	13010000	█			ig3029.t1.p1	chr4:13001599..13002988		4E-91 (71.35)		
SS_chr4_13000000	TajimaD	4	13000000	13010000	█			ig3030.t1.p1	chr4:13008189..13008877	(-)-alpha-pinene synthase-like [<i>P. persica</i>]	2E-74 (95.60)	Fruit flavor	Aharoni et al. (2004)
SS_chr4_13003335	CLR	4	13003335	13013335	█			ig3031.t1.p1	chr4:13010876..13013875		4E-107 (83.33)		
SS_chr4_17357554	Omega	4	17357554	17367554	█								
SS_chr4_17357738	CLR	4	17357738	17367738	█			ig4143.t1.p1 (12Kb)	chr4:17343390..17345449	LRR receptor serine/threonine-protein kinase FLS2 [<i>P. avium</i>]	1E-34 (69.62)	FLS2 is involved in innate immunity.	Gomez-Gomez et al. (2000)
SS_chr4_17360000	Mu	4	17360000	17370000	█			ig4151.t1.p1	chr4:17362456..17365932	Callose synthase 10-like [<i>Juglans regia</i>]	1E-10 (51.76)	Male gametophyte development	Záveská et al. (2017)
SS_chr4_17360000	Mu	4	17360000	17370000	█							Pathogen response	Ellinger and Voigt (2014)
SS_chr5_9110000	TajimaD	5	9110000	9120000	█								
SS_chr5_9110000	Mu	5	9110000	9120000	█			ig12888.t1.p1	chr5:9122145..9127015	NPR4 ankyrin repeat-containing protein isoform X4 [<i>P. persica</i>]	9E-82 (87.04)	Involved in salt stress tolerance and resistance to herbivore	Yuan et al. (2007)
SS_chr5_9110001	FST	5	9110001	9120001	█								
SS_chr5_14730000	Mu	5	14730000	14740000	█			ig14209.t1.p1	chr5:14738597..14738968	Cyclic nucleotide-gated ion channel 1-like [<i>P. avium</i>]	1E-23 (91.23)	Component of the signaling pathways leading to hypersensitive response (HR)	Zhu et al. (2010)
SS_chr5_14734233	Omega	5	14734233	14744233	█			ig14211.t1.p1	chr5:14742272..14743175		8E-53 (40.36)	Activation of TIR-NB-LRR-mediated immune responses and control of photoperiodic flowering	Goraloppia et al. (2017)
SS_chr5_14734301	CLR	5	14734301	14744301	█			ig14213.t1.p1 (5Kb)	chr5:14749524..14754134		3E-32 (54.84)		
SS_chr7_17490000+A43:N45	Mu	7	17490000	17500000	█								
SS_chr7_17490000	TajimaD	7	17490000	17500000	█			ig8862.t1.p1	chr7:17499703..17500332	TMV resistance protein N-like [<i>P. avium</i>]	5E-30 (64.06)	Resistance to Tobacco Mosaic virus	Marathe et al. (2002)
SS_chr7_17492278	Omega	7	17492278	17502278	█								
SS_chr8_3050000	TajimaD	8	3050000	3060000	█								
SS_chr8_3052974	CLR	8	3052974	3062974	█			ig15542.t1.p1	chr8:3047541..3052196	Cycling DOF factor 2 [<i>P. yedoensis</i> var. <i>nudiflora</i>]	4E-50 (66.67)	Control of photoperiodic flowering by repressing 'CONSTANS' expression.	Fomara et al. (2009)
SS_chr8_3052985	Omega	8	3052985	3062985	█								

Footnote: ¹ Chr: chromosome, ² in brackets, distance from the closest selective sweep when the CG is not mapping within the 10Kb sliding window, ³ NCBI BlastX Results: e-value (percent identity), ⁴ QTL name as referred in the Genome database of Rosaceae (in brackets, closest molecular marker linked to the SS and distance in Kb)

CHAPTER III

Table III-4B. Summary of overlap between selective sweeps and annotated candidate genes or quantitative trait loci.

Selective sweep (SS) name	Detected with	Chr	10Kb sliding window		SS detected in			Most interesting candidate gene(s) within the 10Kb sliding window ²	Position in Marouch v2.0	Annotated gene	NBI BlastX ³	Within the potential QTL interval ⁴	Function/Association	Reference
			Start position	End position	EU R	CH N	CA							
SS_chr1_7790000	TajimaD	1	7790000	7800000				jg27483.t1.p1	chr1:7795811..7796599	PPVres locus MATH-domain containing protein [<i>P. armeniaca</i>]	1E-44 (98.31)	PPVres for resistance to PPV (ZP002, 0Kb)	Resistance to Plum Pox Virus	Zuriaga et al. (2013); Decroocq et al. (2014)
SS_chr1_23856202	CLR	1	23856202	23866202				jg31755.t1.p1	chr1:23856997..23858343	Protein FAR-RED IMPAIRED RESPONSE 1 [<i>Arabidopsis thaliana</i>]	3E-35 (41.67)	qDBD.HAxRM-ch1 for bud-breaking (UDAp-414, 80Kb)	Regulation of circadian rhythm, control of flowering time through the photoperiod pathway	Lin et al.(2007)
SS_chr2_5980001	Pi	2	5980001	5990000				jg37916.t1.p1 (26Kb)	chr2:5950042..5954356	TMV resistance protein N isoform X1 [<i>P. persica</i>]	0E-0 (86.24)	RMia for rootknot nematod (AMPA117, 75Kb)	Resistance to rootknot nematods	Duval et al (2014)
SS_chr3_8520000	TajimaD	3	8520000	8530000				jg44958.t1.p1	chr3:8533072..8535747	Putative Myb family transcription factor [<i>P. persica</i>]	3E-51 (92.56)	qFSC.HAxRM-ch3 for fruit skin color (UDAp-446, 2800Kbp)	Fruit skin color	Jin et al. (2016)
SS_chr3_8524133	Omega	3	8524133	8534133										
SS_chr4_2720000	TajimaD	4	2720000	2730000				jg708.t1.p1 (9Kb)	chr4:2739149..2742072	Transcriptional corepressor LEUNIG LUG [<i>P. persica</i>]	3E-49 (95.71)		Control of early stages of floral meristem development	Sitaraman et al. (2008)
								jg721.t1.p1 (46Kb)	chr4:2776139..2777017	DOF zinc finger protein DOF1.4 [<i>P. persica</i>]	1E-149 (97.26)	qDBD.HAxRM-ch4 (PaCITA25, 757Kb)	Control of photoperiodic flowering	Fornara et al. (2009)
								jg722.t1.p1 (54Kb)	chr4:2784757..2786087	DOF zinc finger protein DOF1.4-like [<i>P. avium</i>]	1E-147 (97.06)		Control of photoperiodic flowering	
SS_chr4_12560000	hapfik	4	12560000	12570000				jg2931.t1.p1	chr4:12494365..12495326	NAC domain-containing protein NAC072 [<i>P. persica</i>]	6E-119 (95.70)	qP-FF4.1 for fruit firmness (ss490552928, 22Kb)	Response to drought stress	Tran et al. (2004)
SS_chr6_23360000	Mu	6	23360000	23370000				jg25177.t1.p1	chr6:23362439..23363444	Putative ribonuclease H protein At1g65750 RNase-1 [<i>Rosa chinensis</i>]	0E-0 (41.43)		Control of gametophytic self-incompatibility: pistil S determinant S-ribonuclease gene	Matsumoto and Tao (2016)
								jg25179.t1.p1	chr6:23367761..23368324	S haplotype-specific F-box SFbc gene [<i>P. armeniaca</i>]	3E-145 (100)	SI Self-incompatibility locus	Control of gametophytic self-incompatibility: pollen S determinant F-box gene	
SS_chr7_8436653	CLR	7	8436653	8446653				jg6649.t1.p1	chr7:8435048..8437765	EMB175 pentatricopeptide repeat-containing protein At5g03800 [<i>P. persica</i>]	0 (97.68)	qDBD.HAxRM-ch7 for bud breaking (UDAp-407, 19.5Kb)	Embryo development and seed dormancy	

Table III-4C. Summary of overlap between selective sweeps and annotated candidate genes or quantitative trait loci.

Selective sweep (SS) name	Detected with	Chr ¹	10Kb sliding window		SS detected in			Most interesting candidate gene(s) within the 10Kb sliding window ²	Position in Marouch v2.0	Annotated gene	NBI BlastX ³	Function/Association	Reference
			Start position	End position	EUR	CHN	CA						
C) Selective sweep linked to known candidate genes													
SS_chr1_25060000	TajimaD	1	25060000	25070000	■			jg32067.t1.p1 (9.7Kb)	chr1:25047811..25050257	Floral homeotic protein APETALA 1 [<i>P. pseudocerasus</i>]	6E-29 (100)	Transition of an inflorescence meristem into a floral meristem	Bowman et al.(1993)
SS_chr1_31120000	TajimaD	1	31120000	31130000			■	jg33566.t1.p1 (8.3Kb)	chr1:31111681..31112803	Zinc finger protein CONSTANS-LIKE 4	9E-141 (98.71)	Control of photoperiodic flowering	Shim et al. (2017)
SS_chr1_39090000	TajimaD	1	39090000	39100000			■	jg35412.t1.p1	chr1:39095998..39107201	Dormancy-associated MADS-box DAM6, AGL24 gene [<i>P. persica</i>]	3E-19 (79.45)	Dormancy release	Leida et al. (2010)
SS_chr2_2451745	Omega	2	2451745	2461745	■			jg37049.t1.p1 (5Kb)	chr2:2466071..2466800	Predicted disease resistance protein RPS2 [<i>P. mume</i>]	0E-0 (76.18)	Resistance to <i>Pseudomonas syringae</i>	Mackey et al.(2003)
SS_chr2_13048216	Omega	2	13048216	13058216			■	jg39797.t1.p1	chr2:13049856..13052675	TMV resistance protein N [<i>P. persica</i>]	0E-0 (77.15)	Disease resistance protein which restricts TMV (tobacco mosaic virus) growth	Dinesh et al. (2000)
SS_chr2_15369608	Omega	2	15369608	15379608			■	jg40300.t1.p1	chr2:15376149..15379073	MDIS1-interacting receptor like kinase 2 [<i>P. persica</i>]	0E-0 (88.17)	Involved in the pollen tube perception of the female signal	Wang et al. (2016)
SS_chr2_15370000	Mu	2	15370000	15380000	■	■							
SS_chr3_15447359	CLR	3	15447359	15457359			■	jg46751.t1.p1	chr3:15455506..15457193	7-deoxyloganetin glucosyltransferase [<i>P. persica</i>]	0E-0 (97.74)	Up-regulated during the early stage of fruit ripening, by methyl jasmonate	Nagatoshi et al. (2011)
SS_chr3_15447593	Omega	3	15447593	15457593			■						
SS_chr4_11780000	TajimaD	4	11780000	11790000			■	jg2778.t1.p1 jg2779.t1.p1	chr4:11784942..11785829 chr4:11787872..11789367	LECRK L-type lectin-domain containing receptor kinase IX.1 [<i>P. persica</i>]	0E-0 (94.53) 0E-0 (92.10)	Resistance to oomycetes <i>Phytophthora infestans</i> and <i>Phytophthora capsici</i> .	Wang et al. (2015)
SS_chr4_11780001	FST	4	11780001	11790001	■								
SS_chr4_12640000	TajimaD	4	12640000	12650000	■			jg2962.t1.p1	chr4:12641236..12647639	ADP-ribosylation factor GTPase-activating AGD5 protein [<i>Citrus clementina</i>]	9E647 (67.44)	Required for floral organ shedding	Liu et al. (2013)
SS_chr4_12642971	CLR	4	12642971	12652971	■			jg2963.t1.p2	chr4:12652159..12653520	CFE protein [<i>Gossypium herbaceum</i>]	4E-43 (63.16)	Role in cotton fibre cell initiation and elongation	Lv et al. (2015)
SS_chr4_15990000	TajimaD	4	15990000	16000000	■			jg3765.t1.p1 (31Kb)	chr4:16031078..16033096	FRL1 FRIGIDA-like protein 1 [<i>P. persica</i>]	0E-0 (96.86)	Flowering repressor	Michaels et al. (2004)
SS_chr5_6146311	CLR	5	6146311	6156311	■			jg12150.t1.p1 (20Kb)	chr5:6178600..6177190	DOF zinc finger protein DOF4.6 [<i>P. persica</i>]	7E-142 (95.50)	Control of photoperiodic flowering	Fornara et al. (2009)
SS_chr5_14650000	Mu	5	14650000	14660000		■		jg14193.t1.p1	chr5:14658554..14659126	Topless-related protein 1 [<i>Arabidopsis thaliana</i>]	9E-29 (52.83)	Co-repressor TF acting as hub protein in many developmental and response pathways such as plant immune response and control of photoperiodic flowering	Zhu et al. (2010)
SS_chr5_14654227	CLR	5	14654227	14664227		■							
SS_chr6_10634704	Omega	6	10634704	10644704			■	jg22056.t1.p1	chr6:10642783..10643703	LRR receptor-like serine/threonine-protein kinase FLS2 [<i>P. avium</i>]	6E-31 (66.32)	Initiation of the innate immune MAP kinase signaling cascade	Gomez-Gomez and Boller (2000)
SS_chr6_10634848	CLR	6	10634848	10644848			■						
SS_chr7_4410000	TajimaD	7	4410000	4420000	■			jg5679.t1.p1 jg5680.t1.p1	chr7:4409658..4411079 chr7:4412539..4413933	Leucine-rich repeat receptor-like protein kinase PXL2 [<i>P. yedoensis</i>] Leucine-rich repeat receptor-like protein kinase PXL2 [<i>P. yedoensis</i>]	2E-04 (30.85) 4E-05 (35.87)	Regulation of vascular-tissue development and male gametophyte production	Mou et al. (2017)
SS_chr7_8970001	FST	7	8970001	8980001	■			jg6748.t1.p1	chr7:8982953..8984194	TFL1, CEN-like protein 2 [<i>P. avium</i>]	8E-39 (98.61)	Photoperiod-regulated floral repressor	Mohamed et al. (2010)
SS_chr7_10060000	TajimaD	7	10060000	10070000			■	jg6985.t1.p1 (6Kb)	chr7:10052389..10053330	DOF zinc finger protein DOF3.5 [<i>P. persica</i>]	0E-0 (93.93)	Control of photoperiodic flowering	Fornara et al. (2009)
SS_chr8_1630000	hapflk	8	1630000	1640000	■	■	■	jg15194.t1.p1 (8.5Kb)	chr8:1648550..1647590	DOF zinc finger protein DOF1.2 [<i>P. persica</i>]	0E-0 (95.75)	Control of photoperiodic flowering	Fornara et al. (2009)

CHAPTER III

Table III-5. Convergent candidate, selective sweep regions and genes shared between European cultivated and Chinese cultivated apricots.

Selective sweep (SS) name	detected with	Chr ¹	Start position	End position	EUR	CHN	Most interesting candidate gene(s) within the 10Kb sliding window ²	Position in Marouch v2.0	Annotated gene	NBI BlastX ³	Function/Association	Reference
SS_chr1_0	TajimaD	1	0	10000								
SS_chr1_1560000	TajimaD	1	1560000	1570000			jg26098.t1.p1 jg26099.t1.p1	chr1:1559031..1560952 chr1:1565811..1570094	UGT78G1 Flavonoid 3-O-glucosyltransferase [<i>P. persica</i>] HOP2 Hsp70-Hsp90 organizing protein 2 [<i>P. avium</i>]	0E-0 (92.55) 6E-28 (96.77)	Regulates anthocyanin biosynthesis during fruit coloration Heat-shock protein, involved in stress response	Zhao et al. (2012)
SS_chr1_5890000	Mu	1	5890000	5900000			jg27095.t1.p1	chr1:5893720..5893997	TPR1 Topless-related protein 1	1E-06 (91.67)	Activation of TIR-NB-LRR-mediated immune responses and control of photoperiodic flowering	Goraloppia et al (2017)
SS_chr1_5900000	Mu	1	5900000	5910000			jg27087.t1.p1	chr1:5898636..5900122	S-locus-specific glycoprotein-like [<i>P. persica</i>]	1E-30 (91.30)	Involved in sporophytic self-incompatibility system (the inability of flowering plants to achieve self-fertilization)	
SS_chr1_5910000	Mu	1	5910000	5920000			jg27089.t1.p1	chr1:5905246..5906369	uncharacterized protein LOC107881929 [<i>P. mume</i>]	6E-126 (86.22)		
SS_chr1_5960001	Pi	1	5960001	5970000			jg27099.t1.p1 jg27100.t1.p1	chr1:5965630..5966611 chr1:5967482..5968367	LecRK1.9 Concanavalin A-like lectin protein kinase family protein [<i>Arabidopsis thaliana</i>] Digeranylglyceranyl glycerol phosphate synthase [<i>Actinidia chinensis</i> var. <i>chinensis</i>]	2.6 (46.88) 5E-41 (62.30)	Role in cell wall-plasma membrane adhesion. Involved in Phytophthora resistance. Involved in the pathway glycerophospholipid metabolism, which is part of Membrane lipid metabolism	Bouwmeester et al (2011)
SS_chr1_15264165	Omega	1	15264165	15274165			jg29360.t1.p1	chr1:15265298..15266031	eukaryotic translation initiation factor 5B-like [<i>P. avium</i>]	2E-30 (45.61)	Part of the translation initiation complex	
SS_chr1_15264194	CLR	1	15264194	15274194								
SS_chr1_15600001	Pi	1	15600001	15610000								
SS_chr1_16160000	Mu	1	16160000	16170000			jg29597.t1.p1	chr1:16167840..16169423	LRR receptor-like serine/threonine-protein kinase FLS2 [<i>P. avium</i>]	1E-33 (41.83)	Initiation of the innate immune MAP kinase signaling cascade	Gomez-Gomez and Bolter (2000)
SS_chr1_16170000	Mu	1	16170000	16180000			jg29601.t1.p1	chr1:16176923..16179903	WAKL8 Wall-associated receptor kinase-like 8 [<i>P. mume</i>]	0E-0 (98.16)	Involved in response to pathogens and required for cell wall expansion	Verica and He (2002)
SS_chr1_16180000	Mu	1	16180000	16190000			jg29602.t1.p1	chr1:16181597..16182343	Uncharacterized mitochondrial protein AtMg00810			
SS_chr1_17990000	Mu	1	17990000	18000000			jg30140.t2.p1	chr1:17996063..17999338	Probable LRR receptor-like serine/threonine-protein kinase At3g47570 [<i>P. persica</i>]	0E-0 (89.13)	Potentially involved in defense response to pathogens	
SS_chr1_17994712	Omega	1	17994712	18004712								
SS_chr1_18700000	TajimaD	1	18700000	18710000								
SS_chr1_18704960	CLR	1	18704960	18714960								
SS_chr1_36859000	CLR	1	36859000	36869000			jg34929.t1.p1	chr1:36861072..36866725	LACS9 Long chain acyl-CoA synthetase 9, chloroplastic [<i>P. mume</i>]	3E-100 (64.86)	Activation of long-chain fatty acids (that make up the common structural and storage lipids) synthesis. Involved in seed oil biosynthesis	Schnurr et al (2002)
SS_chr1_36860000	TajimaD	1	36860000	36870000			jg34930.t1.p1	chr1:36868697..36874939	SGT small glutamine-rich tetratricopeptide repeat-containing protein [<i>P. persica</i>]	2e-58 (100)	Co-chaperone, involved in the regulation of the endoplasmic reticulum-associated	
SS_chr1_37530001	Pi	1	37530001	37540000			jg35072.t1.p1	chr1:37534222..37535649	Putative F-box protein At5g49610 [<i>Rosa chinensis</i>]	6E-100 (44.12)	Involved in the pathway of protein ubiquitination, which is part of Protein modification.	
SS_chr2_6590000	Mu	2	6590000	6600000			jg38088.t1.p1	chr2:6588839..6589462	Hypothetical protein Pyn_15652 [<i>P. yedoensis</i> var. <i>nudiflora</i>]	3E-15 (60.42)		
SS_chr2_6590000	TajimaD	2	6590000	6600000								
SS_chr2_6800000	TajimaD	2	6800000	6810000			jg38135.t1.p1	chr2:6800382..6800867	Protein FAR1-RELATED SEQUENCE 5-like [<i>P. avium</i>]	4E-68 (88.71)	Control of photoperiodic flowering	Lin et al (2007)
SS_chr2_6800000	Mu	2	6800000	6810000			jg38137.t1.p1	chr2:6806116..6809640	Ubiquitin-like-specific protease ESD4	+ (58.9)	Control of flowering time by the SUMO pathway	Murtas et al (2003)
SS_chr2_6804458	CLR	2	6804458	6814458								
SS_chr2_8650001	Pi	2	8650001	8660000			jg38698.t1.p1	chr2:8649128..8651996	Putative protein-serine/threonine phosphatase [<i>Rosa chinensis</i>]	5E-42 (32.22)		
SS_chr2_8655565	CLR	2	8655565	8665565			jg38701.t1.p1	chr2:8654691..8656080	Ankyrin repeat-containing protein NPR4 [<i>P. persica</i>]	1E-118 (94.55)	Involved in disease resistance and salt stress tolerance	Yuan et al (2007)
SS_chr2_8660000	Mu	2	8660000	8670000			jg38703.t1.p1	chr2:8658818..8661100	Ankyrin repeat-containing protein NPR4 isoform X4 [<i>P. persica</i>]	4E-49 (52.81)	Involved in disease resistance and salt stress tolerance	Yuan et al (2007)
SS_chr2_8660001	Pi	2	8660001	8670000			jg38704.t1.p1	chr2:8668080..8669584	Zinc finger protein 36, C3H1 type-like 3 [<i>P. mume</i>]	1E-09 (46.99)	most likely functions in regulating the response to growth factors	
SS_chr2_8670000	Mu	2	8670000	8680000			jg38706.t1.p1	chr2:8675571..8677554	Ubiquitin-like-specific protease ESD4 [<i>P. avium</i>]	3E-05 (51.61)	Control of flowering time by the SUMO pathway	Murtas et al (2003)
SS_chr2_8680000	Mu	2	8680000	8690000			jg38710.t1.p1 jg38713.t1.p1	chr2:8685272..8687636 chr2:8688990..8690741	Protein FAR-RED IMPAIRED RESPONSE 1-like [<i>Prunus mume</i>] Protein FAR1-RELATED SEQUENCE 5-like [<i>P. persica</i>]	2E-49 (89.60) 6E-116 (42.98)	Control of photoperiodic flowering Control of photoperiodic flowering	Lin et al (2007) Lin et al (2007)
SS_chr2_9370000	Mu	2	9370000	9380000								
SS_chr2_9390000	Mu	2	9390000	9400000			jg38883.t1.p1	chr2:9381422..9383156	Rust resistance kinase Lr10-like [<i>P. persica</i>]	0E-0 (89.06)	Resistance to rust pathogen in wheat	Feuillet et al. (1997)
SS_chr2_9400000	Mu	2	9400000	9410000			jg38884.t1.p1	chr2:9384405..9385313	Rust resistance kinase Lr10-like [<i>P. persica</i>]	0E-0 (93.89)	Resistance to rust pathogen in wheat	Feuillet et al. (1997)
SS_chr2_9410000	Mu	2	9410000	9420000			jg38887.t1.p1 jg38889.t1.p1 jg38890.t1.p1	chr2:9393728..9394426 chr2:9397100..9404899 chr2:9409975..9411390	Rust resistance kinase Lr10-like [<i>Pyrus x bretschneideri</i>] Uncharacterized protein LOC109950324 [<i>P. persica</i>] Rust resistance kinase Lr10-like [<i>P. persica</i>]	5E-84 (61.97) 0E-0 (91.28) 0E-0 (82.98)	Resistance to rust pathogen in wheat Unknown Resistance to rust pathogen in wheat	Feuillet et al. (1997) Feuillet et al. (1997)

Footnote: flowering pathway (lightyellow), fruit and seed properties (orange), developmental pathways and senescence (pink), response to pathogens (lightgreen), response to abiotic stress (lightblue), grey indicates that no match was found or only transposons.

Table III-5. Convergent candidate, selective sweep regions and genes shared between European cultivated and Chinese cultivated apricots. (Continue)

Selective sweep (SS) name	detected with	Chr ¹	Start position	End position	EUR	CHN	Most interesting candidate gene(s) within the 10Kb sliding window ²	Position in Marouch v2.0	Annotated gene	NBI BlastX ³	Function/Association	Reference
SS_chr2_12220000	Mu	2	12220000	12230000			ig39606.t1.p1	chr2:12216799..12219925	Putative disease resistance protein At1g50180 [<i>P. mume</i>]	0E-0 (84.53)	The LRR repeats probably act as specificity determinant of pathogen recognition	
SS_chr2_12227607	Omega	2	12227607	12237607			ig39608.t1.p1	chr2:12236907..12244361	Putative RNA-directed DNA polymerase [<i>Rosa chinensis</i>]	0E-0 (52.86)	Transposon like	
SS_chr2_16139953	Omega	2	16139953	16149953			ig40493.t1.p1	chr2:16140134..16142282	Sulfite exporter TauE/SafE family protein 3 [<i>M. domestica</i>]	2E-24 (94.44)	Induced by drought and Aluminium stresses	
							ig40495.t1.p1	chr2:16145444..16146369	Probable LRR receptor-like serine/threonine-protein kinase At3g47570 [<i>P. mume</i>]	1E-116 (93.92)	The LRR repeats probably act as specificity determinant of pathogen recognition	
							ig40496.t1.p1	chr2:16147589..16150605	WD repeat-containing protein WRAP73-like [<i>P. mume</i>]	1E-58 (72.79)	Involved in plant innate immune signaling	Miller et al (2016)
SS_chr2_16630001	Pi	2	16630001	16640000			ig40617.t1.p1	chr2:16633471..16636200	PRAF1 PH, RCC1 and FYVE domains-containing protein 1 [<i>Vitis vinifera</i>]	4E-57 (65.43)	Involved in membrane trafficking and signal transduction	Heras and Dröbak (2002)
SS_chr2_16630247	Omega	2	16630247	16640247								
SS_chr2_16700001	Pi	2	16700001	16710000			ig40647.t1.p1	chr2:16705782..16707036	Sugar transporter ERD6-like 7 [<i>P. avium</i>]	8E-19 (52.63)	Protein involved in the transfer of sugars and accumulation of fruit sugars	Wei et al (2014)
SS_chr2_16700000	Mu	2	16700000	16710000			ig40648.t1.p1	chr2:16707102..16708454	Sugar transporter ERD6-like 7 [<i>P. mume</i>]	5E-30 (41.01)	Protein involved in the transfer of sugars and accumulation of fruit sugars	Wei et al (2014)
SS_chr3_1110001	Pi	3	1110001	1120000			ig43267.t1.p1	chr3:1110507..1112526	RPSTA 40S ribosomal protein S7-1 [<i>P. avium</i>]	5E-18 (97.87)	Structural constituent of ribosome	
SS_chr3_1130000	Mu	3	1130000	1140000			ig43270.t1.p1	chr3:1132992..1136849	Hypothetical protein Prudu_016432 [<i>P. dulcis</i>]	4E-148 (87)	Potential transposon	
SS_chr3_1130001	Pi	3	1130001	1140000			ig43271.t1.p1	chr3:1140360..1141726	RPSTA 40S ribosomal protein S7-1 [<i>P. yedoensis</i> var. <i>nudiflora</i>]	4E-31 (97)	Structural constituent of ribosome	
SS_chr3_15560001	Pi	3	15560001	15570000			ig46777.t1.p1	chr3:15562209..15563503	Ferric reduction oxidase 8, mitochondrial [<i>P. yedoensis</i> var. <i>nudiflora</i>]	2E-102 (87.76)	Probably involved in iron reduction in leaf veins for transport	Wu et al (2005)
							ig46778.t1.p1	chr3:15565772..15568875	Ferric reduction oxidase 8, mitochondrial isoform X1 [<i>P. avium</i>]	0E-0 (70.43)	Probably involved in iron reduction in leaf veins for transport	Wu et al (2005)
SS_chr4_30000	Mu	4	30000	40000			ig4610.t1.p1	chr4:30701..31036	Receptor-like serine/threonine-protein kinase S-locus-specific glycoprotein S6-like [<i>P. mume</i>]	4E-69 (91.96)	Involved in the regulation of cellular expansion and differentiation	Pastuglia et al (2002)
SS_chr4_30218	Omega	4	30218	40218			ig6.t1.p1	chr4:31452..32249	Hypothetical protein PRUPE_4G006100, partial [<i>P. persica</i>]	3E-54 (81.40)		
							ig7.t1.p1	chr4:35185..37651	Topless-related protein 1-like isoform X1 [<i>P. yedoensis</i> var. <i>nudiflora</i>]	1E-34 (94.44)	Activation of TIR-NB-LRR-mediated immune responses and control of photoperiodic flowering	Goraloppia et al (2017)
SS_chr4_5080000	Mu	4	5080000	5090000			ig1275.t1.p1	chr4:5077037..5080083	Cyclic nucleotide-gated ion channel 1 [<i>P. avium</i>]	6E-88 (70)	Involved in calcium signal transduction and tolerance to heavy metal	Sunkar et al (2000)
SS_chr4_5080000	TajimaD	4	5080000	5090000			ig1276.t1.p1	chr4:5094307..5095440	WAK2 wall-associated receptor kinase 2 [<i>P. avium</i>]	0E-0 (98.41)	Binding to pectin that controls cell expansion and response to pathogens	Wagner and Kohorn (2001)
SS_chr4_5210000	Mu	4	5210000	5220000			ig1296.t1.p1	chr4:5204317..5205609	WAK2 wall-associated receptor kinase 2 [<i>P. avium</i>]	2E-93 (56.27)	Binding to pectin that controls cell expansion and response to pathogens	Wagner and Kohorn (2001)
SS_chr4_5210000	TajimaD	4	5210000	5220000			ig1299.t1.p1	chr4:5219167..5219484	Cyclic nucleotide-gated ion channel 1 [<i>P. avium</i>]	1E-22 (59.26)	Involved in calcium signal transduction and tolerance to heavy metal	Sunkar et al (2000)
							ig1298.t1.p1	chr4:5220094..5222343	Cyclic nucleotide-gated ion channel 1 [<i>P. avium</i>]	3e-85 (64.94)	Involved in calcium signal transduction and tolerance to heavy metal	Sunkar et al (2000)
SS_chr4_5330000	TajimaD	4	5330000	5340000			ig1324.t1.p1	chr4:5337570..5340960	WAK2 wall-associated receptor kinase 2 [<i>P. avium</i>]	0E-0 (97.64)	Binding to pectin that controls cell expansion and response to pathogens	Wagner and Kohorn (2001)
SS_chr4_5330000	Mu	4	5330000	5340000								
SS_chr4_5340000	Mu	4	5340000	5350000			ig1329.t1.p1	chr4:5350696..5353488	WAK1 wall-associated receptor kinase 1 [<i>P. mume</i>]	0E-0 (99.05)	Binding to pectin that controls cell expansion and response to pathogens	Wagner and Kohorn (2001)
SS_chr4_10010000	Mu	4	10010000	10020000			ig2359.t1.p1	chr4:10007981..10010399	LRR receptor-like serine/threonine-protein kinase FLS2 [<i>P. mume</i>]	0E-0 (84.43)	The LRR repeats probably act as specificity determinant of pathogen recognition	
SS_chr4_10010309	CLR	4	10010309	10020309			ig2361.t1.p1	chr4:10013362..10015014	LRR receptor-like serine/threonine-protein kinase EFR [<i>P. persica</i>]	9E-92 (84.88)	The LRR repeats probably act as specificity determinant of pathogen recognition	
							ig2363.t1.p1	chr4:10019163..10021020	DELLA protein DWARF8 [<i>P. persica</i>]	4E-08 (75.86)	key component of major hormone signaling pathways	Davière et Achard (2016)
SS_chr4_10900000	Mu	4	10900000	10910000			ig2555.t1.p1	chr4:10911154..10913130	Receptor-like protein 3 [<i>P. mume</i>]	0E-0 (98.94)	Involved in the perception of CLV3 and regulate meristems maintenance. Contributes, with WAKL22/RFO1 to resistance to <i>Fusarium oxysporum</i>	Shen et al (2013)
SS_chr4_10901210	CLR	4	10901210	10911210								
SS_chr4_13000000	TajimaD	4	13000000	13010000			ig3029.t1.p1	chr4:13001599..13002988		4E-91 (71.35)		
SS_chr4_13000000	Mu	4	13000000	13010000			ig3030.t1.p1	chr4:13008189..13008877	(-)-alpha-pinene synthase-like [<i>P. persica</i>]	2E-74 (95.60)	Fruit flavor	Aharoni et al (2004)
SS_chr4_13003335	CLR	4	13003335	13013335			ig3031.t1.p1	chr4:13010876..13013875		4E-107 (83.33)		
SS_chr4_13410000	Mu	4	13410000	13420000			ig3114.t1.p1	chr4:13405586..13408047		0E-0 (96.43)		
SS_chr4_13420000	Mu	4	13420000	13430000			ig3124.t1.p1	chr4:13424437..13427316	(-)-alpha-pinene synthase-like [<i>P. persica</i>]	2E-163 (90.58)	Fruit flavor	Aharoni et al (2004)
							ig3125.t1.p1	chr4:13433562..13436071		3E-96 (69.64)		
SS_chr4_16570000	Mu	4	16570000	16580000			ig3902.t1.p1	chr4:16573009..16573520	Probable pectate lyase 12 [<i>P. yedoensis</i> var. <i>nudiflora</i>]	2E-12 (85.37)	Cell wall degradation and fruit softening	Marín-Rodríguez et al (2002)
SS_chr4_16576949	CLR	4	16576949	16586949			ig3902.t1.p1	chr4:16573499..16573737	Probable pectate lyase 12 [<i>P. yedoensis</i> var. <i>nudiflora</i>]	6E-32 (95)	Cell wall degradation and fruit softening	Marín-Rodríguez et al (2002)
							ig3903.t2.p1	chr4:16576885..16581371	Putative ent-kaurene synthase [<i>Rosa chinensis</i>]	1E-81 (45.29)	Associated with the gibberellin phytohormone biosynthesis	Fu et al (2016)

CHAPTER III

Table III-5. Convergent candidate, selective sweep regions and genes shared between European cultivated and Chinese cultivated apricots. (Continue)

Selective sweep (SS) name	detected with	C hr	Start position	End position	EU R	CH N	Most interesting candidate gene(s) within the 10Kb sliding window ²	Position in Marouch v2.0	Annotated gene	NBI BlastX ³	Function/Association	Reference
SS_chr4_17320000	Mu	4	17320000	17330000			jg4139.t1.p1	chr4:17321707..17324662	Uncharacterized protein LOC109950029 [<i>P. persica</i>]	0E-0 (68.77)		
SS_chr4_17330000	Mu	4	17330000	17340000			jg4141.t1.p1	chr4:17330604..17332983	Putative zinc ion binding / nucleic acid binding protein [<i>Rosa chinensis</i>]	2E-52 (34.76)	Unknown	
SS_chr4_17337718	CLR	4	17337718	17347718			jg4142.t1.p1	chr4:17333676..17336169	Putative ribonuclease H protein At1g65750 [<i>P. mume</i>]	0E-0 (93.38)		
SS_chr4_17340000	Mu	4	17340000	17350000			jg4143.t1.p1	chr4:17343390..17345449	LRR receptor-like serine/threonine-protein kinase FLS2 [<i>P. avium</i>]	1E-34 (69.62)	FLS2 is involved in innate immunity.	Gomez-Gomez and Bolter (2000)
							jg4144.t1.p1	chr4:17345477..17346025	LRR receptor-like serine/threonine-protein kinase FLS2 [<i>P. avium</i>]	2E-42 (59.54)	FLS2 is involved in innate immunity.	Gomez-Gomez and Bolter (2000)
							jg4145.t1.p1	chr4:17346213..17348185	Probable E3 ubiquitin-protein ligase HERC2 [<i>P. mume</i>]	2E-14 (51.03)	Involved in abiotic stress response in plants	Cho et al (2017)
SS_chr4_17360000+A87:N89AA87:N89	Mu	4	17360000	17370000			jg4151.t1.p1	chr4:17362456..17365932	Callose synthase 10 [<i>P. mume</i>]	1E-10 (51.76)	Involved in plant development and/or in response to multiple biotic and abiotic stresses	Chen and Kim (2009)
SS_chr4_17367564	Omega	4	17367564	17377564			jg4154.t3.p1	chr4:17369794..17372255	Uncharacterized protein LOC110768889 [<i>P. avium</i>]	0E-0 (95)	Mutator-like transposase	
SS_chr4_17370000	Mu	4	17370000	17380000			jg4157.t1.p1	chr4:17378468..17381497	LRR receptor-like serine/threonine-protein kinase At3g47570 [<i>P. persica</i>]	0E-0 (83.58)	The LRR repeats probably act as specificity determinant of pathogen recognition	
SS_chr4_17380000	Mu	4	17380000	17390000								
SS_chr4_17400000	Mu	4	17400000	17410000			jg4167.t1.p1	chr4:17399901..17401733	Formin-like protein 14 [<i>P. yedoensis</i> var. <i>nudiflora</i>]	1E-71 (58.33)	Mediators of coordination of the cortical actin and microtubule cytoskeletons	
SS_chr4_17407788	CLR	4	17407788	17417788			jg4168.t1.p1	chr4:17402352..17405032	Putative serine/threonine phosphatase [<i>Rosa chinensis</i>]	2E-37 (31.23)		
SS_chr4_17410000	Mu	4	17410000	17420000			jg4170.t1.p1	chr4:17415466..17419289	Hypothetical protein PRUPE_4G254700 [<i>P. persica</i>]	2E-117 (86)		
SS_chr5_390000	Mu	5	390000	400000			jg10731.t1	chr5:391349..391888	ALP1 ANTAGONIST OF LIKE HETEROCHROMATIN PROTEIN 1 [<i>P. persica</i>]	2E-68 (85)	Needed for full reactivation of several floral homeotic genes that are repressed by PcG	Liang et al (2015)
SS_chr5_390000	TajimaD	5	390000	400000			jg10733.t1.p1	chr5:394685..401219	Sulfate transporter 3.1 [<i>P. persica</i>]	3E-38 (83.58)	Play a role in the regulation of sulfate assimilation	Gigolashvili and Kopriva (2014)
SS_chr5_400000	Mu	5	400000	410000			jg10734.t1.p1	chr5:401277..402064	NETWORKED 1D like protein [<i>P. persica</i>]	7E-78 (90)	Plant-specific actin binding protein	Deeks et al (2012)
							jg10736.t1.p1	chr5:402278..402723	ALP1 ANTAGONIST OF LIKE HETEROCHROMATIN PROTEIN 1 [<i>P. avium</i>]	3E-40 (83.16)	Needed for full reactivation of several floral homeotic genes that are repressed by PcG	Liang et al (2015)
							jg10737.t1.p1	chr5:408797..412239	ENHANCED DISEASE RESISTANCE 2 protein [<i>Fragaria vesca</i> subsp. <i>vesca</i>]	2E-67 (47.90)	Negative regulator of the salicylic acid-(SA-) mediated resistance to pathogens	Vorwerk et al (2007)
SS_chr5_9050000	Mu	5	9050000	9060000			jg12873.t1.p1	chr5:9055227..9058945	Ankyrin repeat-containing protein NPR4 [<i>P. persica</i>]	1E-82 (87.36)	Key role in plant defense	Vo et al (2015)
SS_chr5_9050001	Pi	5	9050001	9060000								
SS_chr5_11360000	TajimaD	5	11360000	11370000			jg13470.t1.p1	chr5:11364405..11365860	Putative zinc ion binding / nucleic acid binding protein [<i>Rosa chinensis</i>]	1E-15 (34)	Unknown	
SS_chr5_11361163	CLR	5	11361163	11371163			jg13471.t1.p1	chr5:11373486..11380646	SEC16B transport protein [<i>P. mume</i>]	0E-0 (98.76)	Required for protein transport from the endoplasmic reticulum to the Golgi apparatus	Takagi et al (2013)
SS_chr5_11370000	TajimaD	5	11370000	11380000								
SS_chr5_11371172	CLR	5	11371172	11381172								
SS_chr5_12170000	Mu	5	12170000	12180000			jg13657.t1.p1	chr5:12173442..12174760	LRR receptor-like serine/threonine-protein kinase EFR [<i>P. avium</i>]	9E-173 (93.06)	Involved in the plant innate immune response	Li et al (2009)
SS_chr5_12170001	Pi	5	12170001	12180000								
SS_chr5_12180000	Mu	5	12180000	12190000			jg13659.t1.p1	chr5:12183838..12197196	Zinc finger BED domain-containing protein RICESLEEPER 2 [<i>P. persica</i>]	0E-0 (88.35)	Transposase-like protein that is essential for plant growth and development	Knip et al (2012)
							jg13659.t1.p1	chr5:12185007..12191263	VRN2 Polycomb group protein VERNALIZATION 2 [<i>P. mume</i>]	7E-41 (91.11)	Plays a central role in vernalization by maintaining repressed the homeotic gene FLC, a floral repressor, after a cold treatment	De Lucia et al (2008)
SS_chr5_13560001	Pi	5	13560001	13570000			jg13958.t1.p1	chr5:13558234..13562892	Sister chromatid cohesion protein SCC2 [<i>Morus notabilis</i>]	9E-167 (65.46)	Essential protein required for cell fate determination during embryogenesis	Sebastian et al (2009)
							jg13959.t1.p1	chr5:13567896..13570772	Sister chromatid cohesion protein SCC2 [<i>P. avium</i>]	1E-159 (87.17)	Essential protein required for cell fate determination during embryogenesis	Sebastian et al (2009)

Table III-5. Convergent candidate, selective sweep regions and genes shared between European cultivated and Chinese cultivated apricots. (Continue)

Selective sweep (SS) name	detected with	Chr ¹	Start position	End position	EU R	CH N	Most interesting candidate gene(s) within the 10Kb sliding window ²	Position in Marouch v2.0	Annotated gene	NBI BlastX ³	Function/Association	Reference
SS_chr5_14660000	Mu	5	14660000	14670000			ig14193.t1.p1	chr5:14658554..14659126	Topless-related protein 1 [<i>A. thaliana</i>]	1E-28 (52.83)	Activation of TIR-NB-LRR-mediated immune responses and control of photoperiodic flowering	Zhu et al (2010)
SS_chr5_14664236	CLR	5	14664236	14674236			ig14194.t1.p1	chr5:14668881..14670976	FHY3 FAR-RED ELONGATED HYPOCOTYL 3 protein		Involved in light responses and plant senescence	Lin et al (2007)
SS_chr5_14670000	Mu	5	14670000	14680000								
SS_chr5_16370000	Mu	5	16370000	16380000			ig14580.t1.p1	chr5:16370478..16373515	Arabinosyltransferase XEG113 [<i>P. yedoensis</i> var. nudiflora]	4E-166 (93.50)	Important for cell elongation, root hair cell development, lateral root development and root hair tip growth	Roycewicz et al (2014)
SS_chr5_16380000	Mu	5	16380000	16390000			ig14582.t1.p1	chr5:16378271..16378747	ras guanine nucleotide exchange factor R isoform X1 [<i>P. persica</i>]	1E-95 (89)		
							ig14585.t1.p1	chr5:16386156..16392356	Arabinosyltransferase XEG113 [<i>P. yedoensis</i> var. nudiflora]	5E-165 (92.68)	Important for cell elongation, root hair cell development, lateral root development and root hair tip growth	Roycewicz et al (2014)
SS_chr6_2230001	Pi	6	2230001	2240000			ig20005.t1.p1	chr6:2231082..2237476	THO complex subunit 3, PPR-containing protein [<i>P. yedoensis</i> var. nudiflora]	3E-77 (59.75)	Acts as component of the THO subcomplex of the TREX complex which is thought to couple mRNA transcription, processing and nuclear export	Jauvin et al (2010)
							ig20006.t1.p1	chr6:2237599..2243581	DEXH-box ATP-dependent RNA helicase isoform X3 [<i>P. avium</i>]	1E-74 (89.44)	Involved in exosome-mediated RNA decay, the regulation of potassium deprivation and cold stress response	Xu et al (2011)
SS_chr6_7700000	TajimaD	6	7700000	7710000			ig21215.t1.p1	chr6:7703239..7709824	Protein argonaute 5 [<i>P. persica</i>]	4E-96 (64.74)	Main component of the RNA-induced silencing complex (RISC) that binds to a short guide RNA (miRNA or siRNA).	Takeda et al (2008)
SS_chr6_7703501	CLR	6	7703501	7713501			ig21216.t1.p1	chr6:7711692..7715340	Aspartic proteinase-like protein 2 [<i>P. mume</i>]	2E-87 (83.08)		
SS_chr6_9460000	TajimaD	6	9460000	9470000			ig21685.t1.p1	chr6:9473798..9474556	Putative Myb/SANT-like domain-containing protein [<i>Medicago truncatula</i>]	1E-63 (47)	Unknown	
SS_chr6_9464255	CLR	6	9464255	9474255								
SS_chr6_10310000	TajimaD	6	10310000	10320000			ig21957.t1.p1	chr6:10317228..10319421	F-box/LRR-repeat protein At3g03360-like [<i>P. avium</i>]	3E-63 (65.16)		
SS_chr6_10310000	Mu	6	10310000	10320000			ig21958.t1.p1	chr6:10319506..10321525	Uncharacterized protein LOC109948957 [<i>P. persica</i>]	0E-0 (65.07)		
SS_chr6_10320001	Pi	6	10320001	10330000								
SS_chr6_10324623	CLR	6	10324623	10334623								
SS_chr6_12130001	Pi	6	12130001	12140000								
SS_chr6_19860001	Pi	6	19860001	19870000			ig24264.t1.p1	chr6:19865829..19867108	SRG1 Senescence related gene [<i>P. avium</i>]	2E-62 (93.64)	Involved in leaf senescence	Kim et al (2018)
							ig24267.t1.p1	chr6:19867851..19872301	SRG1 Senescence related gene [<i>P. mume</i>]	2E-110 (81.78)	Involved in leaf senescence	Kim et al (2018)
SS_chr7_4140000	TajimaD	7	4140000	4150000			ig5621.t1.p1	chr7:4136759..4141737	3-oxoacyl-[acyl-carrier-protein] synthase II, chloroplastical [<i>Rosa chinensis</i>]	2E-60 (91.30)	Catalyzes the condensation reaction of fatty acid synthesis during seed metabolism and confers resistance to low temperatures by maintaining chloroplast membranes integrity	Carlsson et al (2002)
							ig5622.t1.p1	chr7:4142732..4150148	RRP12-like, ARM repeat protein [<i>P. avium</i>]	4E-133 (80.70)	Ribosomal RNA Processing	
SS_chr7_17490000	TajimaD	7	17490000	17500000			ig8859.t1.p1	chr7:17492370..17494476	No similarity found			
SS_chr7_17490000	Mu	7	17490000	17500000			ig8861.t1.p1	chr7:17495519..17498195	No similarity found			
SS_chr7_17490001	Pi	7	17490001	17500000			ig8862.t1.p1	chr7:17499703..17500332	TMV resistance protein N-like [<i>P. avium</i>]	5E-30 (64.06)	Disease resistance protein	Dinesh-Kumar et al (2000)
SS_chr7_17492278	Omega	7	17492278	17502278								
SS_chr8_4233845	CLR	8	4233845	4243845			ig15850.t1.p1	chr8:4241641..4242175	Tubby-like F-box protein 8 [<i>P. mume</i>]	1E-73 (97.20)	Involved in ABA signaling pathway and osmotic stress response	Bao et al (2014)
SS_chr8_4240000	TajimaD	8	4240000	4250000			ig15851.t1.p1	chr8:4245450..4248562	ACCELERATED CELL DEATH 6-like protein [<i>P. mume</i>]	1E-126 (99.09)	Activator of the defense response, Regulates the salicylic acid (SA) signaling pathway leading to cell death and modulating cell fate	Rate et al (1999)
							ig15852.t1.p1	chr8:4250284..4256917	Tubby-like F-box protein 8 [<i>P. mume</i>]	5E-155 (99.60)	Involved in ABA signaling pathway and osmotic stress response	Bao et al (2014)
SS_chr8_5780000	Mu	8	5780000	5790000								
SS_chr8_5780001	Pi	8	5780001	5790000								
SS_chr8_6680000	Mu	8	6680000	6690000			ig16530.t1.p1	chr8:6690034..6690654	Uncharacterized protein LOC107880980 [<i>P. mume</i>]	2e-65 (72.97)	Unknown	
SS_chr8_6680001	Pi	8	6680001	6690000								

GENERAL DISCUSSION and CONCLUSION

One primary aim of my PhD work was to identify the processes and the causal factors that have shaped the composition and distribution of apricot (*Prunus armeniaca* L.) over time. Another was to infer the history of wild apricot species divergence and apricot domestication processes, in a phylogenetic context. To this end, we addressed important questions related to the origin of apricot, its hypothetical ancestor, its relationship with other wild species endemic in Central Asia and China and finally the pedo-climatic and geologic events that caused *Armeniaca* species to evolve through time. From then onward, we were ready to investigate how selection has influenced the genetic diversity and the genomic architecture of wild and cultivated apricots.

I- Evolutionary processes with a role in *P. armeniaca* history

A common origin for the Armeniaca species?

Based on microsatellite markers and approximate Bayesian computation, we uncovered the divergence history of wild *Armeniaca* lineages across Central and East Asia. We found genetically differentiated clusters of wild *P. armeniaca* and *P. sibirica* species that may have been formed, from an initial common ancestor, during the Himalayan orogeny, in the Neogene period. Indeed, our ABC nested framework has shown with high confidence that *P. armeniaca* and *P. sibirica* ancestral populations diverged 8 to 16 million years ago (Mya) (Figure I-3A, Chapter I).

This speciation event between *P. armeniaca* and *P. sibirica* coincides with the divergence time between peach and almond (Delplancke et al., 2016; Velasco et al. 2016). Similar to peach and almond, the causal factors for divergence would be the climatic changes consecutive to the Tibetan Plateau uplift and Himalayan orogeny that led to isolation and subsequent divergence of *P. sibirica* (on the Eastern part of Himalaya) and *P. armeniaca* (on the Western part) lineages from a common ancestral species. Nevertheless, at this point, answering more precisely to the question of the common ancestor for those *Armeniaca* species would require a sampling of more ancestral apricot species that we lacked in our ABC analyses, such as *P. mume* and *P. mandshurica*, to avoid simulating an unknown ancestral population. Interestingly, molecular phylogenetic analyses performed with four plastid markers and one nuclear ribosomal ITS region indicated a divergence time for the *Armeniaca* species around 34 Mya and for the *Amygdalus* species, ~47 Mya (Chin et al., 2014). This incongruence is likely due to the combined effects of the type of markers used, the limited number of individuals tested, incomplete lineage sorting and hybridization/introgression. Effectively, we demonstrated that after *P. armeniaca* and *P. sibirica* diverged, ancient hybridization still occurred between them having generated a new

species relatively reproductively isolated from both parent species, *i.e.* the W3 green *P. sibirica* clade which is endemic in North-Western China. This illustrates the first case of hybridization in the *Armeniaca* evolutionary history that seemingly had an important role in shaping apricot diversity, as we will discuss below.

Evolutionary processes towards diversification in wild apricot species

Long after *P. armeniaca* and *P. sibirica* split and concomitant with the *P. sibirica* W3 hybrid speciation, both Central Asia and China acted as diversification centers for wild *Armeniaca* species. We observed in both regions genetically differentiated clusters resulting from the population contraction into refugia during the quaternary glaciations. This last result was confirmed when we inferred the demographic history of wild and cultivated apricots by using whole-genome sequencing data (Figure III-7, Chapter III).

In Central Asia, the phylogeny and structure analysis based on SSR markers revealed that the wild Central Asian apricots were clustered into two groups, W1 (red) and W2 (yellow) respectively. The same clustering was retrieved with the WG (whole-genome) polymorphism dataset (Figure III-6B, Chapter III). The W2 yellow cluster consisted of the largest number of accessions all from the Tian-shan Mountains ranges while the W1 red cluster included wild accessions from Sary Chelek (Kyrgyzstan). We estimated that those two Central Asian clusters diverged about 20 Kya which corresponds to the Last Glaciation Maximum (LGM) (Decroocq et al, 2016; Liu et al, 2019, Chapter I). This would indicate that Sary Chelek (and more generally the Fergana valley) acted as refugia during LGM, from which Central Asian apricot lineage recolonized the Northern regions of Tian-shan Mountains ranges after increase in temperature. A similar scenario of contraction during LGM was proposed as a driver of diversification in other perennial plant species, such as *Amygdalus* (almond; Zeinalabedini et al. (2010) and *Juglans* (walnuts; Aradhya et al 2017).

Meanwhile, we also showed that, on the east side of the Pamir-Himalaya Mountains, the *P. sibirica* W3 lineage resulted from an interspecific hybridization between *P. armeniaca* and *P. sibirica* about 330 Kya, albeit estimate of the hybridization time had a large confidence interval [12~474 Kya]. Those estimates were obtained from genotyping 24 'perfect' microsatellite loci only. Therefore, it would be interesting in the future to infer this hybridization event from the WG polymorphism data. *P. sibirica* W3 and W4 populations are nowadays highly differentiated and display a prominent phylogeographic distribution, North-West to North-East respectively. This probably reflects allopatric division during the

Quaternary climate² upheavals towards distant refugia, followed by subsequent admixture when two populations rejoined after the ice-sheet retreat. While we detected no clear hybridization zone from the two *P. sibirica* genetic clusters, probably due to the lack of sampling the right connecting zone, Wang et al (2017) detected a distinct boundary located over the Yanshan Mountains (West of the Beijing province in North China). In this study that was based on both plastid and nuclear markers, the split time of the two Western and Eastern *P. sibirica* groups ranged from ~179.5 Kya, based on the cpDNA, and 13 Kya, from the nuclear DNA, which correspond to our confidence interval for the hybridization and later speciation of the W3 (Western) *P. sibirica* lineage. It thus appears that diversification in the wild Chinese apricots (mostly *P. sibirica* here) resulted from (i) interspecific hybridization between *P. armeniaca* and *P. sibirica*, followed by (ii) fragmentation into independent refugia during the successive cold periods of the Pleistocene epoch (2.58 Mya to 11.7 Kya), (iii) recolonization of the surrounding areas once climatic conditions improved, leading to (iv) subsequent admixture(s) when the two distinct *P. sibirica* lineages rejoined. This would explain part of the inconsistencies observed when estimating the divergence, hybridization and speciation time all along the *P. sibirica* evolutionary history. Indeed, incongruence between plastid and nuclear phylogenies suggests that the W3 *P. sibirica* lineage originated via an ancient hybridization event (indicated by the cpDNA gene flow) but also, that recurrent admixture between the two divergent lineages happened by population contraction and expansion from separate refugia, following the Pleistocene environmental upheavals. Similar evolutionary processes were observed for other temperate-deciduous tree species native from North China, among which *Juglans manshurica*, the manchurian walnut species. Chloroplastic and nuclear polymorphism analyses showed two different lineages, suggesting that the species distribution was fragmented into two independent refugia, in the past, before experiencing a more recent admixture event (Bai et al., 2010).

In summary, our data show that wild apricot species encountered a complex evolutionary history made of various mechanisms, including divergence, speciation, isolation, hybridization, migration (in the meaning of both gene flow and recolonization) and which all impacted the genetic variation of the current species endemic in Central Asia and in China.

² The Quaternary period is the most recent geological time period, spanning from 2.58 million years ago to today. Global temperatures have shifted between cold glacial periods and warmer interglacial periods.

A proposed model for the evolutionary pathway of cultivated apricots

Domestication started with the conscious or/and unconscious selection, within the wild progenitors, of advantageous natural variation for human ends. It is the mechanism by which most domestic species evolved, among which apricot. Domestication is expected to have happened long after the divergence between wild *P. armeniaca* and *P. sibirica* (several Mya), and after the interspecific hybridization that gave rise to the *P. sibirica* W3 genetic cluster (several hundred Kya). The first step of the domestication process, *i.e.* the sampling of food directly from the fruit forests, should have taken place when humans were still nomadic hunter-gatherers. It was no more than 12,000 years ago that humankind began to consciously harness and select among the genetic diversity of living plants for its own benefit (Diamond, 2002; Puruggannan and Fuller, 2009), thus leading to the second step of the domestication process: the cultivation, propagation and preservation of the most interesting individuals (Figure 7A, for long-lived perennial crop species such as apricot). At this time, the Central Asian and Chinese apricot forests were already differentiated and geographically apart (except over the Xinjiang region), the wild *P. armeniaca* forests on the Western side of the Pamir/Himalaya ranges and the two *P. sibirica* genetic clusters on the Eastern part, together with *P. mume* natural populations. We showed in Chapter I that two distinct domestication events gave rise to European and Chinese cultivated apricots, separately. While the European cultivated apricots most probably originate from the Central Asian wild forests (W2 yellow cluster from the Tian Shan ranges), the origin of the Chinese cultivars is more complicated, coming either from Central Asian *P. armeniaca* progenitors or from the W3 Western *P. sibirica* cluster. The fact that W3 is a hybrid species between *P. armeniaca* and *P. sibirica* could explain the difficulties in clearly assigning the origin of the cultivated apricots in China. Unfortunately, because of a limited number of flowering apricot (*P. mume*) individuals in our samples, we could not assess the contribution of this other *Armeniaca* species which is also native from China and partly shares the *P. sibirica* habitats. This is planned in a very close future, after the release of hundreds of *P. mume* genome sequences (Zhang et al., 2018).

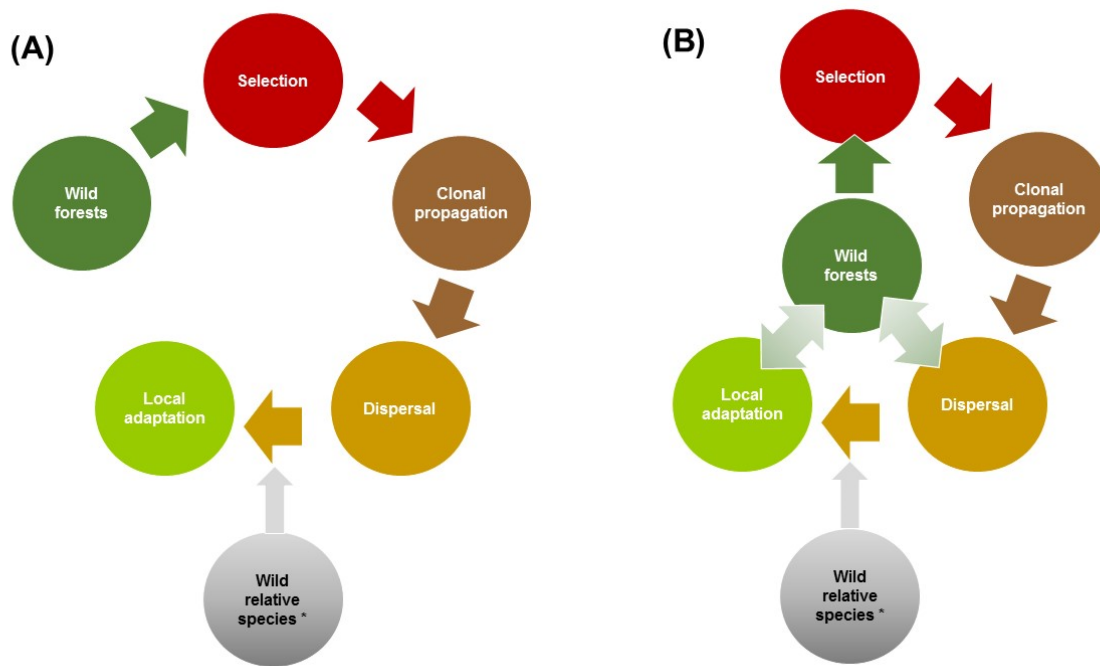


Figure 7. The domestication process in Eurasian cultivated *P. armeniaca*. (A) The process of domestication for European cultivated apricot trees. (B) The process of domestication for Chinese cultivated apricot trees. The figure is an adaptation of Figure 1 of Gaut et al (2015). * Wild related species which do not belong to the section *Armeniaca*. The light green arrows are depicting bi-directional gene flow that might have occurred after the initial steps of selection of domesticated/pre-domesticated forms.

Nevertheless, inference of demographic history for the Central Asian wild and the Chinese/European cultivated apricots, based on WG nucleotide polymorphism (Chapter III of this manuscript), allowed to confirm the previous scenarios, that is to say (i) European apricot cultivars originated from Central Asia and were domesticated a few thousand years after LGM from the Central Asian forests, (ii) Chinese apricots diverged from Central Asian wild apricots long before domestication. They were thus domesticated elsewhere, most probably in China, from the forests of *P. armeniaca* x *P. sibirica* hybrids.

Multiple dispersion routes for the domesticated/pre-domesticated apricots

The domestication process often involves, as a third step, dispersion across long distances and different environmental conditions, combined with intense selection (Figure 7A). Under such a process, multiple opportunities arise for gene flow with locally adapted cultivated or wild forms (Meyer et al., 2013). This was the case for apple (*Malus domestica*) which was domesticated 4,000-10,000 ya from *M. sieversii* in the Central Asian Tian Shan mountains and moved to Europe along the Royal and Silk Roads, hybridizing bidirectionally with wild relatives (*M. orientalis* in Caucasia and far Eastern Europe, *M. sylvestris* in Europe) at different time points on route (Cornille et al., 2014). From our data, we can say that European

cultivated apricots followed more or less the same pattern of dispersion, towards Europe (See Figure 6, General Introduction of this manuscript) but the bi-directional gene flow was more limited (Figure 7A). Wild relatives of *P. armeniaca* are native neither from the Near East, nor from Caucasia, Europe (see Chapter II concerning *P. brigantina*) or North Africa. The only documented hybridization is the one that occurred with species from the subgenus *Prunophora*, i.e. the diploid plum species *P. cerasifera*, giving rise to the *P. x dasycarpa* species (the so called purple or black apricot). In consequence, it is expected that diversification (before modern breeding) in European apricots towards adaptation to new environments and human practices/culture was achieved by selection on new or existing genetic diversity, rather than by crossing with other interfertile species.

The situation is obviously different for Chinese cultivated apricots. While it is not clear yet where the domestication of Chinese apricots happened, and from which species, domesticated apricots in China encountered recurrent hybridization with wild relatives since domestication. Of some interest is that the modern distributions of wild *P. mandshurica*, *P. sibirica* and cultivated *P. armeniaca* in Northern China overlap with each other, which, given that all can readily hybridize, suggests a recurrent gene flow between them.

From forest to orchard, the importance of local adaptation in the domestication process

When our ancestors began to shift from collecting wild plants to actively cultivating them, they imposed intense selective pressures for traits that facilitate human cultivation, crop harvesting and conservation. These intense selection pressures have generated remarkable transformations of plant phenotypes, generating a suite of features that are shared across many crop species, the so called 'domestication syndrome'. Domestication traits in the strict sense may be considered as those that distinguish a crop from its wild relatives. For long-lived perennial crops, the distinction is not always that easy, at least for apricot. The same variation of fruit size and color as well as tree shape and growth habit is observed both in cultivated and wild compartments, with slightly bigger fruits and leaves for the modern cultivars (Zaurov et al., 2013). However, traits favored during the initial stages of domestication are generally those that facilitate fruit harvesting, tree propagation (graft compatibility or multiplication through cuttings) and the ones that are associated to a better taste or fruit drying. These traits include not only those that are likely to have evolved through conscious selection (e.g., increased palatability and productivity), but also changes more likely to reflect unintentional selection (e.g., synchronized flowering and fruit maturity, erect growth to facilitate increased plant density in gardens). Like its natural counterpart,

unconscious selection leading to adaptation under domestication is not limited to visible phenotypes and may have involved substantial physiological or developmental changes in order to adapt to new environments and growth conditions associated with cultivation.

While the distinction between domestication traits and adaptive traits that characterize varietal differences (e.g., variation in fruit pigmentation, adaptation to different climates and latitudes) is not always clear, the latter traits often may be discerned because they remain variable among different varieties or landraces. Those traits are usually related to adaptation to local environmental conditions or cultivation practices. Intrinsically, they are expected to be different for Chinese and European cultivated apricots. However, despite the fundamental role of plant domestication in human history, we still know little about adaptation under domestication and the underlying genes. Opportunities exist provided we compare parallel (or convergent, we will discuss this term later on) adaptive evolution (Ross-Ibarra et al., 2007). The independently evolved domestication traits can be studied at various levels of phylogenetic divergence, including separate lineages within a single crop species (e.g., apricot adaptation to European and Chinese cultural practices or environmental conditions), different crop species within a single genus (e.g., fruit quality traits in *Prunus* crop species), and different genera at higher taxonomic levels (e.g., winter cold requirement in Rosaceae perennial species). Comparisons of independently domesticated crop lineages (e.g., European and Chinese apricots) can thus facilitate inferences into the molecular and developmental underpinnings of parallel adaptation during domestication. This has been tackled in Chapter III of this manuscript and will be largely discussed below (see 2-Comparative genomics of apricot domestication and adaptation).

Similarly to adaptation under domestication, local adaptation occurs in natural populations under the combined action of many selective factors, such as climate, edaphic factors and biotic stresses (Hedrick, 2006, Kawecki and Ebert 2004, Linhart and Grant, 1996). Naturally grown trees have experienced long-term and cyclic climate changes that often have promoted phenotypic variation and in consequence local adaptation (Frejaville et al., 2019). Population genomic analyses of wild Central Asian apricots indicate signatures of positive selection most probably by long-term adaptation (Chapter III). Indeed, footprints of adaptive selection are usually linked to the accumulation of beneficial mutations which will be maintained by balancing selection, as indicated by the significantly positive value of Tajima's *D*. Inference of the wild Central Asian apricot demographic history yielded evidence of a reduction of its population size, most probably due to the last glaciation period (~20 Kya), followed by its expansion (Figure III-7, Chapter III). This result is consistent with the

report on *Vitis vinifera* which positive values of Tajima's D were mostly observed in both wild (D : ~0.89) and cultivated (D : ~1.35) subgroups (Marrano et al., 2018).

Wild-to-crop and/or interspecific introgression as a source of local adaptation and diversification in Central Asian and Chinese cultivated apricots?

Natural hybridization has been defined as 'Successful matings in nature between individuals from two populations, or groups of populations that are distinguishable on the basis of one or more heritable characters' (Arnold et al., 2006). However, the definition of hybridization can be extended to between domesticated forms and their wild relatives that is not facilitated by humans and that led to adaptation during or after domestication (Janzen et al., 2019).

In the past two decades, numerous studies have documented the role of adaptive hybridization in the evolution of domesticated annual species (Matsuoka et al., 2002). In perennial species, examples are scarcer although wild-to-crop introgression may be even more important in perennials than annuals because of their outcrossing mating system, the spatial proximity between wild and cultivated forms and the lifespan of both of them. Hybridization and consecutive gene flow have been shown to have shaped European domesticated apples (Cornille et al., 2014). More recently, Duan et al (2017) demonstrated that not only hybridization occurred between the ancient domesticated apples from Central Asia (*M. sieversii*) and *M. orientalis* in Caucasus or *M. sylvestris* in Europe but also between *M. sieversii* and *M. baccata*, giving rise to orient hybrid species which are now cultivated along the Silk Road, eastward.

In the current manuscript, we showed, in Chapter I, footprints of gene flow between *P. sibirica* and the cultivated apricot landraces in China (Figure I-1B). A similar wild-to-crop hybridization process happens in Central Asia, between the wild *P. armeniaca* (W1 and W2) and the local cultivars (Figure I-3C, Chapter I). However, in both cases, China and Central Asia, we are lacking the information on the phenotypic traits that could have been impacted by such interspecific or wild-to-crop gene flow. Due to genetic similarity between wild populations and local cultivars, it is also sometimes very difficult to discriminate between local adaptation following wild-to-crop introgression and intraspecific crossing between domesticated cultivars and subsequent selection. Nevertheless, few examples of diversification in Chinese cultivated apricots that are linked to wild-to-crop gene flow exist, *i.e.* the kernel apricot used mainly for its seed properties (Fang et al., 2016) and the Xing mei ornamental trees (Zhang et al., 2018). The genetic basis of domestication in flowering apricot –also called *mei* in China- (which corresponds to a distinct *Armeniaca* species, *P. mume*) was recently published, together with details on introgression events from apricot (*P.*

armeniaca) and plum (*P. salicina*) that were essential in mei cultivation (Zhang et al., 2018). The same approach remains to be done for Xing (the Chinese name for *P. armeniaca* apricot). However, we can already speculate on the importance of interspecific hybridization on the genetic diversity, the diversification and local adaptation of Chinese apricots. Whereas in Europe, our data obtained on cultivated *P. armeniaca* and wild *P. brigantina* displayed no evidence for interspecific gene flow (Chapter II), more studies are requested to decipher the impact of adaptive wild introgression in Chinese apricots thus explaining the range of habitats currently occupied. Moreover, such adaptive introgression could have an important impact on the basic study of domestication, affecting estimates of the strength of the domestication bottleneck, the level of genetic diversity, the timing of domestication, the targets of selection during domestication etc... (Figure 7 B). Beyond confounding detection of the true progenitor of Chinese cultivated apricots, it also reminds us that we have to keep a more nuanced view of the following genomic results.

II- Comparative genomics of apricot domestication and adaptation

A large amount of genomic variation resources generated in Chapter III of this manuscript has not only allowed us to gain new insights into apricot evolution but also provided a mean in understanding how the apricot genome has been reshaped by human selections and numerous parallel processes of local adaptation, all along with apricot dispersion and cultivation.

Indeed, both cultivated apricot groups studied here, *i.e.* Chinese and European, showed similar global patterns of genomic diversity (positive Tajima's D value, significant genome-wide nucleotide diversity estimates π). Compared to other domesticated perennial crops, the nucleotide diversity in cultivated apricot is higher than that of peach (1.5×10^{-3}) (Cao et al., 2014), cassava (2.6×10^{-3}) (Kawuki et al., 2009) or apple (2.20×10^{-3}) (Duan et al., 2017), but lower than that of date palm (9.2×10^{-3}) (Hazzouri et al., 2015). It is rather similar to pear (5.5×10^{-3}) (Wu et al., 2018). Domestication in Chinese and European apricots was accompanied by a low reduction of diversity ($\pi_{\text{wild}}/\pi_{\text{CHN}}=1.02$ and $\pi_{\text{wild}}/\pi_{\text{EUR}}=1.21$) which indicated a very weak bottleneck during apricot domestication, consistent with findings in previous studies (Decroocq et al., 2016; Liu et al., 2019, Chapter I of this manuscript). While Chinese apricots retained up to 97% of genome-wide diversity from its Central Asian progenitor, European cultivars encountered a more significant bottleneck, retaining only 82%. Nonetheless, part of the Chinese nucleotide diversity could also result from recurrent hybridization with wild relatives (see discussion above). Those results are comparable with the domestication effect in apple (Duan et al., 2017) ($\pi_{\text{M. sieversii}}/\pi_{\text{M. domestica}}=1.07$, retained 93%

from *M. sieversii*; $\pi_{M. sylvestris}/\pi_{M. domestica}=1.16$ retained 86% from *M. sylvestris*, but contrasts with the significant reduction of nucleotide diversity that happened during peach domestication ($\pi_{wild}/\pi_{landrace}=2.92$) and improvement ($\pi_{wild}/\pi_{improved-cultivars}=3.5$) (Li et al., 2019). Linkage disequilibrium analyses for each apricot cultivated group further supported a very weak and nearly undetectable domestication bottleneck. One in the other, this suggests that, albeit Chinese and European domesticated apricots followed different demographic trajectories since their common Eurasian ancestor, we should be able to investigate and compare the genomic changes underlying domestication in different environments and cultivation systems.

A fine tuning of flowering time during apricot adaptation

Using the *P. armeniaca* ‘Marouch’ reference genome and resequenced individuals spanning the geographic range of wild and cultivated apricot trees, we were able to characterize genomic selections at several levels. First, we detected enrichment of resistance genes involved in environmental responses (biotic stresses, mostly). Second, we uncovered a key role for transcription factors controlling flowering time and dormancy release, especially in natural populations of Central Asia and in domesticated European apricots.

For European cultivars, it appears that selection has acted during apricot dispersion in new environments like in Western and Eastern Europe, by rewiring the timing and cross-talk between the vegetative and flowering processes. Such an example of selection that acted upon the shift of flowering time was demonstrated in *Arabidopsis thaliana*, more particularly on the vernalization requirement locus *FRIGIDA* (*FRI*), leading to an increase of early-flowering allele frequency (Toomajian et al., 2006). The apricot *FRIGIDA*-like candidate gene (SS_chr4_15990000) is also under positive selection in European cultivated apricots. This might reflect a change in winter chilling requirement in determining the flowering time, especially for cultivars adapted to Southern, Mediterranean latitudes (Shindo et al., 2005; Stinchcombe et al., 2004). Within the five flowering pathways that involve more than 300 different genes in *A. thaliana* (Bouché et al., 2015), two other TFs and two transcriptional co-repressors are also under positive selection, *i.e.* TFL1 (SS_chr7_8970001), APETALA-1 (SS_chr1_25060000), LEUNIG and DOF-1 (SS_chr4_2720000). Another genetic factor under positive selection (SS_chr5_12180000), this time in both cultivated Chinese and European apricots, is the vernalization response gene *VRN2*, a key factor which controls the requirement of a cold period to switch from the vegetative to reproductive phase (Gendall et al., 2001). All things considered, it thus appears that the dispersal of apricots westward and eastward is concomitant with a selection having likely re-shaped the function of some

specific flowering factors, readjusting it depending on the environment and the local conditions. This would correspond more to selection acting on an adaptive trait than on a domestication trait since it resulted in variants adapted to different agronomic conditions than the ones encountered usually in its area of origin. Indeed, apricot trees are now able to grow under a wide range of agro-climatic regions, from latitude 50° to 30° North and 50° to 30° South. Its adaptation to new environments, in which the length of growing season is influenced by factors such as drought, cold or heat, was achieved by selection for earlier or later flowering, with or without a cold requirement in winter. In the future, studying the allelic variation at the above loci in relation to plant phenology would help determine which allelic combinations are most beneficial in a particular growing region or in optimizing plant phenology for the changing climate.

In the case of the Central Asian wild apricots, the most important transcription factors under selection are related to the control of dormancy (*DAM -SVP-like-* genes) and of the photoperiodic flowering (*CONSTANS*, *CO gene*). As rightly stated by Brambilla et al (2017), this illustrates ‘The importance of Being on Time...’ for triggering reproductive growth. This is particularly true for perennial species growing naturally in regions characterized by only two seasons: Winter and Summer, as Central Asia. Species adapted to higher latitudes promote flowering during seasons characterized by LD (Long-Day), indicative of the warm days of spring and summer. When day length exceeds a specific critical threshold, flowering is promoted in those LD species. Moreover, plants that initiated flowering buds before winter (as is the case for perennial fruit trees), often also need to satisfy a ‘chilling’ requirement (exposure to low temperatures for several weeks, that coincides with the ‘dormancy period’ in trees) to become competent to respond to photoperiodic induction. In our apricot model, both the *CONSTANS* and *DAM* genes are under natural selection in the Central Asian forests. While *CONSTANS* is clearly a central regulator within the photoperiodic flowering pathway (Shim et al., 2017), the peach *DAM* gene products were shown to be involved in the integration of environmental cues that regulate the transition into and out of dormancy (Falavigna et al., 2019; Li et al., 2009). Since those two elements, Day Length and cold requirement, appear to be under selection in natural populations, it might indicate that they are key drivers into the cross-talk between photoperiodic response and dormancy in apricot forests where properly timing the floral transition is crucial for reproductive success. Ding and Nilsson (2016) have shown that the genetic pathways controlling growth and dormancy cycles in forest trees are highly conserved with the pathways regulating flowering time in annual plants. Here, we postulate that they also correspond to important adaptive traits in fruit trees.

The molecular and physiological control of flowering is complex, it brings into play up to 306 genes in the model plant *Arabidopsis thaliana*, in eight interconnected pathways, namely the photoperiodic, vernalization, circadian clock, ageing, hormones, ambient temperatures, sugars and autonomous pathways (Bouché et al., 2015). Noteworthy, it appears that the control of flowering time in apricots involves distinct pathways, depending on if it is cultivated or not. While in domesticated apricots, key factors of the FRI and PRC2 (through VRN2) complexes that both regulate *FLC* (*Flowering Locus C*) expression within the vernalization pathway are under strong positive selection, our data indicate the importance of the photoperiodic pathway (with the CO 'hub' factor) in fine-tuning the timing of the reproductive phase in Central Asian forests.

Convergent genomic signatures of domestication in Chinese and European apricots

It has been hypothesized that domestication is mediated by similar sets of genes in different taxa (Fuller et al., 2014; Lenser and TeiBen, 2014). Recent studies have revealed that convergent phenotypic evolution is often based on molecular changes in orthologous genes or pathways (Martin and Orgogozo, 2013). Of note, we will be talking here of 'convergence' instead of 'parallelism'. Convergence assumes that when a given phenotype evolves, the underlying genetic mechanisms are different in distantly related species. Though it will be called 'parallelism' when the underlying genetic mechanisms are similar in closely related species. However, several examples show that the same phenotype might evolve among closely related species by changes in different genes (e.g., Li and Gill, 2006). Conversely, similar phenotypes might evolve in distantly related species by changes in the same gene (e.g., *TFL1* for seasonal flowering in Soybean, Barley, roses, strawberry and more recently kiwifruit. Iwata et al, 2012; Voogd et al. 2017). For simplicity, we chose the term 'convergence' that also includes 'parallelism' for evolution at both phenotypic and genetic levels, as advocated by Arendt and Reznick (Arendt and Reznick, 2008), but we agree that this could be subjective.

Following the theory of convergence, domestication of two distinct taxa or genetic clusters may have largely recruited variation at the same genes. In consequence, if convergence proves to be the rule rather than the exception, it will facilitate the transfer of genetic information among crop species. From the literature, many of those genes linked to convergent traits are controlling flowering time (see *TFL1* example above), grain properties and more specifically, fruit dehiscence (*Sh1* gene in rice, maize and sorghum, Lin et al, 2012), color (e.g., orthologous *MYB R2R3* genes determining fruit skin and flesh color, Allan

and Espley. 2018; Fournier-Level et al, 2010), resistance to abiotic stress (*ALMT* for heavy metal tolerance in Rye and Wheat, Liu and Zhou, 2018).

In perennial fruit crops, we summarized in Table 1 (General introduction of this manuscript) the most recently identified 'domestication traits'. Basically, up to date, the most commonly identified traits are related to fruit properties (acidity, sugar, flavor, size, color, and secondary metabolism), plant growth and hormone synthesis, flowering time and chilling requirement. However, while, clearly, common targets of selection related to domestication and improvement exist, very little convergence was documented, yet, except for the case study of the *MYB R2R3* gene in Kiwifruit, blood oranges and grapes (Butelli et al., 2012; Fournier-Level et al., 2010; Li et al., 2017) and *TFL1* in roses and strawberries (Iwata et al., 2012).

In our study, three main loci under selection in both European and Chinese apricots have been identified through three different tests, *SS_chr2_6800000*, *SS_chr4_5300000* to *SS_chr4_5320000* and *SS_chr4_13000000*, namely. While the first one maps over two candidate genes involved in the photoperiodic control of flowering (*FAR1-like* and Ubiquitin-like-specific protease *ESD4* genes), the two loci on chromosome 4 appear to be related to fruit properties, *i.e.* fruit size and flavor. Indeed, *WAK2* proteins were shown to be implicated in cell expansion together with biotic stress response (Verica and He (2002), whereas *FaNES1*, the gene coding for (-)-alpha-pinene synthase is highly expressed and only present in the wild strawberries (Aharoni et al., 2004). Besides those three loci and when describing patterns of selection linked to apricot domestication, we found a significant number (7.75% of the total number of selective sweeps identified) of genomic signatures of selection shared between Chinese and European cultivars. Most of them relate to response to biotic or abiotic stresses, fruit properties and the control of flowering time, plant growth and leaf senescence (Table III-5).

At this stage, we have no indication if the selection in these genes occurred before the divergence of the Central Asian and Chinese groups or after, during the two independent domestication events that gave rise to European and Chinese apricots or at an early improvement step. Nevertheless, these insights into convergent domestication traits may still contribute to defining candidate genes for genetic improvement in apricot. Withal, we also identified distinct and unique signatures of selection either in Chinese or in European cultivars, therefore suggesting adaptive geographical or cultural divergence in this species. One prominent example might be the positive selection acting over the *S1* (self-incompatibility) locus but only in Chinese cultivated apricots. A similar contrasting pattern of selection was observed over the *S1* locus when comparing self-fertile (SC for self-compatible)

peach and self-incompatible (SI) almond accessions (Akagi et al., 2016; Velasco et al., 2016). Indeed, in general, a shift from SI to SC is strongly selected during domestication (Rowlands, 1964). Apricots from the Central Asian natural populations are self-incompatible while most of the modern and traditional European cultivars are, in contrast, self-compatible. This change in mating system was linked with the emergence of a major, dominant allele at the *SI* locus (a 358-bp long insertion resulting in a loss-of-function mutation within the pollen gene *SFB*), during apricot dispersion over the Caucasian/Eastern Anatolian region towards Europe, long before the southern and northern European dissemination routes of apricot diverged (Halász et al., 2007). In comparison, only 10% of the Chinese cultivars are self-compatible (He et al., 2007) and the origin of the *S_c* allele in the Chinese germplasm is still unknown. This current shift in the mating system in China could come from a recent introduction of *P. armeniaca* *S_c* allele(s) from Europe (through modern breeding programs), from an introgression of *P. mume* *S_c* allele (a 6.8 Kb insertion in the middle of the *SFB_f* coding region, Ushijima et al. 2004) or be conferred by an independent mutation affecting either the RNase or the F-box proteins. In regards with our results, it would be interesting to deepen our knowledge on those genes that were specifically under selection either in European apricots or in Chinese cultivars. Most probably, this would provide unique insights into the domestication traits targeted during apricot evolution but also into the adaptive traits under domestication.

Outlook and outstanding questions for the future

The release of the *P. armeniaca* genome, its annotation and comparative population genomics data sets provide an unprecedented opportunity for the identification and utilization of adaptive and domestication traits that are important for apricot cultivation. Moreover, our data show that apricot is an excellent model to unravel evolutionary processes at play during divergence and domestication of long-lived perennial fruit trees. Local adaptation, domestication and diversification related genes could be better characterized by integrating knowledge on the underlying biological pathways, perhaps through further investigation on nowadays genetic diversity among the section *Armeniaca* or even the subgenus *Prunus*. For example, it would be interesting to analyse the allelic variation of the above candidate genes for cold requirement and photoperiod response in a diverse panel of high and low latitude apricot germplasm and to validate the association of this allelic variation with phenologic and agronomic traits. We must be careful not to assume adaptation simply because a gene correlates with a trait of agronomic importance, therefore the correlation between allelic and phenotypic variation would help to confirm or not

GENERAL DISCUSSION and CONCLUSION

preliminary results obtained during my PhD. Nevertheless, those adaptive alleles could be used in the future to develop new apricot varieties better adapted to a fluctuating environment.

In addition to the above considerations, a few outstanding questions remain, as follow:

1) What was the genome architecture and the geographic distribution of the hypothetical ancestral population of *Armeniaca* species? 2) Which *Armeniaca* lineage contributed mostly to the Chinese cultivated apricots? 3) What is the extent and impact of the genome-wide wild-to-crop and/or interspecific introgression in apricot local adaptation? 4) Which level of genetic load (cost of domestication as an accumulation of deleterious mutations) in cultivated European and Chinese apricots? 5) What is the impact of domestication and other evolutionary processes on large-scale chromosomal variations in the apricot genome? How is this related to adaptation?

Finally, although the deciphering of the biological pathways involved in different processes of apricot evolution is still in its infancy, the preliminary results presented in this PhD manuscript are expected to be of potential agronomic and fundamental benefits, providing a unique opportunity to identify the genetic basis of long-lived perennials' adaptation and domestication.

References

- Abdala-Roberts, L., Moreira, X., Rasmann, S., Parra-Tabla, V., & Mooney, K. A. (2016). Test of biotic and abiotic correlates of latitudinal variation in defences in the perennial herb *Ruellia nudiflora*. *Journal of Ecology*, *104*(2), 580-590. doi:10.1111/1365-2745.12512
- Adams, W. T., & Burczyk, J. (2000). Magnitude and implications of gene flow in gene conservation reserves (pp. 215-224). Wallingford: CABI Publishing.
- Aharoni, A., Giri, A. P., Verstappen, F. W. A., Berteaux, C. M., Sevenier, R., Sun, Z., . . . Bouwmeester, H. J. (2004). Gain and loss of fruit flavor compounds produced by wild and cultivated strawberry species. *The Plant Cell*, *16*(11), 3110-3131. doi:10.1105/tpc.104.023895
- Ait Mouheb, H., Kadik, L., Albert, C. H., Berrached, R., & Prinzing, A. (2018). How do steppe plants follow their optimal environmental conditions or persist under suboptimal conditions? The differing strategies of annuals and perennials. *Ecology and Evolution*, *8*(1), 135-149. doi:10.1002/ece3.3664
- Akagi, T., Hanada, T., Yaegaki, H., Gradziel, T. M., & Tao, R. (2016). Genome-wide view of genetic diversity reveals paths of selection and cultivar differentiation in peach domestication. *DNA Research*, *23*(3), 271-282. doi:10.1093/dnares/dsw014
- Allan, A. C., & Espley, R. V. (2018). MYBs drive novel consumer traits in fruits and vegetables. *Trends in plant science*, *23*(8), 693-705. doi:10.1016/j.tplants.2018.06.001
- Amasino, R. (2009). Floral induction and monocarpic versus polycarpic life histories. *Genome biology*, *10*(7), 228. doi:10.1186/gb-2009-10-7-228
- Andersson, L. (1990). The driving force: species concepts and ecology. *Taxon*, *39*(3), 375-382. doi:10.2307/1223084
- Aradhya, M., Velasco, D., Ibrahimov, Z., Toktoraliev, B., Maghradze, D., Musayev, M., . . . Preece, J. E. (2017). Genetic and ecological insights into glacial refugia of walnut (*Juglans regia* L.). *PLoS One*, *12*(10), e0185974. doi:10.1371/journal.pone.0185974
- Arendt, J., & Reznick, D. (2008). Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? *Trends in Ecology & Evolution*, *23*(1), 26-32. doi:10.1016/j.tree.2007.09.011
- Arnold, M. L. (1997). *Natural hybridization and evolution*: Oxford University Press on Demand.
- Arnold, M. L., & Burke, J. M. (2006). Natural hybridization. In C. W. F. a. J. B. Wolf (Ed.), *Evolutionary Genetics: Concepts and Case Studies* (pp. 399-413): Oxford University Press.
- Bai, W. N., Liao, W. J., & Zhang, D. Y. (2010). Nuclear and chloroplast DNA phylogeography reveal two refuge areas with asymmetrical gene flow in a temperate walnut tree from East Asia. *New Phytologist*, *188*(3), 892-901. doi:10.1111/j.1469-8137.2010.03407.x
- Bailey, C. H., & Hough, L. F. (1975). Apricots. In J. Janick & J. N. Moore (Eds.), *In Advances in Fruit Breeding* (pp. 367-383.). West Lafayette, In: Purdue University Press.
- Bailey, L. H. (1916). Prunus. In J. H. McFarland (Ed.), *The standard cyclopedia of horticulture* (Vol. V. P-R. Mount, pp. 2822-2845). Harrisburg, PA: Pleasant Press.
- Bailey, L. H. (1927). *The standard cyclopedia of horticulture*. New York: Macmillan Press.
- Barrett, S. C. H., Harder, L. D., & Worley, A. C. (1996). The comparative biology of pollination and mating in flowering plants. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *351*(1345), 1271-1280.
- Bennett, B. C. (2010). Plant domestication and the origins of agriculture. In B. Bennett (Ed.), *Economic Botany in Encyclopedia of Life Support Systems (EOLSS), Developed under the Auspices of the UNESCO*. Paris: Eolss.
- Berg, B. O., & Lahav, E. (1996). "Avocados," in Fruit Breeding. In J. Janick & J. N. Moore (Eds.), *Tree and Tropical Fruits* (pp. 113-166). Hoboken: John Wiley & Sons.

GENERAL DISCUSSION and CONCLUSION

- Bock, W. J. (2004). Species: the concept, category and taxon. *Journal of Zoological Systematics and Evolutionary Research*, 42(3), 178-190.
- Bosco, F. A., Castro, D., & Briones, M. R. (2012). Neutral and stable equilibria of genetic systems and the Hardy–Weinberg principle: limitations of the chi-square test and advantages of auto-correlation functions of allele frequencies. *Frontiers in genetics*, 3, 276. doi:10.3389/fgene.2012.00276
- Bouché, F., Lobet, G., Tocquin, P., & Périlleux, C. (2015). FLOR-ID: an interactive database of flowering-time gene networks in *Arabidopsis thaliana*. *Nucleic Acids Research*, 44(D1), D1167-D1171. doi:10.1093/nar/gkv1054
- Bourguiba, H., Audergon, J. M., Krichen, L., Trifi-Farah, N., Mamouni, A., Trabelsi, S., . . . Khadari, B. (2012). Loss of genetic diversity as a signature of apricot domestication and diffusion into the Mediterranean Basin. *BMC Plant Biology*, 12(1), 49. doi:10.1186/1471-2229-12-49
- Brambilla, V., Gomez-Ariza, J., Cerise, M., & Fornara, F. (2017). The importance of being on time: regulatory networks controlling photoperiodic flowering in cereals. *Frontiers in Plant Science*, 8(665). doi:10.3389/fpls.2017.00665
- Butelli, E., Licciardello, C., Zhang, Y., Liu, J. J., Mackay, S., Bailey, P., . . . Martin, C. (2012). Retrotransposons control fruit-specific, cold-dependent accumulation of anthocyanins in blood oranges. *The Plant Cell*, 24(3), 1242-1255. doi:10.1105/tpc.111.095232
- Campbell, N., & Reece, J. (2002). *Biology*. Benjamin Cummings.
- Campoy, J. A., Le Dantec, L., Barreneche, T., Dirlwanger, E., & Quero-García, J. (2015). New insights into fruit firmness and weight control in sweet cherry. *Plant Molecular Biology Reporter*, 33(4), 783-796. doi:10.1007/s11105-014-0773-6
- Cao, K., Li, Y., Deng, C. H., Gardiner, S. E., Zhu, G. R., Fang, W. C., . . . Wang, L. R. (2019). Comparative population genomics identified genomic regions and candidate genes associated with fruit domestication traits in peach. *Plant Biotechnology Journal*. doi:10.1111/pbi.13112.
- Cao, K., Zheng, Z. J., Wang, L. R., Liu, X., Zhu, G. R., Fang, W. C., . . . Wang, X. W. (2014). Comparative population genomics reveals the domestication history of the peach, *Prunus persica*, and human influences on perennial fruit crops. *Genome biology*, 15(7), 415. doi:10.1186/s13059-014-0415-1
- Charlesworth, B., Nordborg, M., & Charlesworth, D. (1997). The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genetics Research*, 70(2), 155-174.
- Chen, J. J. (2010). The Hardy-Weinberg principle and its applications in modern population genetics. *Frontiers in Biology*, 5(4), 348-353. doi:10.1007/s11515-010-0580-x
- Chin, S.-W., Shaw, J., Haberle, R., Wen, J., & Potter, D. (2014). Diversification of almonds, peaches, plums and cherries – Molecular systematics and biogeographic history of *Prunus* (Rosaceae). *Molecular Phylogenetics and Evolution*, 76, 34-48. doi:10.1016/j.ympev.2014.02.024
- Cornille, A., Giraud, T., Smulders, M., Roldán-Ruiz, I., & Gladieux, P. (2014). The domestication and evolutionary ecology of apples. *Trends in Genetics*, 30(2), 57-65. doi:10.1016/j.tig.2013.10.002
- Cornille, A., Gladieux, P., Smulders, M. J., Roldan-Ruiz, I., Laurens, F., Le Cam, B., . . . Feugey, L. (2012). New insight into the history of domesticated apple: secondary contribution of the European wild apple to the genome of cultivated varieties. *PLoS genetics*, 8(5), e1002703. doi:10.1371/journal.pgen.1002703
- Cracraft, J. (1983). Species concepts and speciation analysis *Current Ornithology* (pp. 159-187): Springer.
- Cracraft, J. (1989). Speciation and its ontology: the empirical consequences of alternative species concepts for understanding patterns and processes of differentiation. *Speciation and its Consequences*, 28, 59.
- Cronquist, A. (1978). *Once again, what is a species?* *Biosystematics in agriculture*. Paper presented at the Beltsville Symposia in Agricultural Research.

- Cultrera, N. G. M., Sarri, V., Lucentini, L., Ceccarelli, M., Alagna, F., Mariotti, R., . . . Baldoni, L. (2019). High levels of variation within gene sequences of *olea europaea* L. *Frontiers in Plant Science*, 9, 1932. doi:10.3389/fpls.2018.01932
- Darwin, C., & Irvine, C. (1904). *The origin of species by means of natural selection: or, the preservation of favoured races in the struggle for life*. D. Appleton.
- Darwin, C., & Wallace, A. (1858). On the tendency of species to form varieties; and on the perpetuation of varieties and species by natural means of selection. *Journal of the proceedings of the Linnean Society of London. Zoology*, 3(9), 45-62.
- David, E. Z., Thomas, J. M., Sasha, W. E., Timothy, M. F., Ravza, F. M., John, M. C., . . . Joseph, C. G. (2013). Genetic resources of apricots (*Prunus armeniaca* L.) in Central Asia. *HortScience*, 48(6), 681-691. doi:10.21273/HORTSCI.48.6.681
- de Tournefort, J. P. (1700). *Institutiones rei herbariae*: s.n.
- Decroocq, S., Cornille, A., Tricon, D., Babayeva, S., Chague, A., Eyquard, J. P., . . . Decroocq, V. (2016). New insights into the history of domesticated and wild apricots and its contribution to Plum pox virus resistance. *Molecular Ecology*, 25(19), 4712-4729. doi:10.1111/mec.13772
- Delplancke, M., Yazbek, M., Arrigo, N., Espíndola, A., Joly, H., & Alvarez, N. (2016). Combining conservative and variable markers to infer the evolutionary history of *Prunus* subgen. *Amygdalus* s.l. under domestication. *Genetic Resources and Crop Evolution*, 63(2), 221-234. doi:10.1007/s10722-015-0242-6
- Diamond, J. (2002). Evolution, consequences and future of plant and animal domestication. *Nature*, 418(6898), 700-707. doi:10.1038/nature01019
- Ding, J., & Nilsson, O. (2016). Molecular regulation of phenology in trees-because the seasons they are a-changin'. *Current Opinion in Biotechnology*, 29, 73-79. doi:10.1016/j.pbi.2015.11.007
- Du, F., Xu, J. N., Li, D., & Wang, X. Y. (2015). The identification of novel and differentially expressed apple-tree genes under low-temperature stress using high-throughput Illumina sequencing. *Mol Biol Rep*, 42(3), 569-580. doi:10.1007/s11033-014-3802-5
- Duan, N., Bai, Y., Sun, H., Wang, N., Ma, Y., Li, M., . . . Chen, X. (2017). Genome re-sequencing reveals the history of apple and supports a two-stage model for fruit enlargement. *Nature Communications*, 8(1), 249. doi:10.1038/s41467-017-00336-7
- Falavigna, V. d. S., Guitton, B., Costes, E., & Andrés, F. (2019). I want to (Bud) break free: the potential role of DAM and SVP-Like genes in regulating dormancy cycle in temperate fruit trees. *Frontiers in Plant Science*, 9(1990). doi:10.3389/fpls.2018.01990
- Fan, S. Q., Liang, T., Yu, H. Y., Bi, Q. X., Li, G. T., & Wang, L. B. (2016). Kernel characteristics, oil contents, fatty acid compositions and biodiesel properties in developing Siberian apricot (*Prunus sibirica* L.) seeds. *Industrial Crops and Products*, 89, 195-199. doi:10.1016/j.indcrop.2016.05.012
- FAO. (2017). Save and Grow. *A policymaker's guide to the sustainable intensification of smallholder crop production*. Retrieved from <http://www.fao.org/docrep/014/i2215e/i2215e.pdf>
- Faust, M., S. D., & N, F. (1998). Origin and dissemination of apricot *Horticultural Reviews* (Vol. 22, pp. 225-260). NEW YORK: WESTPORT.
- Fournier-Level, A., Lacombe, T., Le Cunff, L., Boursiquot, J., & This, P. (2010). Evolution of the VvMybA gene family, the major determinant of berry colour in cultivated grapevine (*Vitis vinifera* L.). *Heredity*, 104(4), 351. doi:10.1038/hdy.2009.148
- Frejaville, T., Fady, B., Kremer, A., Ducouso, A., & Garzon, M. B. (2019). Inferring phenotypic plasticity and local adaptation to climate across tree species ranges using forest inventory data. *bioRxiv*, 527390. doi:10.1101/527390
- Friedman, J., & Rubin, M. J. (2015). All in good time: understanding annual and perennial strategies in plants. *American Journal of Botany*, 102(4), 497-499. doi:10.3732/ajb.1500062

GENERAL DISCUSSION and CONCLUSION

- Fuller, D. Q., Harvey, E., & Qin, L. (2015). Presumed domestication? Evidence for wild rice cultivation and domestication in the fifth millennium BC of the Lower Yangtze region. *Antiquity*, 81(312), 316-331. doi:10.1017/S0003598X0009520X
- Gao, Z. H., & Luo, W. J. (2019). Origin and evolution of *Prunus mume*. In Z. Gao (Ed.), *The Prunus mume Genome* (pp. 5-7). Cham: Springer International Publishing.
- Garnier, E., & Laurent, G. (1994). Leaf anatomy, specific mass and water content in congeneric annual and perennial grass species. *New Phytologist*, 128(4), 725-736.
- Gaut, B. S. (2015). Evolution is an experiment: assessing parallelism in crop domestication and experimental evolution: (Nei Lecture, SMBE 2014, Puerto Rico). *Molecular biology and evolution*, 32(7), 1661-1671. doi:10.1093/molbev/msv105
- Gendall, A. R., Levy, Y. Y., Wilson, A., & Dean, C. (2001). The VERNALIZATION 2 gene mediates the epigenetic regulation of vernalization in Arabidopsis. *Cell*, 107(4), 525-535. doi:10.1016/s0092-8674(01)00573-6
- Gentile, A., Monticelli, S., & Damiano, C. (2002). Adventitious shoot regeneration in peach [*Prunus persica* (L.) Batsch]. *Plant and Cell Physiology*, 20(11), 1011-1016. doi:10.1007/s00299-002-0451-2
- Gepts, P. (2014). The contribution of genetic and genomic approaches to plant domestication studies. *Current opinion in plant biology*, 18, 51-59. doi:10.1016/j.pbi.2014.02.001
- Gepts, P., & Papa, R. (2003). Evolution during Domestication. *Encyclopedia of Life Sciences*. doi:10.1038/npg.els.0003071
- Godfrey, J. M., Ferguson, L., Sanden, B. L., Tixier, A., Sperling, O., Grattan, S. R., & Zwieniecki, M. A. (2019). Sodium interception by xylem parenchyma and chloride recirculation in phloem may augment exclusion in the salt tolerant Pistacia genus: context for salinity studies on tree crops. *Tree Physiology*. doi:10.1093/treephys/tpz054
- Graf, D. F. (1994). *The Persian royal road system* (Vol. 8).
- Gros-Balthazard, M., Besnard, G., Sarah, G., Holtz, Y., Leclercq, J., Santoni, S., . . . Khadari, B. (2019). Evolutionary transcriptomics reveals the origins of olives and the genomic changes associated with their domestication. *The Plant Journal*. doi:10.1111/tpj.14435
- Gross, B. L., & Olsen, K. M. (2010). Genetic perspectives on crop domestication. *Trends in plant science*, 15(9), 529-537. doi:10.1016/j.tplants.2010.05.008
- Guerra-García, A., Suárez-Atilano, M., Mastretta-Yanes, A., Delgado-Salinas, A., & Piñero, D. (2017). Domestication genomics of the open-pollinated scarlet runner bean (*Phaseolus coccineus* L.). *Frontiers in Plant Science*, 8(1891). doi:10.3389/fpls.2017.01891
- Guevara, M. Á., Soto, A., Collada, C., Plomion, C., Savolainen, O., B Neale, D., . . . Cervera, M. T. (2005). Genomics applied to the study of adaptation in pine species. *Investigación agraria. Sistemas y recursos forestales*, 14, 292-306. doi:10.5424/srf/2005143-00918
- H, B. L. (1927). *The standard cyclopedia of horticulture*. New York: Macmillan Press.
- Hagen, L., Khadari, B., Lambert, P., & Audergon, J. M. (2002). Genetic diversity in apricot revealed by AFLP markers: species and cultivar comparisons. *Theoretical and Applied Genetics*, 105(2), 298-305. doi:10.1007/s00122-002-0910-8
- Halász, J., Pedryc, A., & Hegedűs, A. (2007). Origin and dissemination of the pollen-part mutated SC haplotype which confers self-compatibility in apricot (*Prunus armeniaca*). *New Phytologist*, 176(4), 792-803. doi:10.1111/j.1469-8137.2007.02220.x
- Hardy, G. H. (1908). *Mendelian proportions in a mixed population*. NJ: Englewood Cliffs.
- Harlan, J. R., Gepts, P., Famula, T. R., Bettinger, R. L., Brush, S. B., Damania, A. B., . . . Qualset, C. O. (2012). *Biodiversity in agriculture: domestication, evolution, and sustainability*. Cambridge University Press.

- Harrison, R. G. (1990). Hybrid zones: windows on evolutionary process. *Oxford surveys in evolutionary biology*, 7, 69-128.
- Harrison, R. G. (1993). Hybrids and hybrid zones: historical perspective. *Hybrid zones and the evolutionary process*, 3-12.
- Hartl, D. L., Clark, A. G., & Clark, A. G. (1997). *Principles of population genetics* (Vol. 116): Sinauer associates Sunderland, MA.
- Hausdorf, B. (2011). Progress toward a general species concept. *Evolution: International Journal of Organic Evolution*, 65(4), 923-931. doi:10.1111/j.1558-5646.2011.01231.x
- Hazzouri, K. M., Flowers, J. M., Visser, H. J., Khierallah, H. S., Rosas, U., Pham, G. M., . . . Masmoudi, K. (2015). Whole genome re-sequencing of date palms yields insights into diversification of a fruit tree crop. *Nature Communications*, 6, 8824. doi:10.1038/ncomms9824
- He, T. M., Chen, X. S., Zhang, D. H., Xu, L., Liu, N., Gao, J. S., & Xu, Z. (2007). Frequency distribution of several biological characters in different apricot eco-geographical groups native to China. *Acta Horticulturae Sinica*, 34(1), 17.
- Hedrick, P. W. (2006). Genetic polymorphism in heterogeneous environments: the age of genomics. *Annual Review of Ecology, Evolution, and Systematics*, 37(1), 67-93. doi:10.1146/annurev.ecolsys.37.091305.110132
- Heslop-Harrison, J. S., & Schwarzacher, T. (2007). Domestication, genomics and the future for banana. *Annals of botany*, 100(5), 1073-1084. doi:10.1093/aob/mcm191
- Hilu, K. W. (1993). Polyploidy and the evolution of domesticated plants. *American Journal of Botany*, 80(12), 1494-1499.
- Hormaza, J. I., Yamane, H., & Rodrigo, J. (2007). Apricot. In C. Kole (Ed.), *Fruits and Nuts* (pp. 171-187). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Huang, J., Zhang, C. M., Zhao, X., Fei, Z. J., Wan, K. K., Zhang, Z., . . . Sun, X. Q. (2016). The jujube genome provides insights into genome evolution and the domestication of sweetness/acidity taste in fruit trees. *PLoS genetics*, 12(12), e1006433. doi:10.1371/journal.pgen.1006433
- Hughes, C. E., Govindarajulu, R., Robertson, A., Filer, D. L., Harris, S. A., & Bailey, C. D. (2007). Serendipitous backyard hybridization and the origin of crops. *Proceedings of the National Academy of Sciences*, 104(36), 14389-14394. doi:10.1073/pnas.0702193104
- Husband, B. C., & Schemske, D. W. (1996). Evolution of the magnitude and timing of inbreeding depression in plants. *Evolution*, 50(1), 54-70.
- Iorizzo, M., Senalik, D. A., Ellison, S. L., Grzebelus, D., Cavagnaro, P. F., Allender, C., . . . Simon, P. W. (2013). Genetic structure and domestication of carrot (*Daucus carota* subsp. *sativus*) (Apiaceae). *American Journal of Botany*, 100(5), 930-938. doi:10.3732/ajb.1300055
- Iwata, H., Gaston, A., Remay, A., Thouroude, T., Jeauffre, J., Kawamura, K., . . . Foucher, F. (2012). The TFL1 homologue KSN is a regulator of continuous flowering in rose and strawberry. *The Plant Journal*, 69(1), 116-125. doi:10.1111/j.1365-313X.2011.04776.x
- Janick, J., & Moore, J. N. (1996). *Fruit breeding, tree and tropical fruits* (Vol. 1): John Wiley & Sons.
- Janzen, G. M., Wang, L., & Hufford, M. B. (2019). The extent of adaptive wild introgression in crops. *New Phytologist*, 221(3), 1279-1288. doi:10.1111/nph.15457
- Kalkman, C. (2008). The phylogeny of the Rosaceae. *Botanical Journal of the Linnean Society*, 98(1), 37-59. doi:10.1111/j.1095-8339.1988.tb01693.x
- Kawecki, T. J., & Ebert, D. (2004). Conceptual issues in local adaptation. *Ecology Letters*, 7(12), 1225-1241. doi:10.1111/j.1461-0248.2004.00684.x
- Kawuki, R. S., Ferguson, M., Labuschagne, M., Herselman, L., & Kim, D.-J. (2009). Identification, characterisation and application of single nucleotide polymorphisms for diversity assessment in

GENERAL DISCUSSION and CONCLUSION

- cassava (*Manihot esculenta* Crantz). *Molecular Breeding*, 23(4), 669-684. doi:10.1007/s11032-009-9264-0
- Kerschbamer, E. (2015). Identification of selective sweeps in domesticated apple (*Malus× domestica* Borkh.). doi:10.1007/s11032-009-9264-0
- Kislev, M. E., Hartmann, A., & Bar-Yosef, O. (2006). Early domesticated fig in the Jordan Valley. *Science*, 312(5778), 1372-1374. doi:10.1126/science.1125910
- Kostina, K. F. (1964). Application of phytogeographical method to apricot classification. *Trud. Nikit. Bot. Sad*, 37, 45–63.
- L, L., G, C., L, C., C, A., B, B., R, B. A., & A, S. S. (2003). *Flora of China*. Beijing: Science Press and Missouri Botanical Garden Press.
- Labonte, N. R., Zhao, P., & Woeste, K. (2018). Signatures of selection in the genomes of Chinese chestnut (*Castanea mollissima* Blume): the roots of nut tree domestication. *Frontiers in Plant Science*, 9, 810. doi:10.3389/fpls.2018.00810
- Lenser, T., & Theißen, G. (2013). Molecular mechanisms involved in convergent crop domestication. *Trends in plant science*, 18. doi:10.1016/j.tplants.2013.08.007
- Lerma, L. C., García, M. P. S., Cabanás, V. M. F., Sillero, A. M. M., González, R. J., & Morillo, P. R. (2014). From the juvenile to the adult vegetative phase in olive seedlings: the transition along the stem axis. *Spanish journal of agricultural research*(4), 1149-1157. doi:10.5424/sjar/2014124-6363
- Li, L. F., & Olsen, K. M. (2016). To have and to hold: selection for seed and fruit retention during crop domestication *Current Topics in Developmental Biology* (Vol. 119, pp. 63-109): Elsevier.
- Li, W., & Gill, B. S. (2006). Multiple genetic pathways for seed shattering in the grasses. *Funct Integr Genomics*, 6(4), 300-309. doi:10.1007/s10142-005-0015-y
- Li, W. B., Ding, Z. H., Ruan, M. B., Yu, X. L., Peng, M., & Liu, Y. F. (2017). Kiwifruit R2R3-MYB transcription factors and contribution of the novel AcMYB75 to red kiwifruit anthocyanin biosynthesis. *Scientific Reports*, 7(1), 16861. doi:10.1038/s41598-017-16905-1
- Li, X. L., Liu, L., Ming, M. L., Hu, H. J., Zhang, M. Y., Fan, J., . . . Wu, J. (2019). Comparative transcriptomic analysis provides insight into the domestication and improvement of pear (*P. pyrifolia*) fruit. *Plant physiology*, 180(1), 435-452. doi:10.1104/pp.18.01322
- Li, X. Y., Shangguan, L. F., Wang, Y. J., Korir, N., Wang, C., Zhang, J., . . . Fang, J. G. (2011). Evidence of pre-zygotic reproductive isolation between *Prunus armeniaca* and *Prunus mume*. *Evolutionary Ecology Research*, 13, 725-746.
- Li, Y., Cao, K., Zhu, G. R., Fang, W. C., Chen, C. W., Wang, X. W., . . . Wang, L. R. (2019). Genomic analyses of an extensive collection of wild and cultivated accessions provide new insights into peach breeding history. *Genome biology*, 20(1), 36. doi:10.1186/s13059-019-1648-9
- Li, Z. G., Reighard, G. L., Abbott, A. G., & Bielenberg, D. G. (2009). Dormancy-associated MADS genes from the EVG locus of peach [*Prunus persica* (L.) Batsch] have distinct seasonal and photoperiodic expression patterns. *Journal of Experimental Botany*, 60(12), 3521-3530. doi:10.1093/jxb/erp195
- Liang, Z. C., Duan, S. C., Sheng, J., Zhu, S. S., Ni, X. M., Shao, J. H., . . . Fan, P. G. (2019). Whole-genome resequencing of 472 *Vitis* accessions for grapevine diversity and demographic history analyses. *Nature Communications*, 10. doi:10.1038/s41467-019-09135-8
- Lin, Z., Li, X., Shannon, L. M., Yeh, C. T., Wang, M. L., Bai, G., . . . Yu, J. (2012). Parallel domestication of the Shattering1 genes in cereals. *Nature Genetics*, 44(6), 720-724. doi:10.1038/ng.2281
- Linhart, Y. B., & Grant, M. C. (1996). Evolutionary significance of local genetic differentiation in plants. *Annual Review of Ecology and Systematics*, 27(1), 237-277. doi:10.1146/annurev.ecolsys.27.1.237
- Linnaeus, C., & Clifford, G. (1737). *Hortus cliffortianus*.

- Liu, J., & Zhou, M. X. (2018). The ALMT gene family performs multiple functions in plants. *Agronomy*, 8, 20. doi:10.3390/agronomy8020020
- Liu, W., Liu, N., Zhang, Y., Yu, X., Sun, M., M, X., . . . Liu, S. (2012). Kernel-using apricot resources and its utilization. In A. A (Ed.), *XV International Symposium on Apricot Breeding and Culture* (Vol. 966, pp. 189–191). Leuven, Belgium: Soc. Horticultural Science.
- Mallet, J. (1995). A species definition for the modern synthesis. *Trends in Ecology & Evolution*, 10(7), 294-299. doi:10.1016/0169-5347(95)90031-4
- Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends in Ecology & Evolution*, 20(5), 229-237. doi:10.1016/j.tree.2005.02.010
- Marrano, A., Micheletti, D., Lorenzi, S., Neale, D., & Grando, M. S. (2018). Genomic signatures of different adaptations to environmental stimuli between wild and cultivated *Vitis vinifera* L. *Horticulture research*, 5(1), 34. doi:10.1038/s41438-018-0041-2
- Martin, A., & Orgogozo, V. (2013). The Loci of repeated evolution: A catalog of genetic hotspots of phenotypic variation. *Evolution*, 67(5), 1235-1250. doi:10.1111/evo.12081
- Matsuoka, Y., Vigouroux, Y., Goodman, M. M., Sanchez, J., Buckler, E., & Doebley, J. (2002). A single domestication for maize shown by multilocus microsatellite genotyping. *Proceedings of the National Academy of Sciences*, 99(9), 6080-6084. doi:10.1073/pnas.052125199
- Maynard, D. N. (1999). Horticultural reviews. *HortScience*, 22(34), 1131–1131.
- Mayr, E. (1942). *Systematics and the origin of species*. New York: Columbia University Press.
- Mayr, E. (1991). *One long argument: Charles Darwin and the genesis of modern evolutionary thought* (Vol. 2): Harvard University Press.
- McClure, K. A., Sawler, J., Gardner, K. M., Money, D., & Myles, S. (2014). Genomics: a potential panacea for the perennial problem. *American Journal of Botany*, 101(10), 1780-1790. doi:10.3732/ajb.1400143
- McCown, B. H. (2000). Special symposium: In vitro plant recalcitrance of woody and herbaceous perennial plants: Dealing with genetic predeterminism. *In Vitro Cellular & Developmental Biology-Plant*, 36(3), 149-154. doi:10.1007/s11627-000-0030-6
- Mehlenbacher, S. A., Cociu, V., & Hough, F. L. (1991). APRICOTS (PRUNUS). *Acta Horticulturae*, 290, 65-110. doi:10.17660/ActaHortic.1991.290.3
- Melnyk, C. W., & Meyerowitz, E. M. (2015). Plant grafting. *Current Biology*, 25(5), R183-R188. doi:10.1016/j.cub.2015.01.029
- Meyer, R. S., DuVal, A. E., & Jensen, H. R. (2012). Patterns and processes in crop domestication: an historical review and quantitative analysis of 203 global food crops. *New Phytologist*, 196(1), 29-48. doi:10.1111/j.1469-8137.2012.04253
- Meyer, R. S., & Purugganan, M. D. (2013). Evolution of crop species: genetics of domestication and diversification. *Nature Reviews Genetics*, 14, 840. doi:10.1038/nrg3605
- Miller, A. J., & Gross, B. L. (2011). From forest to field: Perennial fruit crop domestication. *American Journal of Botany*, 98(9), 1389-1414. doi:10.3732/ajb.1000522
- Myles, S., Boyko, A. R., Owens, C. L., Brown, P. J., Grassi, F., Aradhya, M. K., . . . Buckler, E. S. (2011). Genetic structure and domestication history of the grape. *Proceedings of the National Academy of Sciences*, 108(9), 3530-3535. doi:10.1073/pnas.1009363108
- Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences*, 70(12), 3321-3323. doi:10.1073/pnas.70.12.3321
- Nelson, G., & Platnick, N. I. (1981). *Systematics and biogeography*. Columbia Univ. Press.
- Nielsen, R. (2005). Molecular signatures of natural selection. *The Annual Review of Genetics*, 39, 197-218. doi:10.1146/annurev.genet.39.073003.112420

GENERAL DISCUSSION and CONCLUSION

- Paauw, M., Koes, R., & Quattrocchio, F. M. (2019). Alteration of flavonoid pigmentation patterns during domestication of food crops. *Journal of Experimental Botany*, *70*(15), 3719-3735. doi:10.1093/jxb/erz141
- Peil, A., Dunemann, F., Richter, K., Hofer, M., Király, I., & Flachowsky, H. (2008, 18th February to 20th February 2008). *Resistance breeding in apple at Dresden-Pillnitz*. Paper presented at the Ecofruit-13th International Conference on Cultivation Technique and Phytopathological Problems in Organic Fruit-Growing, Weinsberg.
- Pignatti, S. (1982). *Flora d'Italia* (Vol. I, II, III.). Bologna, Italy.: Edagricole.
- Price, T. D., & Bar-Yosef, O. (2011). The origins of agriculture: new data, new ideas: an introduction to supplement 4. *Current Anthropology*, *52*(S4), S163-S174. doi:10.1086/659964
- Purugganan, M. D., & Fuller, D. Q. (2009). The nature of selection during plant domestication. *Nature*, *457*(7231), 843. doi:10.1038/nature07895
- Quintana-Murci, L. (2016). Genetic and epigenetic variation of human populations: An adaptive tale. *Comptes rendus biologiques*, *339*(7-8), 278-283. doi:10.1016/j.crv.2016.04.005
- Quintana-Murci, L., & Clark, A. G. (2013). Population genetic tools for dissecting innate immunity in humans. *Nature Reviews Immunology*, *13*, 280. doi:10.1038/nri3421
- Rao, M. N., Soneji, J. R., Chen, C., Huang, S., & Gmitter, F. G. (2008). Characterization of zygotic and nucellar seedlings from sour orange-like citrus rootstock candidates using RAPD and EST-SSR markers. *Tree Genetics & Genomes*, *4*(1), 113-124. doi:10.1007/s11295-007-0092-2
- Raven, P. H., & Johnson, G. B. (2002). *Five agents of evolutionary change*. In *Biology* (6th ed.). Boston, MA: McGraw-Hill.
- Rehder, A. (1940). Manual of cultivated trees and shrubs hardy in North America: exclusive of the subtropical and warmer temperate regions. In: The Macmillan Company. Ridley, M. (1993). Evolution. *Journal of Evolutionary Biology*, *6*, 615-617.
- Ross-Ibarra, J., Morrell, P. L., & Gaut, B. S. (2007). Plant domestication, a unique opportunity to identify the genetic basis of adaptation. *Proceedings of the National Academy of Sciences*, *104*(suppl 1), 8641-8648. doi:10.1073/pnas.0700643104
- Rowlands, D. G. (1964). Self-incompatibility in sexually propagated cultivated plants. *Euphytica*, *13*(2), 157-162. doi:10.1007/BF00033304
- Saumitou-Laprade, P., Vernet, P., Vekemans, X., Billiard, S., Gallina, S., Essalouh, L., . . . Baldoni, L. (2017). Elucidation of the genetic architecture of self-incompatibility in olive: Evolutionary consequences and perspectives for orchard management. *Evolutionary Applications*, *10*(9), 867-880. doi:10.1111/eva.12457
- Schwenk, K., & Spaak, P. (1995). Evolutionary and ecological consequences of interspecific hybridization in cladocerans. *Experientia*, *51*(5), 465-481.
- Shim, J. S., Kubota, A., & Imaizumi, T. (2017). Circadian clock and photoperiodic flowering in Arabidopsis: CONSTANS is a hub for signal integration. *Plant physiology*, *173*(1), 5-15. doi:10.1104/pp.16.01327
- Shindo, C., Aranzana, M. J., Lister, C., Baxter, C., Nicholls, C., Nordborg, M., & Dean, C. (2005). Role of FRIGIDA and FLOWERING LOCUS C in determining variation in flowering time of Arabidopsis. *Plant physiology*, *138*(2), 1163-1173. doi:10.1104/pp.105.061309
- Simpson, G. G. (1961). Principles of animal taxonomy.
- Singh, R. K., Miskolczi, P., Maurya, J. P., & Bhalerao, R. P. (2019). A tree ortholog of SHORT VEGETATIVE PHASE floral repressor mediates photoperiodic control of bud dormancy. *Current Biology*, *29*(1), 128-133.e122. doi:10.1016/j.cub.2018.11.006
- Soltis, P. S., & Soltis, D. E. (2009). The role of hybridization in plant speciation. *Annual review of plant biology*, *60*, 561-588. doi:10.1146/annurev.arplant.043008.092039

- Stinchcombe, J. R., Weinig, C., Ungerer, M., Olsen, K. M., Mays, C., Halldorsdottir, S. S., . . . Schmitt, J. (2004). A latitudinal cline in flowering time in *Arabidopsis thaliana* modulated by the flowering time gene FRIGIDA. *Proceedings of the National Academy of Sciences*, 101(13), 4712-4717. doi:10.1073/pnas.0306401101
- Stoeckel, S., Grange, J., FERNÁNDEZ-MANJARRES, J. F., Bilger, I., FRASCARIA-LACOSTE, N., & Mariette, S. (2006). Heterozygote excess in a self-incompatible and partially clonal forest tree species—*Prunus avium* L. *Molecular Ecology*, 15(8), 2109-2118. doi:10.1111/j.1365-294X.2006.02926.x
- Toomajian, C., Hu, T. T., Aranzana, M. J., Lister, C., Tang, C., Zheng, H., . . . Nordborg, M. (2006). A nonparametric test reveals selection for rapid flowering in the *Arabidopsis* genome. *PLoS biology*, 4(5), e137. doi:10.1371/journal.pbio.0040137
- Ushijima, K., Yamane, H., Watari, A., Takechi, E., Ikeda, K., Hauck, N. R., . . . Tao, R. (2004). The S haplotype-specific F-box protein gene, SFB, is defective in self-compatible haplotypes of *Prunus avium* and *P. mume*. *The Plant Journal*, 39(4), 573-586. doi:10.1111/j.1365-313X.2004.02154.x
- Van Valen, L. (1976). Ecological species, multispecies, and oaks. *Taxon*, 233-239.
- Vavilov, N. I. (1951). Phytogeographic basis of plant breeding. The origin, variation, immunity and breeding of cultivated plants (tr. K.S. Chester). *Chron Bot*, 13: 366 p.
- Velasco, D., Hough, J., Aradhya, M., & Ross-Ibarra, J. (2016). Evolutionary genomics of peach and almond domestication. *G3 (Bethesda, Md.)*, 6(12), 3985-3993. doi:10.1534/g3.116.032672
- Verica, J. A., & He, Z. H. (2002). The cell wall-associated kinase (WAK) and WAK-like kinase gene family. *Plant physiology*, 129(2), 455-459. doi:10.1104/pp.011028
- Villars, D., Chovin, J. A., & Robert, J. (1789). Histoire des plantes de Dauphiné: contenant une préface historique, un dictionnaire des termes de botanique, les classes, les familles, les genres, & les herborisations des environs de Grenoble, de la Grande Chartreuse, de Briançon, de Gap & de Montelimar (Chez l'Auteur & Chez les Libraires).
- Vit, P., Wolfova, K., Urfus, T., Tajek, P., & Suda, J. (2014). Interspecific hybridization between rare and common plant congeners inferred from genome size data: assessing the threat to the Czech serpentine endemic *Cerastium alsinifolium*. *Preslia*, 86(1), 95-117.
- Voogd, C., Brian, L. A., Wang, T., Allan, A. C., & Varkonyi-Gasic, E. (2017). Three FT and multiple CEN and BFT genes regulate maturity, flowering, and vegetative phenology in kiwifruit. *Journal of Experimental Botany*, 68(7), 1539-1553. doi:10.1093/jxb/erx044
- Vukicevich, E., Lowery, T., Bowen, P., Úrbez-Torres, J. R., & Hart, M. (2016). Cover crops to increase soil microbial diversity and mitigate decline in perennial agriculture. A review. *Agronomy for Sustainable Development*, 36(3), 48. doi:10.1007/s13593-016-0385-7
- Walbank, F. W. (2019). Alexander the Great. Retrieved from <https://www.britannica.com/biography/Alexander-the-Great>.
- Wang, Z., Kang, M., Liu, H. B., Gao, J., Zhang, Z. D., Li, Y. Y., . . . Pang, X. M. (2014). High-Level genetic diversity and complex population structure of siberian apricot (*Prunus sibirica* L.) in China as revealed by nuclear SSR markers. *PLoS One*, 9(2), e87381. doi:10.1371/journal.pone.0087381
- Wang, Z., Zeng, Y. F., Zhang, Z. D., Sheng, S. B., Tian, J., Wu, R. L., & Pang, X. M. (2017). Phylogeography study of the siberian apricot (*Prunus sibirica* L.) in Northern China assessed by chloroplast microsatellite and DNA makers. *Frontiers in Plant Science*, 8, 1989-1989. doi:10.3389/fpls.2017.01989
- Warschewsky, E. J., Klein, L. L., Frank, M. H., Chitwood, D. H., Londo, J. P., von Wettberg, E. J. B., & Miller, A. J. (2016). Rootstocks: diversity, domestication, and impacts on shoot phenotypes. *Trends in plant science*, 21(5), 418-437. doi:10.1016/j.tplants.2015.11.008
- Weinberg, W. (1908). ber den Nachweis der Vererbung beim Menschen. *Jahres. Wiertt. Ver. Vaterl. Natkd.*, 64, 369-382.

GENERAL DISCUSSION and CONCLUSION

- Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, 38(6), 1358-1370. doi:10.2307/2408641
- Wiley, E. O. (1978). The evolutionary species concept reconsidered. *Systematic zoology*, 27(1), 17-26.
- Wiley, E. O. (1981). *Phylogenetics: The theory and practice of phylogenetics* (1st ed.). John Wiley and Sons, N. York.
- Wright, S. (1978). *Evolution and the genetics of populations. A treatise in four volumes. Volume 4. Variability within and among natural populations*. Chicago, London: University of Chicago Press.
- Wu, J., Wang, Y. T., Xu, J. B., Korban, S. S., Fei, Z. J., Tao, S. T., . . . Postman, J. D. (2018). Diversification and independent domestication of Asian and European pears. *Genome biology*, 19(1), 77. doi:10.1186/s13059-018-1452-y
- Wu, M., Lewis, J., & Moore, R. C. (2017). A wild origin of the loss-of-function lycopene beta cyclase (CYC-b) allele in cultivated, red-fleshed papaya (*Carica papaya*). *American Journal of Botany*, 104(1), 116-126. doi:10.3732/ajb.1600301
- Yu, Y., Fu, J., Xu, Y. G., Zhang, J. W., Ren, F., Zhao, H. W., . . . Xie, H. (2018). Genome re-sequencing reveals the evolutionary history of peach fruit edibility. *Nature Communications*, 9(1), 5404. doi:10.1038/s41467-018-07744-3
- Zaurov, D., Molnar, T., Eisenman, S., M. Ford, T., Mavlyanova, R., M. Capik, J., . . . C. Goffreda, J. (2013). Genetic resources of apricots (*Prunus armeniaca* L.) in Central Asia. *HortScience*, 48, 681-691. doi:10.21273/HORTSCI.48.6.681
- Zhang, Q., Wu, G., Zhao, Y., Lv, Y., Cheng, T., & Luo, L. (2010). Distribution of resources of wild mei (*Prunus mume* Sieb et Zucc) flower in Anhui, Jiangxi, Sichuan and Yunnan provinces of China. *Beijing For. Univ.*, 32, 8–13.
- Zhang, Q., Zhang, H., Sun, L., Fan, G., Ye, M., Jiang, L., . . . Cheng, T. (2018). The genetic architecture of floral traits in the woody plant *Prunus mume*. *Nature Communications*, 9(1), 1702. doi:10.1038/s41467-018-04093-z
- Zhang, Q. X., Chen, W. B., Sun, L. D., Zhao, F. Y., Huang, B. Q., Yang, W. R., . . . Wang, J. (2012). The genome of *Prunus mume*. *Nature Communications*, 3, 1318. doi:10.1038/ncomms2290
- Zhou, Y., Massonnet, M., Sanjak, J. S., Cantu, D., & Gaut, B. S. (2017). Evolutionary genomics of grape (*Vitis vinifera* ssp. *vinifera*) domestication. *Proceedings of the National Academy of Sciences*, 114(44), 11715-11720. doi:10.1073/pnas.1709257114
- Zohary, D., & Hopf, M. (2000). *Domestication of plants in the Old World: the origin and spread of cultivated plants in West Asia, Europe and the Nile Valley*. Oxford University Press.
- Zohary, D., & Spiegel-Roy, P. (1975). Beginnings of fruit growing in the old world. *Science*, 187(4174), 319-327. doi:10.1126/science.187.4174.319

ANNEXES

Annex 1: Chapter I supporting information

Supplemental notes

Supplementary note 1: Cautionary notes on the assignment of apricot individuals as ‘wild’ or ‘cultivated’ (landraces and/or modern cultivars)

The identification of genuine wild *Armeniaca* samples (including *P. armeniaca*, *P. sibirica*, *P. manshurica* and *P. mume*) can be challenging because of the existence of feral forms, the interfertility between wild and cultivated *Armeniaca* and the common use of wild seedlings for ornamental/rootstock purposes. To be conservative, we classified as wild only the samples resulting from collection expeditions in the mountains of Central Asia and China by S.L., D.T. and V.D. (Table I-1). Table I-S1 provides details of the *Armeniaca* (cultivated and wild) samples used in the present study. Latitude, longitude and elevation can be provided on request. For the *P. armeniaca* samples collected in Southern Central Asia, Pakistan, Turkmenistan and Afghanistan (Table I-S1, numbers US015 to US241), not enough information was available on the strategy used for collection, and on their wild/cultivated status. We classified them as ‘landraces’ because they carried names referring to localities or local varieties and their coordinates mapped along roads or in villages. Information on those samples can be accessed via the USDA Germplasm Resources Information Network (GRIN: <http://www.ars-grin.gov/>). Similarly, all *P. mume* samples were classified as ‘cultivated’ since they were not collected directly in the wild but were instead obtained from the Nanjing germplasm repository (Nanjing Agricultural University, China), thus most likely corresponding to ornamental landraces (also called ‘Mei’) as confirmed by the curator. Finally, misclassification is unlikely to affect our results as we checked assignment with the STRUCTURE analyses.

Supplementary note 2: General framework for defining focal population for ABC analyses

Supplementary note 2.1: Groups of samples used for running the ABC analysis, step 1 (“Wild divergence”)

We first investigated the evolutionary history of wild apricots, as the first step of ABC analyses. We performed new STRUCTURE analyses using only the 288 wild individuals (Figure I-S4) and the 25 ‘perfect’ microsatellites (Table I-S2). We recovered a pattern highly similar to the previous analyses, with differentiation of the wild *P. armeniaca* in South Kyrgyzstan (red, PaKGZ_W1), *P. armeniaca* in Northern Central Asia (yellow, PaNCA_W2), and the wild Northwestern (green, Psib_nw_W3) and Northeastern (dark blue, Psib_ne_W4) Chinese *P. sibirica* (Figure I-S4, Table I-S5). While additional clusters (white and purple)

split from Central Asian *P. armeniaca* at $K=6$, their sample size and differentiation from the yellow cluster were too small ($Jost's D=0.078$, $F_{ST}=0.032$) to be considered in ABC analyses for reconstructing their history. We also excluded *P. mume* from the ABC analyses because (i) it was the most genetically divergent species (Tables I-2 and I-3), and (ii) the sample size was too small ($n=9$). After excluding admixed genotypes (i.e. genotypes that did not assign to any cluster with a probability higher than 85%), we ended up with 234 individuals clustered into four main genetic groups for the ABC analysis step 1: PaKGZ_W1 ($n=19$), PaNCA_W2 ($n=138$), Psib_nw_W3 ($n=30$), Psib_ne_W4 ($n=47$) (Table I-S5).

Supplementary note 2.2: Groups of samples used for running the ABC analysis steps 2 and 3 (“Domestication” and “Bottleneck”, respectively):

In the second and third rounds of ABC analyses, we added all non-wild individuals ($n=262$, Tables I-1 and I-S1) together with the 234 wild individuals (Table I-S5, except for *P. mume*, *P. brigantina* and *P. mandshurica*). We performed STRUCTURE analyses based on the 25 ‘perfect’ microsatellite markers (Table I-S2 and Figure I-S5) that are used in the ABC analyses. The clustering of the 262 non-wild individuals was highly similar as the one found with the 34 microsatellite markers (Figure I-S6), indicating that $K=10$ was the most likely clustering solution. We therefore used the membership coefficient inferred for 25 SSR at $K=10$ to choose the individuals to be used in steps 2 and 3 of the ABC analysis. At $K=10$, four cultivated clusters appeared well delimited. We then again pruned admixed individuals using a threshold of 0.85 in membership assignment. Finally, we retained 116 cultivated apricots individuals assigned to three cultivated clusters for ABC analyses of apricot domestication, the cultivated *P. armeniaca* clusters from Europe ($n=63$, grey, PaEUR_C1), South Central Asia ($n=25$, light blue, PaSCA_C2) and China ($n=28$, brown, China_C4), together with the 234 wild individuals from ABC analysis step 1. In total, we therefore defined seven genetic groups representing a total of 350 individuals including the same wild groups as above (supplementary note 2.1), PaKGZ_W1 ($n=19$), PaNCA_W2 ($n=138$), Psib_nw_W3 ($n=30$), Psib_ne_W4 ($n=47$), PaEUR_C1 ($n=63$), PaSCA_C2 ($n=25$), China_C4 ($n=28$). We excluded PaXJ_C3 because of its small sampling size, i.e. $n=5$, and high admixture level with other clusters.

Supplementary note 3: Sets of scenarios to be compared using ABC

Supplementary note 3.1: ABC analysis step 1 - The evolution of wild apricots

Based on the clustering analyses (234 individuals clustered in four main genetic groups for the ABC analysis step 1, Supplementary note 2.1), we designed twelve different possible

scenarios for the wild species divergence history (Figure I-S2, Tables I-S3 and I-S4). We assumed an unknown ancestral population with an effective population size of $n=50$ chosen based on preliminary ABC analyses. The 12 evolutionary scenarios included eight models without gene flow (*scA* to *scH*) and four models (*scE_GF* to *scH_GF*) with gene flow (Figure I-S2, Tables I-S3 and I-S4).

Supplementary note 3.2: ABC analysis steps 2 and 3 - Testing scenarios of apricot domestication (step 2), including bottlenecks (step 3).

To test the apricot domestication history, we introduced the cultivated clusters inferred in the STRUCTURE analyses (Figure I-S5, a total of 350 individuals clustered into seven main genetic groups for the ABC analysis steps 2 and 3, Supplementary note 2.2). In total, we tested eighteen different scenarios of apricot domestication using as backbone the most likely scenarios of wild species divergence inferred in the ABC analyse step 1 (i.e. *scF_GF*) (Figure I-S3). The domestication scenarios included nine models assuming different origins of each of the three cultivated groups (*scF_GF_1* to *scF_GF_9*, Tables I-S3 and I-S4, see results). The nine additional models had the same backbones but assumed gene flow among the South Central Asian *P. armeniaca* and the Irano-Caucasian/European *P. armeniaca* (PaEUR_C1) on the one hand and Chinese landraces (China_C4) on the other hand.

Using the ABC-RF framework, we then choose the most likely scenarios (i.e. *scF_GF_5_GF* and *scF_GF_8_GF*, see results part) from the previous step and tested the occurrence of bottleneck during domestication (Figures I-S3 and I-S10, Tables I-S3 and I-S4). For the ABC analysis step 3 (bottleneck), we tested a total of fourteen scenarios, to which we added two scenarios without bottleneck (Figure I-S3). In total we therefore tested 16 scenarios for apricot domestication (Figure I-S10, Tables I-S3 and I-S4).

Supplementary note 4: Linear Discriminant Analyses (LDA)

The results of the ABC-RF analyses are presented in Table I-S3 for the three nested ABC steps. For each ABC step, the projection of the summary statistic datasets considering all simulated scenarios on the first two LDA axes provided a visual indication that the observed summary statistics fell within the cloud of the summary statistics simulated under the various models (Figure I-S9, Step 1: "Wild divergence" and Step 2: "Domestication"), indicating that our scenarios could provide good simulations of the observed data. However, for the third step ("Bottleneck"), the simulations under different groups of scenarios assuming or not bottlenecks during domestication strongly overlapped (i.e. scenarios presenting different

numbers and strengths of bottlenecks, or no bottleneck, for the three cultivated apricot lineages), indicating a lack of power to identify the number and the strength of bottlenecks, or the respective cultivated lineages affected (Figure I-S9, Step 3: “Bottleneck”).

Supplementary note 5: Parameter estimates inferred from ABC analyses

Parameter estimates are presented in Table I-S13 for the two most likely scenarios, *scF_GF_5_GF* and *scF_GF_8_GF*, and averaged over the two scenarios hereafter (except for T_{W3-C4} that we present separately because it corresponds to the contrasting hypotheses tested by the scenarios). The ancient divergence between the red *P. armeniaca* and the dark blue *P. sibirica* lineages was estimated to have occurred 11,742,317 years ago (ya) (~12 Mya) (average estimate of divergence times across the two scenarios, T_{W1-ANC} : 12,138,226 ya [95% Confidence Interval (CI): 8,119,315-15,864,840]) and of the dark blue *P. sibirica* lineage (average estimate T_{W4-ANC} : 11,346,407 ya [CI: 8,034,770-15,804,315]). The origin of the green *P. sibirica* lineage from an admixture event between the dark blue *P. sibirica* lineage and the red *P. armeniaca* lineage was inferred to have occurred 339,055 ya (average $T_{W3-W1-W4}$, CI: 12,615-474,265]), followed by a split of the yellow wild *P. armeniaca* lineage from the red *P. sibirica* lineage 246,699 ya (average T_{W2-W1} , CI: 6,262-379,215).

According to the *scF_GF_8_GF* scenario, the brown Chinese cultivated apricot lineage derived from the green wild *P. sibirica* lineage 92,972 ya (T_{W3-C4} , CI: 2,730-281,730) (Figure I-4b, brown dotted arrow on the right). Assuming the *scF_GF_5_GF* scenario, the brown Chinese cultivated apricot lineage originated from the yellow wild *P. armeniaca* lineage 65,264 ya (T_{W2-C4} , CI: 1,770-272,190) (Figure I-4b, brown dotted arrow on the left). The domestication event yielding the light blue South Central Asian cultivated apricots from the red wild *P. armeniaca* lineage has occurred 9,543 ya (average T_{C2-W1} , CI: 420-27,810), and the domestication of the grey European cultivated apricot lineages from the yellow wild *P. armeniaca* lineage 3,981 ya (average T_{C1-W2} , CI: 195-19,440) (Figure I-4b). Migration rate estimates are given in Table I-S12. Mutation rate estimates were on average across the two scenarios of $1e^{-4}$ /bp/gen (CI: $4.74e^{-5}$ - $1.97e^{-4}$).

Chapter I supplemental Figures

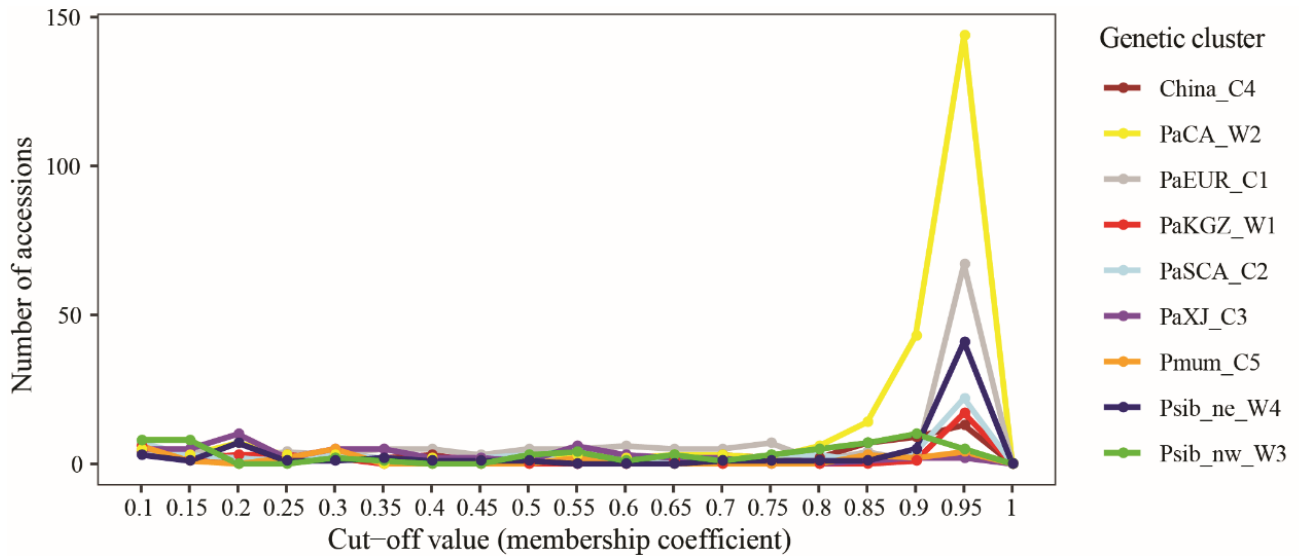


Figure I-S1. The distribution of coefficient memberships for the various genetic groups. The nine genetic clusters correspond to the ones depicted in Figures I-1 and I-S6.

Step 1 : Wild divergence

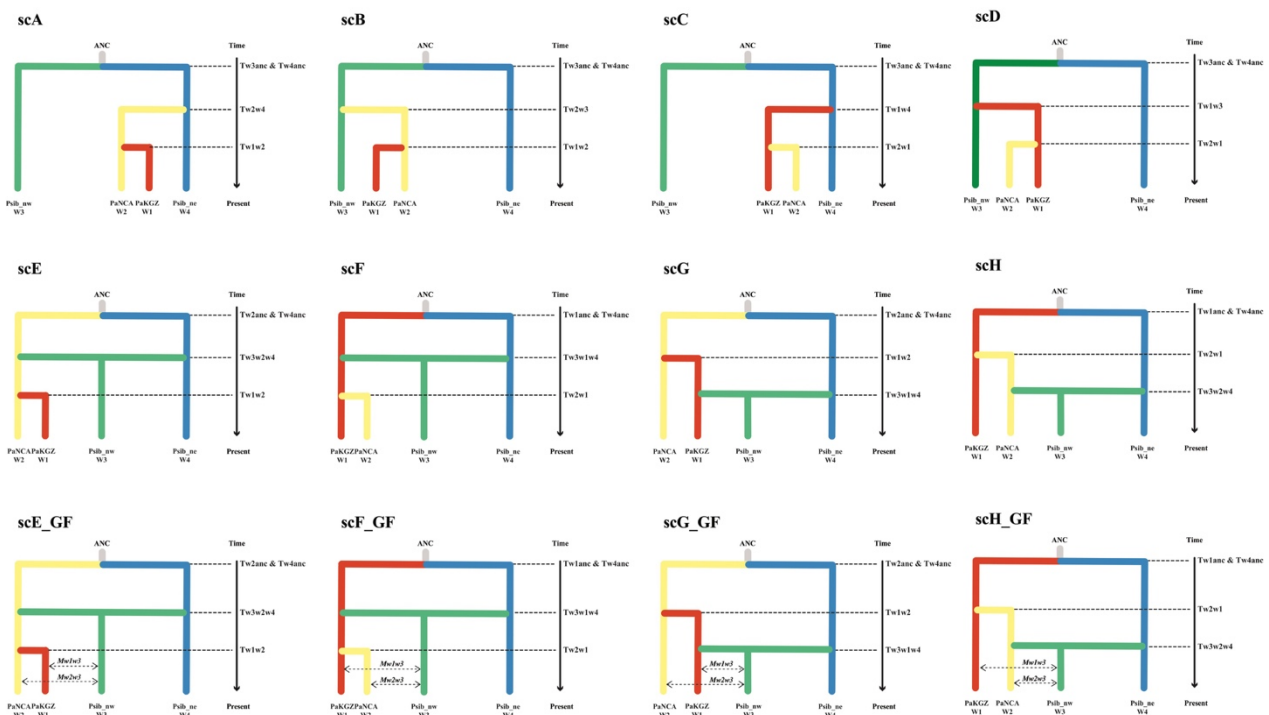


Figure I-S2. The twelve competing scenarios of divergence history among wild *Prunus* populations tested by random forest approximate Bayesian computation (ABC analysis step 1– Wild divergence). Each branch is colored according to its assigned cluster at $K=11$, and PaKGZ_W1 (red), PaNCA_W2 (yellow), Psib_nw_W3 (green), and Psib_ne_W4 (dark blue) represent the four wild clusters (Figure I-S5). The parameters $Mw1w3$ and $Mw2w3$ represent migration rates between PaKGZ_W1 and Psib_nw_W3, and between PaNCA_W2 and Psib_nw_W3, respectively. More detailed descriptions of the models are presented in Tables I-S3 and I-S4.

Step 2: Domestication

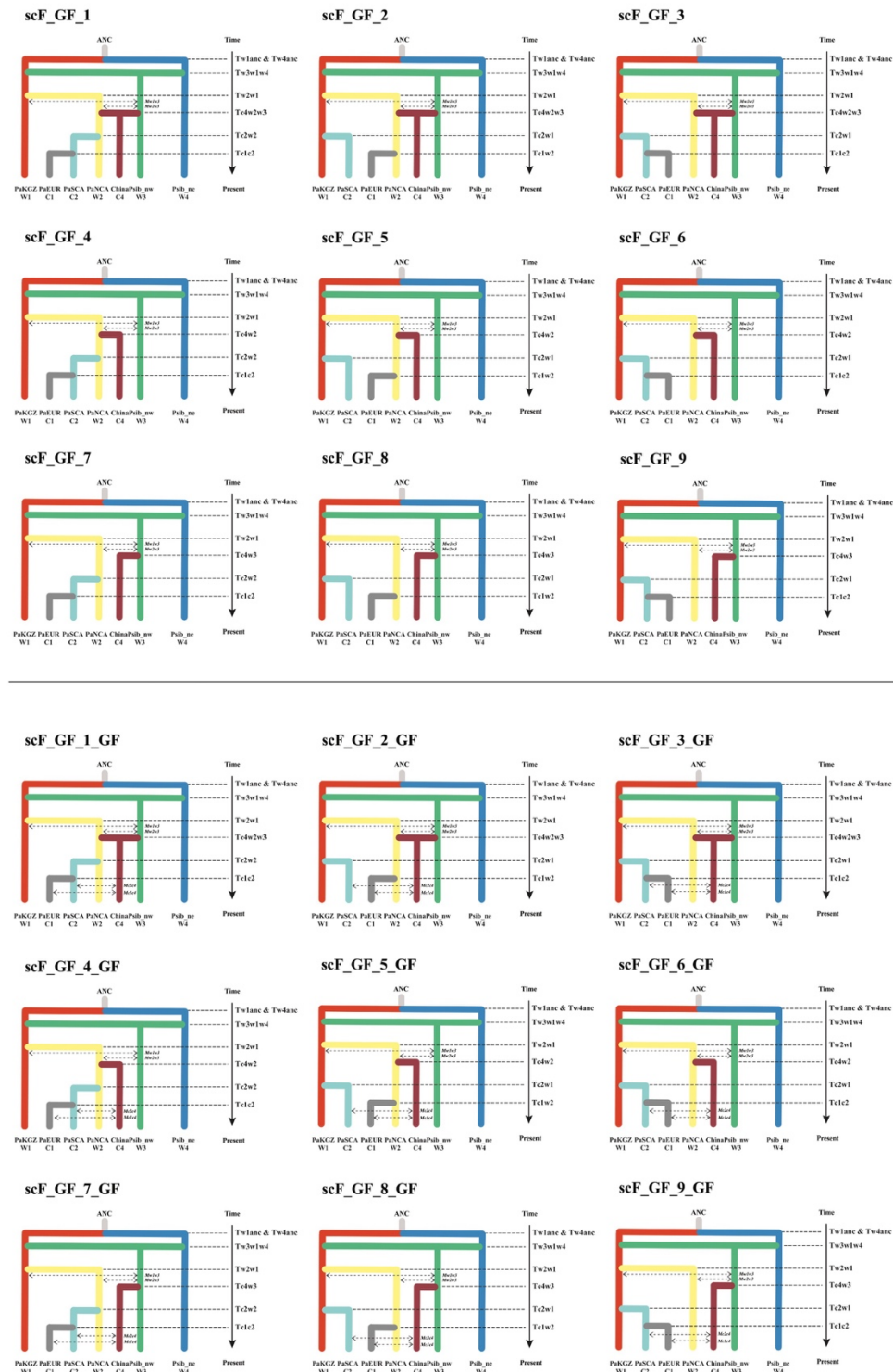


Figure I-S3. Eighteen competing scenarios of apricot domestication tested by random forest approximate Bayesian computation (ABC analysis steps 1 and 2). Each branch is colored according to its assigned cluster at $K=11$, and PaKGZ_W1 (red), PaNCA_W2 (yellow), Psib_nw_W3 (green), Psib_ne_W4 (dark blue), PaEUR_C1, *P. armeniaca* European and Irano-Caucasian cultivars; PaSCA_C2, *P. armeniaca* South Central Asian cultivated apricots; China_C4, Chinese landraces represent the four wild and three cultivated apricot clusters, respectively (Figure I-S4). We set the backbone of the eighteen tested models for scenarios of domestication based on the model chosen in the previous ABC step 1 (i.e. the scF_GF scenario in the “wild divergence” step Figure I-S1 and supplementary notes 2.2 and 3.1). *Mc2c4* and *Mc1c4* represent migration rates between PaSCA_C2 and China_C4, and between PaEUR_C1 and China_C4, respectively. More detailed descriptions of the models are presented in Tables I-S3 and I-S4. GF: gene flow.

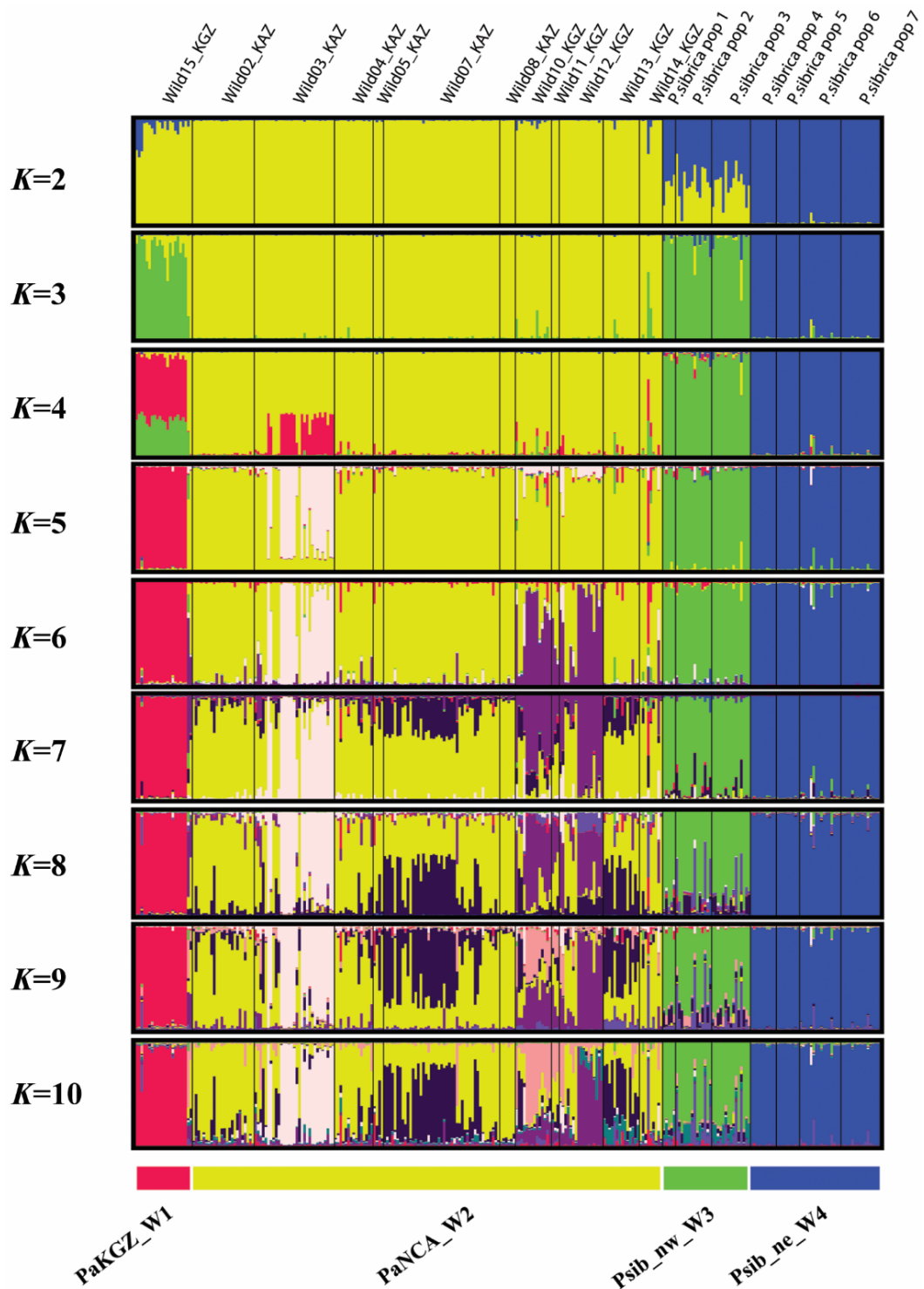


Figure I-S4. Bayesian clustering inferred with STRUCTURE for $K=2$ to $K=10$ for the 288 wild apricots ($N=288$) sampled in 19 natural populations of Central Asia and China. The sampling used to infer population structure with STRUCTURE included 204 *Prunus armeniaca* individuals from 12 sampling sites in Central Asia, and 84 *Prunus sibirica* individuals from seven natural populations in China. Each individual is represented by a vertical bar, partitioned into K segments representing the proportions of ancestry of its genome in K clusters.

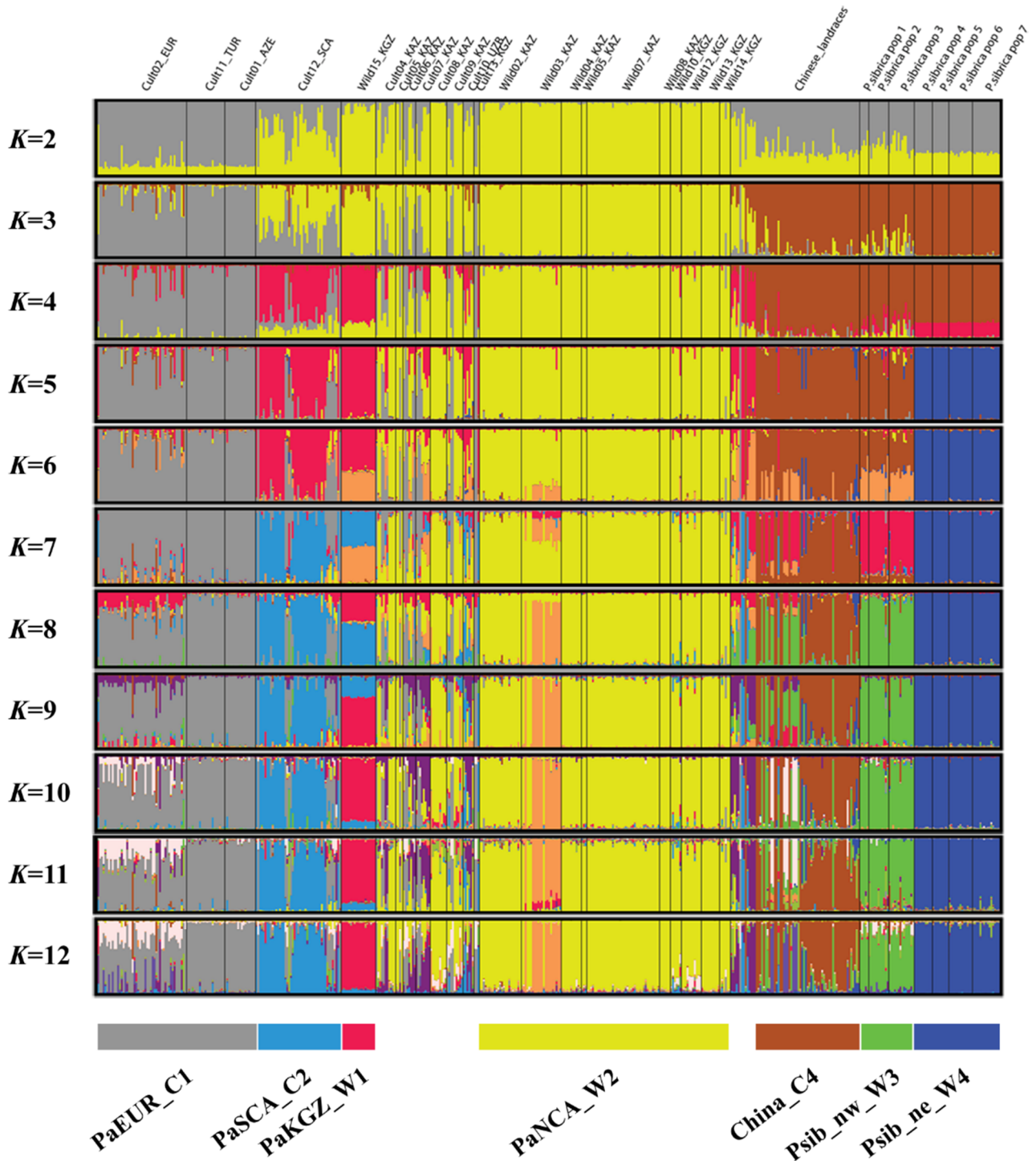


Figure I-S5. Bayesian clustering inferred with STRUCTURE for $K=2$ to $K=12$ based on the whole apricot dataset (excluding *Prunus mume*, *Prunus mandshurica* and *Prunus brigantina*, thus $n=496$) using 25 microsatellite markers. The sampling used to infer population structure, included non-wild ($N=262$) and wild individuals ($N=234$) (see details Tables I-1 and I-S5). Each individual is represented by a vertical bar, partitioned into K segments representing the proportions of ancestry of its genome in K clusters. At the bottom of the figure are shown the seven main clusters, representing a total of 350 individuals, which were assigned to each specific cluster at $K=10$ with a membership coefficient ≥ 0.85 , and then used for ABC. We colored the clusters according to the ones depicted in Figure I-1.

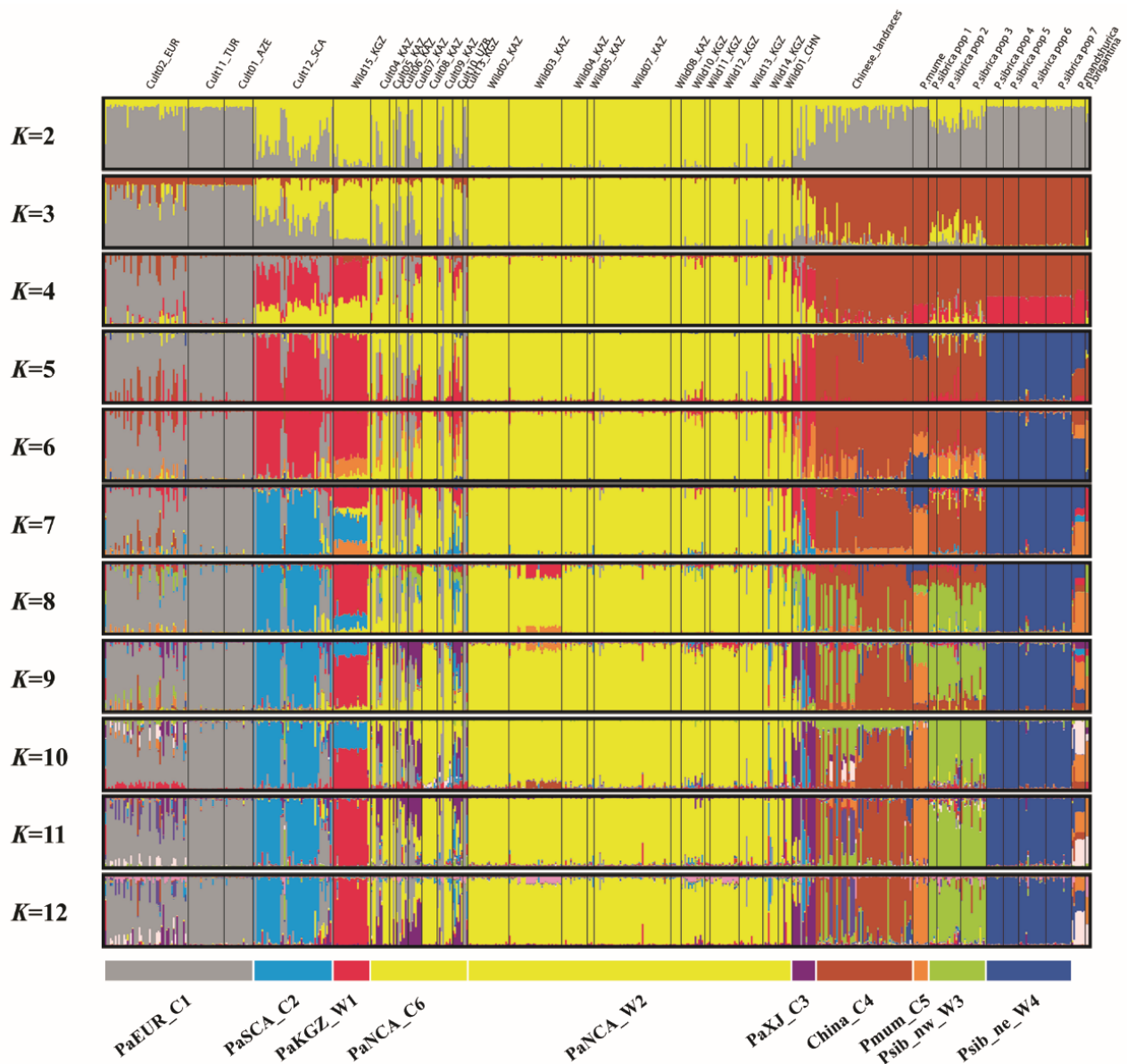


Figure I-S6. Bayesian clustering obtained with STRUCTURE for the wild and cultivated apricots using 34 microsatellite markers, from $K=2$ to $K=12$, based on the whole dataset ($n=577$). The dataset includes *Prunus armeniaca* ($N=403$), *Prunus sibirica* ($N=84$), *Prunus mume* ($N=9$), *Prunus mandshurica* ($N=8$) and *Prunus brigantina* ($N=2$). Each individual is represented by a vertical bar, partitioned into K segments representing the inferred proportions of ancestry of its genome. At the bottom of the figure are represented the ten main clusters, including a total of 416 individuals which were assigned to a cluster at $K=11$ with a membership coefficient ≥ 0.85 ; the barplot at $K=11$ corresponds to the clustering depicted in Figure I-1. The PaEUR_C1 (grey) and PaSCA_C2 (light blue) clusters corresponded to the cultivated *P. armeniaca* populations, occurring respectively in Europe and in the Irano-Caucasian region (i.e. Turkey, Azerbaijan and Iran) for the grey cluster and in South Central Asia (i.e. Pakistan, Turkmenistan, and Afghanistan) for the light blue cluster. The PaKGZ_W1 (red) cluster corresponded to a single natural *P. armeniaca* population in Sary Chelek (Kyrgyzstan) and the PaNCA_W2 and PaNCA-C6 (yellow) clusters to the wild and cultivated *P. armeniaca* populations in North Central Asia (Kazakhstan, Uzbekistan, and Kyrgyzstan), respectively. The PaXJ_C3 (purple) cluster corresponded to the cultivated *P. armeniaca* in Western China (Xinjiang) and the China_C4 (brown) cluster to other Chinese cultivars and landraces. Pmum_C5 (orange) corresponded to the ornamental species *Prunus mume*. The wild species *Prunus sibirica* clustered into Northwestern and Northeastern Chinese populations, Psib_nw_W3 (green) and Psib_ne_W4 (dark blue), respectively.

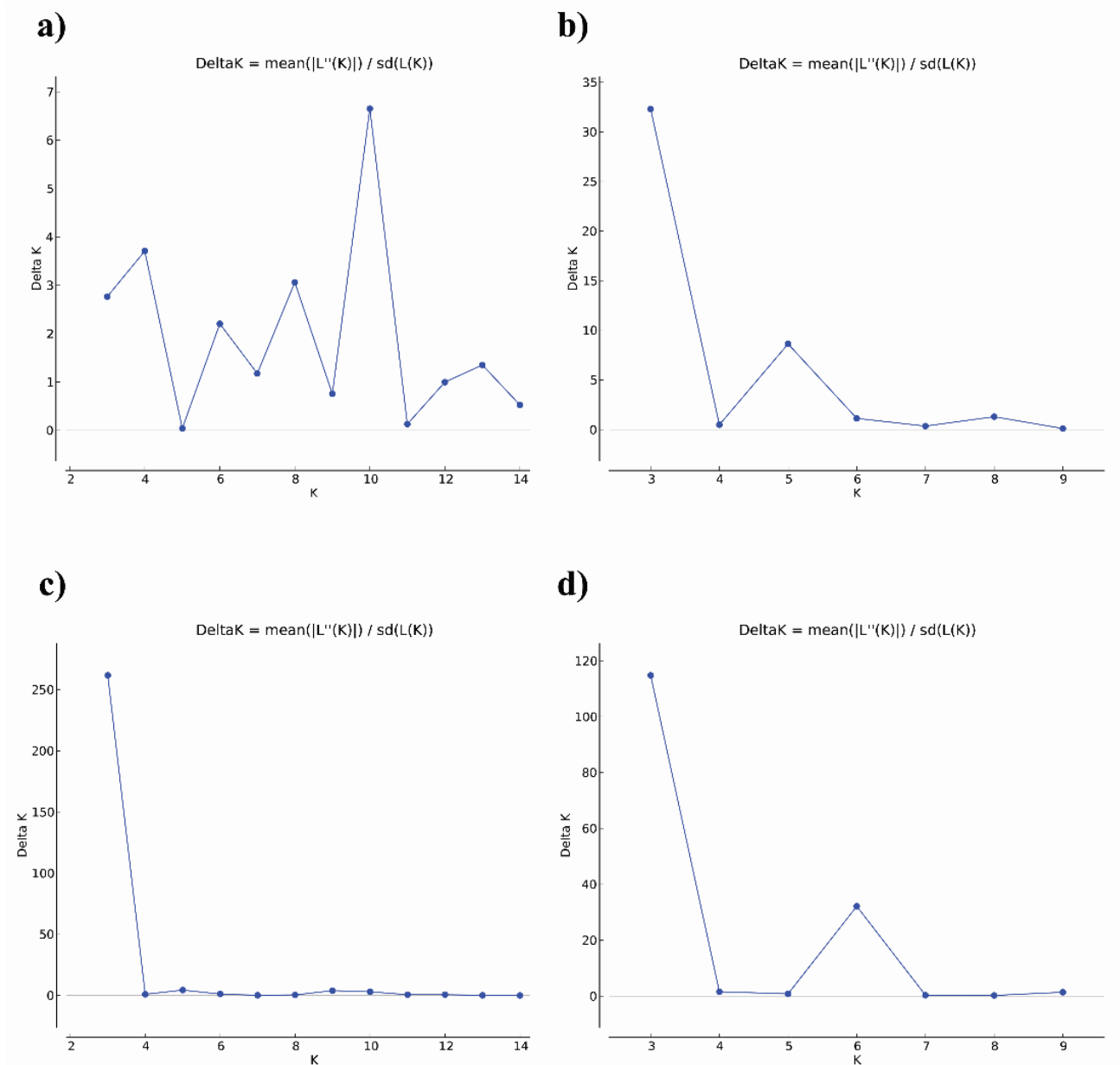


Figure I-S7. Delta K plotted against K values for each set of STRUCTURE analyses. The delta K (ΔK) was estimated for a) the entire apricot dataset of 577 individuals (STRUCTURE results based on 34 microsatellite markers corresponding to barplots in Figure I-S6), b) only the wild *Prunus sibirica* dataset (STRUCTURE results based on 25 microsatellite markers to barplots in Figure I-S8), c) the dataset of wild and cultivated 496 apricots (STRUCTURE results based on 25 microsatellite markers corresponding to barplots in Figure I-S5), and d) the sub-dataset of 288 wild apricot individuals (STRUCTURE results based on 25 microsatellite markers corresponding to barplots in Figure I-S4).

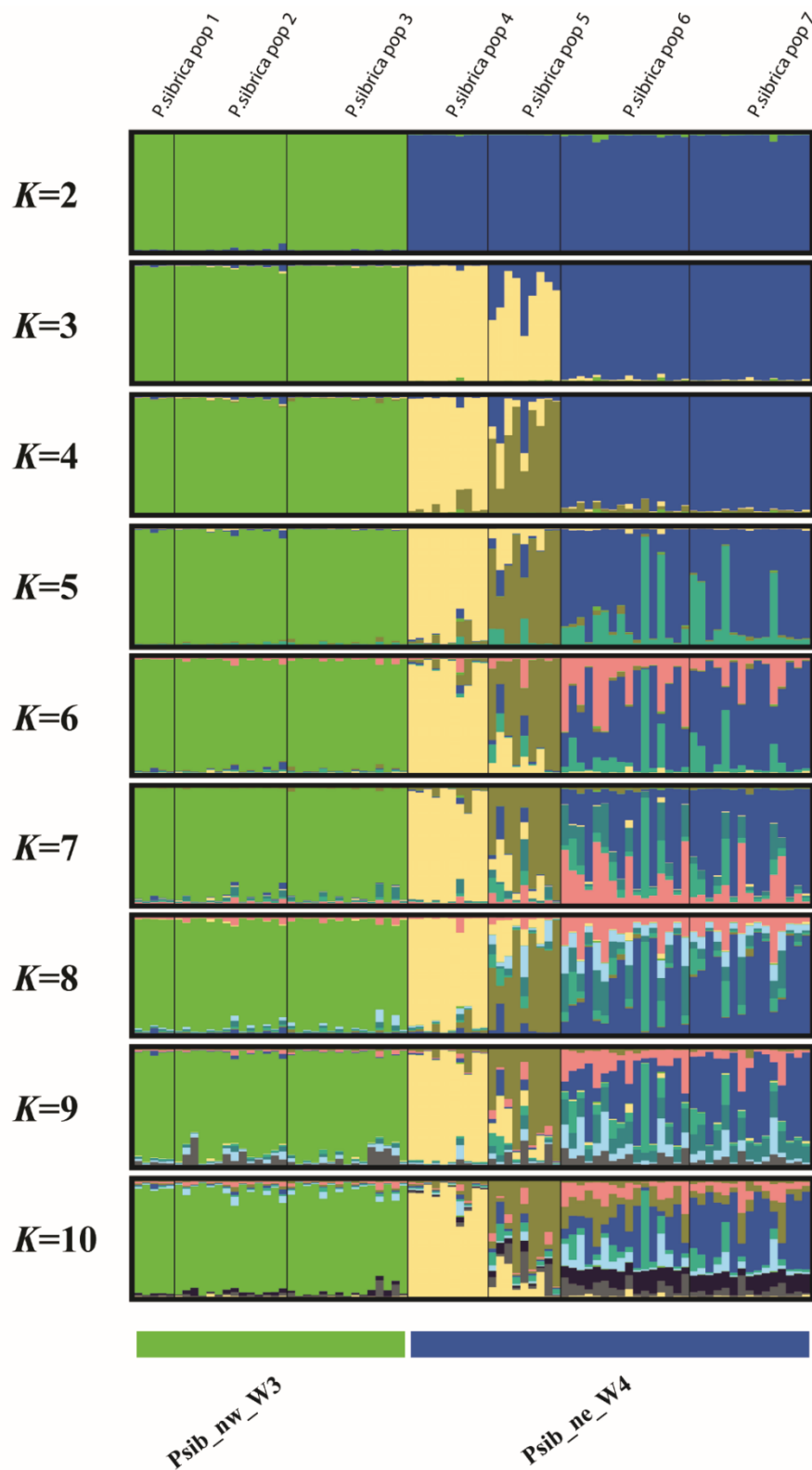


Figure I-S8. Bayesian clustering for $K=2$ to $K=10$ obtained with STRUCTURE for *Prunus sibirica* individuals ($n=84$) using 25 microsatellite markers originating from seven natural populations of Northern China. Each individual is represented by a vertical bar, partitioned into K segments representing the proportions of ancestry of its genome in K clusters. At the bottom are represented the two main *Prunus sibirica* genetic clusters at $K=2$.

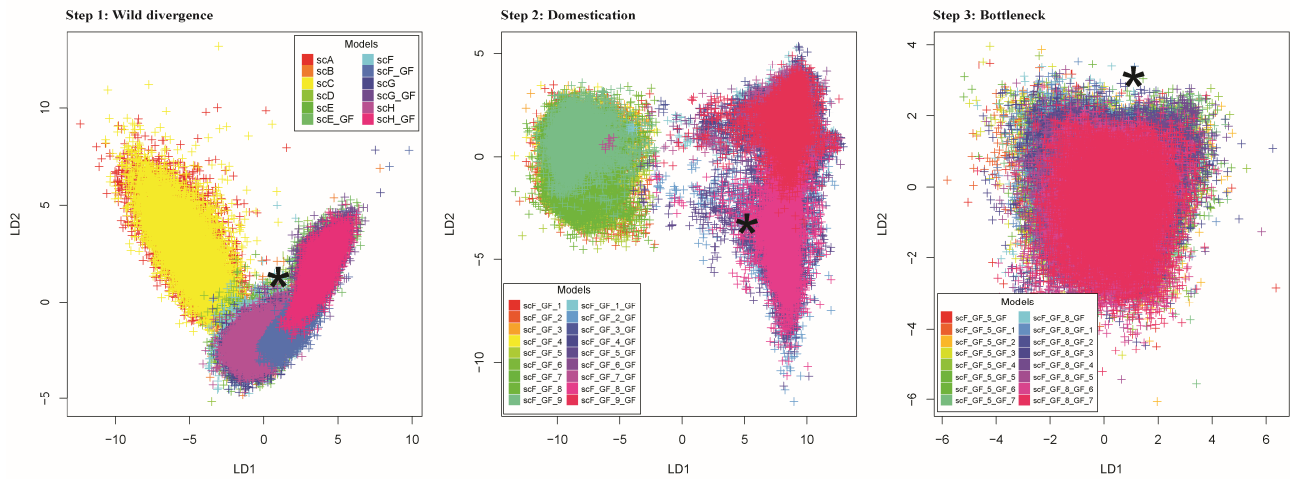


Figure I-S9. Projection of the microsatellite apricot simulated datasets on the first two linear discriminant analysis (LDA) axes to infer the apricot speciation and domestication histories for each ABC step. The location of the additional (observed) dataset is represented by a black star. Step 1: wild divergence (twelve scenarios); Step 2: domestication (eighteen scenarios); Step 3: bottleneck (sixteen scenarios). GF: gene flow.

Step 3: Bottleneck

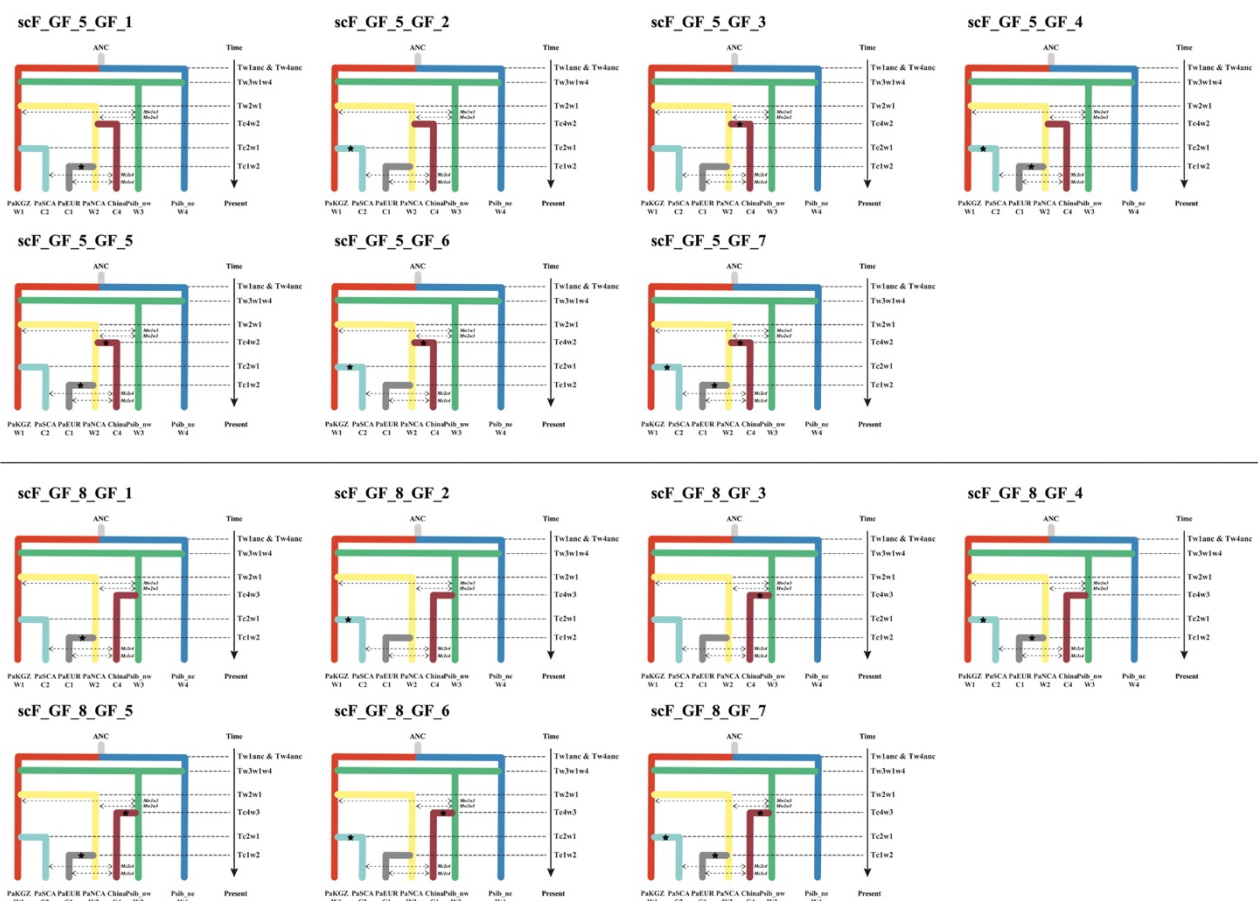


Figure I-S10. Fourteen scenarios with the occurrence of bottleneck during apricot domestication tested with approximate Bayesian computation (ABC analysis step 3). Backbones of the fourteen tested models based on the two models chosen at the previous ABC step 2 (*i.e.*, the *scF_GF_5_GF* and *scF_GF_8_GF* scenarios in the “domestication” step, Figure I-S3 and supplementary note 2.2) with the ABC analyses. The bottleneck is indicated by solid dots. Detailed descriptions of the models are presented in Tables I-S3 and I-S4. GF: gene flow.

Chapter I supplemental Tables

Table I-S1. Description of the *Prunus* cultivated and wild trees sampled for the purpose of this study, with their geographic origin and, when applicable, provider. INRA data portal: <https://doi.org/10.15454/FFKGUP>.

Table I-S2. List of microsatellite markers used in this study and their characteristics: chromosome on which is located the marker, given name of the locus, repeat motif and whether repeats were perfect, primer sequences, number of different alleles (*Na*) and number of effective alleles (*Ne*).

Chr	Locus	Perfect repeat motif	Repeat motif	Forward primer sequence	Reverse primer sequence	<i>Na</i>	<i>Ne</i>
Markers used for the analyses in this study (total: 34)		(total: 25)					
1	AMPA109		(TG)11(AG)9	GTGTCCCGAATTCCAATATCC	TTTGTCTCAACACTTTCCTCTC	6.083	3.569
1	aprigms18	x	(CT)25	TCTGAGTTCAGTGGGTAGCA	ACAGAATGTGCGTTGCTTTA	6.278	3.666
1	G22SSR	x	(TC)12	GGAGTGAGCAGTGAAGTTGTT	TTTTACACGCAGAGCTAGAATATG	2.944	2.248
1	SSRLg1_11m52a	x	(GA)22	TAGATAAGCCCACCAATTGTCA	GCATATACATCCAAAGGAAGCC	7.472	4.520
1	N86B11ssr3		(AT)7(AC)7	CCAATAACCGCTCCTCCAG	TGGATGCCTTATCCACCTGT	5.139	3.452
1	PacB26	x	(CA)19	CCAATCATGAAATCATAAAGCAA	TGGGATGTCCTATTGTTTTCA	7.167	4.416
1	pchgms03	x	(CT)14	ACGGTATGTCCGTACACTCTCCATG	CAACCTGTGATTGCTCCTATTAAC	7.500	4.766
1	PGS-1.21	x	(TC)26	CCCTGGTGTTCTGCTCTCTC	CATCCACAAATGGGAAGCAT	8.861	5.907
1	SSR04AG51		(AT)12(CT)26	TCAAGGATGGTTCCTCCAGAG	ATTGTTGGTTTGGCTATTGG	5.528	3.559
1	SSR5piso4E		(CCT)3(CTT)5	GACCCACCGTATCAAGTCAG	CATTGTTGTCTTGTGGAAACC	1.972	1.590
1	SSR5piso4Ga	x	(AG)11	GGAGGAGATTGCACGCCTAC	GGGCGTTTGGTATCGTGG	5.278	2.781
1	UDAp-414	x	(GA)22	CAAGCACAAGCGAACAAAAT	GGTGGTTTCTTATCCGATGC	7.083	4.461
1	UDP96-018	x	(AC)21	TTCTAATCTGGGCTATGGCG	GAAGTTCACATTTACGACAGGG	2.917	1.510
2	BPPCT004	x	(CT)22	CTGAGTGATCCATTTGCAGG	AGGGCATCTAGACCTCATTGTT	7.750	5.329
2	BPPCT030	x	(AG)25	AATTGTACTIONGCAATGCTATGA	CTGCCTTCTGCTCACACC	5.750	3.872
3	AMPA101		(TC)11(AC)12	CAGTTTGATTTGTGTGCCTCTC	GATCCACCCTTTCATAAAAATC	8.250	5.404
3	EPPCU0532	x	(AG)13	AAAGGGCGATGTTTCCAGAGTG	TGACGAGTTTGTGCGTTTTG	3.361	2.009
3	SSR02iso4G		(TC)18(AAGA)5	TTATGAAGGAAGGAGACAGACG	CAGAGAGAAGGGGTTGGC	5.861	3.405
4	UDA-002		(CT)17(CA)18	AAACGTGAGGTCTCACTCTCTC	GCCATTTAAGGGTCTGGTCA	8.528	5.617
4	AMPA103	x	(AG)10	GAAGGAGACGAAGCTGTGAAAG	CAACACCATCCAATAACAAGC	6.806	4.360
4	BPPCT040	x	(GA)14	ATGAGGACGTGTCTGAATGG	AGCCAAACCCCTCTTATACG	6.417	3.918
4	UDA-021		(TG)28(AC)28	GCACACGTACATTGTGACTGC	TTTGTGTAATGCCACAGATGC	7.806	5.169
4	UDAp-480	x	(GA)30	GGTTCAACCAGACCAGCATT	TGGTTTGGTAGTTGATCATTGG	8.083	5.258
5	BPPCT038	x	(GA)25	TATATTGTTGGCTTCTTGATG	TGAAAGTGAACAATGGAAGC	8.056	5.071
6	AMPA100	x	(AG)12	TGTTTAGTTGAGGGTAACCTTGG	CCCTTCCTTTCTGTGTCTCAC	6.944	4.810
6	BPPCT008	x	(GA)36	ATGGTGTGTATGGACATGATGA	CCTCAACCTAAGACACCTTCACT	6.611	4.220

Table I-S2. List of microsatellite markers used in this study and their characteristics: chromosome on which is located the marker, given name of the locus, repeat motif and whether repeats were perfect, primer sequences, number of different alleles (*Na*) and number of effective alleles (*Ne*). (Continued)

Chr	Locus	Perfect repeat motif	Repeat motif	Forward primer sequence	Reverse primer sequence	<i>Na</i>	<i>Ne</i>
Markers used for the analyses in this study (total : 34)		(total : 25)					
6	BPPCT025	x	(GA)29	TCCTGCGTAGAAGAAGGTAGC	CGACATAAAGTCCAAATGGC	4.222	2.483
6	CPPCT30	x	(CT)30	TGAATATTGTTCTCAATTC	CTCTAGGCAAGAGATGAGA	7.694	5.094
6	UDP98-412	x	(AG)28	AGGGAAAGTTTCTGCTGCAC	GCTGAAGACGACGATGATGA	8.333	5.810
7	CPPCT22		(CT)28CAA(CT)20	CAATTAGCTAGAGAGAATTATTG	GACAAGAAGCAAGTAGTTTG	9.194	6.519
7	CPPCT33	x	(CT)16	TCAGCAAAGTAGAAACAAACC	TTGCAATCTGGTTGATGTT	6.556	4.185
7	CPSCT004	x	(GA)8	GCTCTGAAGCTCTGCATTGA	TTTGGAAATGGCTATGGAGTACG	2.250	1.545
8	CPPCT06	x	(CT)16	AATTAAGTCCAACAGCTCCA	ATGGTTGCTTAATTCAATGG	8.917	5.421
8	UDP98-409	x	(AG)19	CGGACTCTTATCCTCTATCAACA	GCTGATGGGTTTTATGGTTTTTC	8.694	5.646
	Mean					6.480	4.164
Markers excluded (total: 14)							
1	Cd211SSR	-	-	CCTCTGATGTATTATCTTTCTGGC	TTTCAACAGCTCCAAATTCAC	-	-
1	PGS-1.24	-	-	GTAATGAGTGCCTGCGTGT	TGCGAGAGTTGTGATTGATG	-	-
1	ZP002	-	-	AACATTTTCTGATTCAATGCCA	TGTATCCTCCAGCTTCAAAGTC	-	-
1	ssrPaCITA05	-	-	GTTGTGTTTACTTTTTCTTAACGG	GTATCACAAGTGAGAACATAAGAGG	-	-
1	aprigms24	-	-	ATCTGCTCTTCCCTCACCT	GATTATCCCTCAACCCATCC	-	-
1	Cd195SSR	-	-	CCTCTGATGTATTATCTTTCTGGC	TTTCAACAGCTCCAAATTCAC	-	-
1	SSR12iso4G	-	-	TGACTTTTGAAGACCGGAT	TCGTGTGAGCAATCGAGG	-	-
1	aprigms10	-	-	CAATACAAAATGGCCATGC	AGCCCGTTCATTGATTTT	-	-
2	ssrPaCITA27	-	-	GATCCCTCAACTGAATCTCTC	CGTCACAACAATAGATGCCAAGG	-	-
2	BPPCT001	-	-	AATCCCAAAGGATGTGTATGAG	CAGGTGAATGAGCCAAAGC	-	-
2	AMPA116	-	-	ATTGAAGGCCCTTATGTGAG	CAAAAAGGCCGTTACAGATGATG	-	-
3	UDA-002	-	-	AAACGTGAGGTCTCACTCTCTC	GCCATTTAAGGGTCTGGTCA	-	-
3	AMPA119	-	-	GTGCCCACTTACCTGTTTTAGG	TCGACGATCAGACTTGCTACAG	-	-
5	BPPCT017	-	-	TTAAGAGTTTGTGATGGGAACC	AAGCATAATTTAGCATAACCAAGC	-	-

Footnotes: Details of microsatellites could be found in Genome Database for Rosacea (GDR, <https://www.rosaceae.org/>). The symbol X indicates the loci used in ABC analysis displaying "perfect" repeat motifs.

ANNEXE 1

Table I-S3. Summary the model choice analyses, and the corresponding results, that were carried out successively using random forest approximate Bayesian computation to reconstruct: 1) the divergence history of wild apricots, 2) the apricot domestication history, and 3) to test for the occurrence of bottleneck in cultivated groups.

ABC step	Focal populations	Group of scenario	Scenarios included in the group	Nb of tested scenarios	Hypothesis	Hypothesis on GF	Best group model	Percentage of votes*	Posterior probability	Prior error rate in %
1. Wild divergence	ANC, PaKGZ_W1, PaNCA_W2, Psib_nw_W3, Psib_ne_W4	First round ABC: Gene flow or not?	scA, scB, scC, scD	12	Divergence of PaKGZ_W1/PaNCA_W2 from either Psib_nw_W3 or Psib_ne_W4	No		11%	0.84 (+0.02)	16%
			scE, scF, scG, scH		Psib_nw_W3 resulted from an admixture between (PaNCA_W2+Psib_ne_W4) or (PaKGZ_W1+Psib_ne_W4)	No		16%		
			scE_GF, scF_GF, scG_GF, scH_GF		Psib_nw_W3 resulted from an admixture between (PaNCA_W2+Psib_ne_W4) or (PaKGZ_W1+Psib_ne_W4)	for each scenario: Psib_nw_W3 ↔ PaNCA_W2 and Psib_nw_W3 ↔ PaKGZ_W1	X	72%		
		Second round ABC: Histories of divergence?	scE_GF	4	PaNCA_W2 diverged from PaKGZ_W1, Psib_nw_W3 resulted from an admixture between PaNCA_W2+Psib_ne_W4	for each scenario: Psib_nw_W3 ↔ PaNCA_W2 and Psib_nw_W3 ↔ PaKGZ_W1		15%	0.54 (+0.02)	41%
			scF_GF		PaKGZ_W1 diverged from PaNCA_W2, Psib_nw_W3 resulted from an admixture between (PaKGZ_W1+Psib_ne_W4)	for each scenario: Psib_nw_W3 ↔ PaNCA_W2 and Psib_nw_W3 ↔ PaKGZ_W1	X	37%		
			scG_GF		Psib_ne_W4 diverged from Psib_nw_W3, Psib_nw_W3 resulted from an admixture between (PaKGZ_W1+Psib_ne_W4)	for each scenario: Psib_nw_W3 ↔ PaNCA_W2 and Psib_nw_W3 ↔ PaKGZ_W1		16%		
			scH_GF		Psib_nw_W3 diverged from W4, Psib_nw_W3 resulted from an admixture between (PaNCA_W2+Psib_ne_W4)	for each scenario: Psib_nw_W3 ↔ PaNCA_W2 and Psib_nw_W3 ↔ PaKGZ_W1		32%		
2. Domestication	ANC, PaKGZ_W1, PaNCA_W2, Psib_nw_W3, Psib_ne_W4, PaEUR_C1, PaSCA_C2, China_C4	Third round ABC analysis: Gene flow between cultivated groups?	scF_GF_1, ..., to scF_GF_9	18	NA	No		10%	0.96 (+0.008)	0.01%
			scF_GF_1_GF, ..., to scF_GF_9_GF		NA	for each scenario: PaSCA_C2 ↔ China_C4 and PaEUR_C1 ↔ China_C4	X	90%		
		Fourth round ABC analysis: Origin of PaEUR_C1 and PaSCA_C2?	scF_GF_1_GF, scF_GF_4_GF, scF_GF_7_GF	9	Domestication of PaEUR_C1 from PaSCA_C2, and of PaSCA_C2 from PaNCA_W2	for each scenario: PaSCA_C2 ↔ China_C4 and PaEUR_C1 ↔ China_C4		17%	0.72 (+0.02)	19.7%
			scF_GF_2_GF, scF_GF_5_GF, scF_GF_8_GF		Independent domestications of PaEUR_C1 from PaNCA_W2, and of PaSCA_C2 from PaKGZ_W1	for each scenario: PaSCA_C2 ↔ China_C4 and PaEUR_C1 ↔ China_C4	X	63%		
			scF_GF_3_GF, scF_GF_6_GF, scF_GF_9_GF		Domestication of PaEUR_C1 from PaSCA_C2, and of PaSCA_C2 from PaKGZ_W1	for each scenario: PaSCA_C2 ↔ China_C4 and PaEUR_C1 ↔ China_C4		20%		
		Fifth round ABC analysis: Origin of China_C4 from PaNCA_W2 or Psib_nw_W3 or (PaNCA_W2+Psib_nw_W3)?	scF_GF_2_GF	3	China_C4 diverged from an admixture between (PaNCA_W2+Psib_nw_W3)	for each scenario: PaSCA_C2 ↔ China_C4 and PaEUR_C1 ↔ China_C4		25%	0.55 (+0.04)	38.3%
			scF_GF_5_GF		China_C4 diverged from PaNCA_W2	for each scenario: PaSCA_C2 ↔ China_C4 and PaEUR_C1 ↔ China_C4	X?	37%		
		Sixth round ABC analysis: Origin of China_C4 from PaNCA_W2 or Psib_nw_W3?	scF_GF_8_GF	2	China_C4 diverged from Psib_nw_W3	for each scenario: PaSCA_C2 ↔ China_C4 and PaEUR_C1 ↔ China_C4		38%	0.53 (+0.04)	30.5%
			scF_GF_5_GF		China_C4 diverged from PaNCA_W2	for each scenario: PaSCA_C2 ↔ China_C4 and PaEUR_C1 ↔ China_C4	X?	49%		
					scF_GF_8_GF		China_C4 diverged from Psib_nw_W3	for each scenario: PaSCA_C2 ↔ China_C4 and PaEUR_C1 ↔ China_C4		51%
3. Bottleneck		Seventh round ABC analyses: Bottleneck or not?	scF_GF_5_GF, scF_GF_8_GF	16	No bottleneck	for each scenario: PaSCA_C2 ↔ China_C4 and PaEUR_C1 ↔ China_C4	X?	50%	0.47 (+0.01)	12.21%
			scF_GF_5_GF_2, ... to scF_GF_5_GF_7,		Bottleneck	for each scenario: PaSCA_C2 ↔ China_C4 and PaEUR_C1 ↔ China_C4	X?	50%		
			scF_GF_8_GF_2, ... to scF_GF_8_GF_7							

Footnotes: *over the 10 replicates; ANC: Ancestral unknown population; GF: gene flow, all genetic cluster abbreviations are explained in the main text and in Figure I-1.

Table I-S4. Description of each scenario and hypothesis tested with approximate Bayesian computation to reconstruct the wild divergence and domestication histories of apricots.

ABC step	Focal genetic clusters	Scenario name	Tested hypotheses		
Step 1: Wild divergence	PaKGZ_W1 PaNCA_W2 Psib_nw_W3 Pasib_ne_W4	scA	Single introduction from Northeast Chinese <i>P. sibirica</i> to Central Asian <i>P. armeniaca</i> , and single migration from north Central Asian <i>P. armeniaca</i> to Kyrgyzstan <i>P. armeniaca</i> .		
		scB	Single introduction from Northwest Chinese <i>P. sibirica</i> to Central Asian <i>P. armeniaca</i> , and single migration from north Central Asian <i>P. armeniaca</i> to Kyrgyzstan <i>P. armeniaca</i> .		
		scC	Single introduction from Northeast Chinese <i>P. sibirica</i> to Central Asian <i>P. armeniaca</i> , and single migration from Kyrgyzstan <i>P. armeniaca</i> to north Central Asian <i>P. armeniaca</i> .		
		scD	Single introduction from Northwest Chinese <i>P. sibirica</i> to Central Asian <i>P. armeniaca</i> , and single migration from Kyrgyzstan <i>P. armeniaca</i> to north Central Asian <i>P. armeniaca</i> .		
		scE	Admixture event between north Central Asian <i>P. armeniaca</i> and Northeast Chinese <i>P. sibirica</i> , and single migration from north Central Asian <i>P. armeniaca</i> to Kyrgyzstan <i>P. armeniaca</i> .		
		scF	Admixture event between Kyrgyzstan <i>P. armeniaca</i> and Northeast Chinese <i>P. sibirica</i> , and single migration from Kyrgyzstan <i>P. armeniaca</i> to north Central Asian <i>P. armeniaca</i> .		
		scG	Single migration from north Central Asian <i>P. armeniaca</i> to Kyrgyzstan <i>P. armeniaca</i> , and admixture event between Kyrgyzstan <i>P. armeniaca</i> and Northeast Chinese <i>P. sibirica</i> .		
		scH	Single migration from Kyrgyzstan <i>P. armeniaca</i> to north Central Asian <i>P. armeniaca</i> , and admixture event between north Central Asian <i>P. armeniaca</i> and Northeast Chinese <i>P. sibirica</i> .		
		scE_GF	scE + gene flow between Northwest Chinese <i>P. sibirica</i> , and north Central Asian <i>P. armeniaca</i> and Kyrgyzstan <i>P. armeniaca</i> .		
		scF_GF	scF + gene flow between Northwest Chinese <i>P. sibirica</i> , and north Central Asian <i>P. armeniaca</i> and Kyrgyzstan <i>P. armeniaca</i> .		
		scG_GF	scG + gene flow between Northwest Chinese <i>P. sibirica</i> , and north Central Asian <i>P. armeniaca</i> and Kyrgyzstan <i>P. armeniaca</i> .		
		scH_GF	scH + gene flow between Northwest Chinese <i>P. sibirica</i> , and north Central Asian <i>P. armeniaca</i> and Kyrgyzstan <i>P. armeniaca</i> .		
		Step 2: Domestication	PaEUR_C1 PaSCA_C2 China_C4	scF_GF_1	Admixture event between north Central Asian <i>P. armeniaca</i> and Northwest Chinese <i>P. sibirica</i> , and successive <i>P. armeniaca</i> domestication events from north to south Central Asia and later to Irano-Caucasian/European regions.
				scF_GF_2	Admixture event between north Central Asian <i>P. armeniaca</i> and Northwest Chinese <i>P. sibirica</i> , and two independent <i>P. armeniaca</i> domestication events from north Central Asia to Irano-Caucasian/European regions, and from Kyrgyzstan to south Central Asia.
scF_GF_3	Admixture event between north Central Asian <i>P. armeniaca</i> and Northwest Chinese <i>P. sibirica</i> , and successive <i>P. armeniaca</i> domestication events from Kyrgyzstan to south Central Asia and then to Irano-Caucasian/European regions.				
scF_GF_4	Single independent <i>P. armeniaca</i> domestication event from north Central Asia to China, and successive <i>P. armeniaca</i> domestication events from north to south Central Asia and later to Irano-Caucasian/European regions.				
scF_GF_5	Single independent <i>P. armeniaca</i> domestication event from north Central Asia to China, and two independent <i>P. armeniaca</i> domestication events from north Central Asia to Irano-Caucasian/European regions, and from Kyrgyzstan to south Central Asia.				
scF_GF_6	Single independent <i>P. armeniaca</i> domestication event from north Central Asia to China, and successive <i>P. armeniaca</i> domestication events from Kyrgyzstan to south Central Asia and later to Irano-Caucasian/European regions.				
scF_GF_7	Single <i>P. sibirica</i> domestication event in Northwest China, and successive <i>P. armeniaca</i> domestication events from north to south Central Asia and later to Irano-Caucasian/European regions.				
scF_GF_8	Single <i>P. sibirica</i> domestication event in Northwest China, and two independent <i>P. armeniaca</i> domestication events from north Central Asia to Irano-Caucasian/European regions, and from Kyrgyzstan to south Central Asia.				
scF_GF_9	Single <i>P. sibirica</i> domestication event in Northwest China, and successive <i>P. armeniaca</i> domestication events from Kyrgyzstan to south Central Asia and later to Irano-Caucasian/European regions.				
scF_GF_1_GF	scF_GF_1 with bidirectional gene flow between Chinese landraces, and south Central Asian <i>P. armeniaca</i> and Irano-Caucasian/European <i>P. armeniaca</i> .				
scF_GF_2_GF	scF_GF_2 with bidirectional gene flow between Chinese landraces, and south Central Asian <i>P. armeniaca</i> and Irano-Caucasian/European <i>P. armeniaca</i> .				
scF_GF_3_GF	scF_GF_3 with bidirectional gene flow between Chinese landraces, and south Central Asian <i>P. armeniaca</i> and Irano-Caucasian/European <i>P. armeniaca</i> .				
scF_GF_4_GF	scF_GF_4 with bidirectional gene flow between Chinese landraces, and south Central Asian <i>P. armeniaca</i> and Irano-Caucasian/European <i>P. armeniaca</i> .				
scF_GF_5_GF	scF_GF_5 with bidirectional gene flow between Chinese landraces, and south Central Asian <i>P. armeniaca</i> and Irano-Caucasian/European <i>P. armeniaca</i> .				
scF_GF_6_GF	scF_GF_6 with bidirectional gene flow between Chinese landraces, and south Central Asian <i>P. armeniaca</i> and Irano-Caucasian/European <i>P. armeniaca</i> .				
scF_GF_7_GF	scF_GF_7 with bidirectional gene flow between Chinese landraces, and south Central Asian <i>P. armeniaca</i> and Irano-Caucasian/European <i>P. armeniaca</i> .				
scF_GF_8_GF	scF_GF_8 with bidirectional gene flow between Chinese landraces, and south Central Asian <i>P. armeniaca</i> and Irano-Caucasian/European <i>P. armeniaca</i> .				
scF_GF_9_GF	scF_GF_9 with bidirectional gene flow between Chinese landraces, and south Central Asian <i>P. armeniaca</i> and Irano-Caucasian/European <i>P. armeniaca</i> .				
Step 3: Bottleneck	PaEUR_C1 PaSCA_C2 China_C4			scF_GF_5_GF_1	scF_GF_5_GF with a single bottleneck event on Irano-Caucasian/European <i>P. armeniaca</i> .
				scF_GF_5_GF_2	scF_GF_5_GF with a single bottleneck event on South Central Asian <i>P. armeniaca</i> .
				scF_GF_5_GF_3	scF_GF_5_GF with a single bottleneck event on Chinese landraces.
		scF_GF_5_GF_4	scF_GF_5_GF with two bottleneck events on both south Central Asian <i>P. armeniaca</i> and Irano-Caucasian/European <i>P. armeniaca</i> .		
		scF_GF_5_GF_5	scF_GF_5_GF with two bottleneck events on both Chinese landraces and Irano-Caucasian/European <i>P. armeniaca</i> .		
		scF_GF_5_GF_6	scF_GF_5_GF with two bottleneck events on both south Central Asian <i>P. armeniaca</i> and Chinese landraces.		
		scF_GF_5_GF_7	scF_GF_5_GF with three bottleneck events on Irano-Caucasian/European <i>P. armeniaca</i> , south Central Asian <i>P. armeniaca</i> and Chinese landraces.		
		scF_GF_8_GF_1	scF_GF_8_GF with single bottleneck event on Irano-Caucasian/European <i>P. armeniaca</i> .		
		scF_GF_8_GF_2	scF_GF_8_GF with single bottleneck event on south Central Asian <i>P. armeniaca</i> .		
		scF_GF_8_GF_3	scF_GF_8_GF with single bottleneck event on Chinese landraces.		
		scF_GF_8_GF_4	scF_GF_8_GF with two bottleneck events on both South Central Asian <i>P. armeniaca</i> and Irano-Caucasian/European <i>P. armeniaca</i> .		
		scF_GF_8_GF_5	scF_GF_8_GF with two bottleneck events on both Chinese landraces and Irano-Caucasian/European <i>P. armeniaca</i> .		
		scF_GF_8_GF_6	scF_GF_8_GF with two bottleneck events on both south Central Asian <i>P. armeniaca</i> and Chinese landraces.		
		scF_GF_8_GF_7	scF_GF_8_GF with three bottleneck events on Irano-Caucasian/European <i>P. armeniaca</i> , south Central Asian <i>P. armeniaca</i> and Chinese landraces.		

Footnotes: GF: gene flow

ANNEXE 1

Table I-S5. Genetic variation within each of the four clusters inferred with STRUCTURE at $K=6$ for the wild apricots ($n=234$) based on 25 microsatellite markers.

Cluster	N	H_o	H_E	A_R	A_P
PaKGZ_W1	19	0.671 (0.212)	0.694 (0.192)	5.186	0.381
PaNCA_W2	138	0.651 (0.167)	0.705 (0.189)	5.983	0.749
Psib_nw_W3	30	0.720 (0.177)	0.794 (0.164)	8.006	1.697
Psib_ne_W4	47	0.641 (0.271)	0.730 (0.278)	7.877	2.673

Footnotes: N : number of individuals, H_E and H_o : expected and observed heterozygosities, A_r and A_p : allelic richness and private allelic richness. In brackets, standard deviations.

Table I-S6. Prior distributions used for approximate Bayesian computations for inferring the wild and cultivated apricot divergence histories.

	Parameter	Lower bound	Upper bound
<i>ABC analysis 1 (wild only)</i>	μ	10 ⁻⁵	10 ⁻³
	N_X	10	10,000
	T_{Wx-ANC}	8,000,000	16,000,000
	T_{Wx-Wy}	1.000	300.000
	m_{X-Y}	0.005	0.05
<i>ABC analyses 2 and 3 (wild and cultivated)</i>	$T_{Wx-Culty}$	1	30000
	μ	10 ⁻⁵	10 ⁻³
	N_X	10	10,000
	T_{Wx-ANC}	8,000,000	16,000,000
	T_{Wx-Wy}	1.000	300.000
	m_{X-Y}	0.005	0.05

Footnotes: Prior distributions are uniform and log uniform between lower and upper bound. Parameters are introduced in Figures I-S1 and 3. N_x : effective population size of group x; T_{Wx-ANC} : Divergence time between the wild group x and the ancestral unknown population; T_{Wx-Wy} : Divergence time between the wild group x and the wild group y; m_{X-Y} : migration rates between groups x and y; $T_{Wx-Culty}$: Divergence time between the wild group x and the cultivated group y; μ : mutation rate.

Table I-S7. Numbers of wild and cultivated apricots assigned to each, or admixed between, the clusters inferred with STRUCTURE at K=11 (n=577 individuals).

Cultivated or wild groups	N	Genetic clusters inferred with STRUCTURE										Total	% of admixed	Site_ID		
		PaEUR_	PaSCA_	PaXJ_	China_	Pmum_	PaKGZ_	PaNCA_	Psib_nw_	Psib_ne_	other					
		C1	C2	C3	C4	C5	W1	W2	W3	W4						
European/Irano-Caucasian cultivated apricots	87	79 (22)		1 (1)	2 (2)		2 (2)						3 (3)	87 (30)	34.50%	Cult02_EUR, Cult11_TUR, Cult01_AZE
South Central Asian cultivated apricots	47	8 (8)	37 (9)	1 (1)					1 (1)					47 (19)	40.40%	Cult12_SCA
Central Asian wild apricots (Sary Chelek)	22						21 (2)	1 (1)						22 (2)	9.10%	Wild15_KGZ
North Central Asian cultivated apricots	57	14 (8)	3 (3)	8 (8)				32 (6)						57 (25)	43.90%	Cult(04, 05, 06, 07, 08, 09)_KAZ, Cult10_UZB, Cult13_KGZ
North Central Asian wild apricots	190		1 (0)	1 (0)				188 (19)						190 (19)	10%	Wild(02, 03, 04, 05, 07, 08)_KAZ, Wild(08, 10, 11, 12, 13, 14)_KGZ, Wild01_CHN
Chinese cultivated apricots (Xinjiang)	14	1 (0)	2 (2)	10 (6)	1 (0)									14 (8)	57.10%	ch_113, ch_114, ch_117, ch_119, ch_120, ch_121, ch_136, ch_137, CH211, CH213, CH214, CH215, CH216, CH217
Chinese apricot landraces	57		1 (1)		39 (12)				8 (5)	1 (1)			8 (8)	57 (27)	47.40%	Chinese_landraces
<i>Prunus mume</i>	9					9 (1)								9 (1)	11.10%	<i>Prunus mume</i>
<i>Prunus sibirica</i> from northwest China	34								34 (15)					34 (15)	44.10%	<i>Prunus sibirica</i> pop 1, 2, 3
<i>Prunus sibirica</i> from northeast China	50									50 (3)				50 (3)	6%	<i>Prunus sibirica</i> pop 4, 5, 6, 7
<i>Prunus mandshurica</i>	8					2 (2)								2 (2)	100%	<i>Prunus mandshurica</i>
<i>Prunus brigantina</i>	2									2 (2)		6 (6)	8 (8)	100%	<i>Prunus brigantina</i>	

Footnotes: Numbers in brackets indicate the number of accessions not assigned to given genetic cluster with membership > 85% threshold. N=number of individuals. Site_ID: identification name of the site included in each eco-geographical group.

ANNEXE 1

Table I-S8. Percentage and number of fully assigned (i.e. individuals that were assigned at $\geq 85\%$ to a given cluster), or admixed, wild and cultivated apricot individuals to each cluster defined with STRUCTURE for $2 \leq K \leq 12$.

<i>Bayesian clustering of 577 cultivated and wild Armeniaca individuals based on 34 SSR loci</i>											
	K=2	K=3	K=4	K=5	K=6	K=7	K=8	K=9	K=10	K=11	K=12
Percentage of assigned individuals	74.2%	70.2%	57.8%	78.4%	67.1%	69.2%	60.9%	64.0%	66.4%	72.1%	70.2%
Percentage of admixed individuals	25.8%	29.8%	42.2%	21.6%	32.9%	30.8%	39.1%	36.0%	33.6%	27.9%	29.8%

<i>Bayesian clustering of 84 P. sibirica individuals based on 34 SSR loci</i>									
	K=2	K=3	K=4	K=5	K=6	K=7	K=8	K=9	K=10
Percentage of assigned individuals	100.0%	92.9%	91.7%	77.4%	69.0%	58.3%	51.2%	36.9%	40.5%
Percentage of admixed individuals	0.0%	7.1%	8.3%	22.6%	31.0%	41.7%	48.8%	63.1%	59.5%

<i>Bayesian clustering of 496 cultivated and wild Armeniaca individuals, based on 25 SSR loci displaying perfect repeat motifs</i>											
	K=2	K=3	K=4	K=5	K=6	K=7	K=8	K=9	K=10	K=11	K=12
Percentage of assigned individuals	54.03%	75.00%	59.88%	84.07%	66.53%	66.33%	62.30%	65.32%	70.56%	58.47%	57.06%
Percentage of admixed individuals	46.0%	25.0%	40.1%	15.9%	33.5%	33.7%	37.7%	34.7%	29.4%	41.5%	42.9%

<i>Bayesian clustering of 288 wild Armeniaca individuals, based on 25 SSR loci displaying perfect repeat motifs</i>									
	K=2	K=3	K=4	K=5	K=6	K=7	K=8	K=9	K=10
Percentage of assigned individuals	87.5%	94.1%	81.3%	88.6%	81.3%	59.4%	50.0%	46.6%	42.1%
Percentage of admixed individuals	12.5%	5.9%	18.7%	11.4%	18.7%	40.6%	50.0%	53.4%	57.9%

Footnotes: Each value gives the percentage of individuals that had a proportion ≥ 0.85 of membership assignment in a given cluster detected with the different sets of STRUCTURE analyses (supplementary note 2).

Table I-S9. Summary of the Wilcoxon signed rank tests to investigate differences in genetic diversity between cultivated and wild genetic clusters presented in Table I-2. Significant value ($FDR_Q < 0.05$) are highlighted in light blue.

Genetic cluster	Value 2	Compare with :															
		PaKGZ_W1				PaNCA_W2				Psib_nw_W3				Psib_ne_W4			
Value 1		H_o	H_E	A_R	A_P	H_o	H_E	A_R	A_P	H_o	H_E	A_R	A_P	H_o	H_E	A_R	A_P
PaEUR_C1	value 1	0.569	0.639	2.713	0.362	0.569	0.639	2.713	0.362	0.569	0.639	2.713	0.362	0.569	0.639	2.713	0.362
	value1 std	0.199	0.214	0.776	0.420	0.199	0.214	0.776	0.420	0.199	0.214	0.776	0.420	0.199	0.214	0.776	0.420
	value 2	0.652	0.695	2.984	0.329	0.663	0.724	3.160	0.385	0.706	0.783	3.510	0.610	0.629	0.738	3.371	1.044
	value2 std	0.209	0.167	0.655	0.355	0.163	0.170	0.701	0.294	0.200	0.198	0.815	0.476	0.252	0.262	1.016	0.686
	p-value	0.010	0.064	0.038	0.849	0.001	0.003	0.001	0.110	0.000	0.000	0.000	0.000	0.012	0.001	0.000	0.000
	FDR_Q	0.094	0.145	0.019	0.899	0.023	0.011	0.000	0.158	0.004	0.000	0.000	0.002	0.075	0.007	0.000	0.000
PaSCA_C2	value 1	0.645	0.685	2.947	0.298	0.645	0.685	2.947	0.298	0.645	0.685	2.947	0.298	0.645	0.685	2.947	0.298
	value1 std	0.189	0.171	0.663	0.281	0.189	0.171	0.663	0.281	0.189	0.171	0.663	0.281	0.189	0.171	0.663	0.281
	value 2	0.652	0.695	2.984	0.329	0.663	0.724	3.160	0.385	0.706	0.783	3.510	0.610	0.629	0.738	3.371	1.044
	value2 std	0.209	0.167	0.655	0.355	0.163	0.170	0.701	0.294	0.200	0.198	0.815	0.476	0.252	0.262	1.013	0.686
	p-value	0.912	0.704	0.660	0.881	0.653	0.049	0.032	0.157	0.042	0.000	0.000	0.001	0.936	0.020	0.004	0.000
	FDR_Q	0.966	0.905	0.550	0.881	0.940	0.117	0.014	0.215	0.127	0.001	0.000	0.002	0.963	0.065	0.001	0.000
PaXJ_C3	value 1	0.679	0.681	2.864	0.325	0.679	0.681	2.864	0.325	0.679	0.681	2.864	0.325	0.679	0.681	2.864	0.325
	value1 std	0.256	0.168	0.063	0.327	0.256	0.168	0.063	0.327	0.256	0.168	0.063	0.327	0.256	0.168	0.063	0.327
	value 2	0.652	0.695	2.984	0.329	0.663	0.724	3.160	0.385	0.706	0.783	3.510	0.610	0.629	0.738	3.371	1.044
	value2 std	0.209	0.167	0.655	0.355	0.163	0.170	0.701	0.294	0.200	0.198	0.815	0.476	0.252	0.262	1.013	0.686
	p-value	0.889	0.180	0.177	0.865	0.818	0.037	0.020	0.384	0.509	0.000	0.000	0.003	0.617	0.027	0.005	0.000
	FDR_Q	0.969	0.282	0.118	0.890	0.982	0.101	0.008	0.494	0.833	0.001	0.000	0.006	0.926	0.081	0.002	0.000
China_C4	value 1	0.675	0.694	3.010	0.537	0.675	0.694	3.010	0.537	0.675	0.694	3.010	0.537	0.675	0.694	3.010	0.537
	value1 std	0.214	0.200	0.819	0.342	0.214	0.200	0.819	0.342	0.214	0.200	0.819	0.342	0.214	0.200	0.819	0.342
	value 2	0.652	0.695	2.984	0.329	0.663	0.724	3.160	0.385	0.706	0.783	3.510	0.610	0.629	0.738	3.371	1.044
	value2 std	0.209	0.167	0.655	0.355	0.163	0.170	0.701	0.294	0.200	0.198	0.815	0.476	0.252	0.262	1.013	0.686
	p-value	0.726	0.873	0.741	0.004	0.952	0.168	0.223	0.021	0.162	0.000	0.000	0.516	0.459	0.015	0.014	0.000
	FDR_Q	0.902	0.924	0.700	0.007	0.952	0.302	0.154	0.033	0.342	0.000	0.000	0.599	0.827	0.053	0.005	0.000
Pmum_C5	value 1	0.568	0.720	3.097	1.222	0.568	0.720	3.097	1.222	0.568	0.720	3.097	1.222	0.568	0.720	3.097	1.222
	value1 std	0.229	0.186	0.865	0.877	0.229	0.186	0.865	0.877	0.229	0.186	0.865	0.877	0.229	0.186	0.865	0.877
	value 2	0.652	0.695	2.984	0.329	0.663	0.724	3.160	0.385	0.706	0.783	3.510	0.610	0.629	0.738	3.371	1.044
	value2 std	0.209	0.167	0.655	0.355	0.163	0.170	0.701	0.294	0.200	0.198	0.815	0.476	0.252	0.262	1.013	0.686
	p-value	0.049	0.841	0.503	0.000	0.014	0.689	0.865	0.000	0.004	0.004	0.004	0.001	0.082	0.197	0.142	0.424
	FDR_Q	0.135	0.918	0.391	0.000	0.073	0.919	0.841	0.000	0.051	0.015	0.001	0.002	0.196	0.296	0.087	0.508

Footnotes: H_E and H_o : expected and observed heterozygosities, A_R : allelic richness, value1: genetic diversity estimated for the genetic cluster presented in the column; value 2: genetic diversity estimated for the genetic cluster presented in the column; std standard deviation; a significance level of 0.05 was chosen for the two-tailed test; FDR_Q: adjusted P-value calculated using the Benjamini-Hochberg method.

ANNEXE 1

Table I-S10. Summary statistics of genetic diversity in the eco-geographical / genetic group of wild and cultivated apricots ($n=577$ individuals, 34 SSR markers).

Eco-geographical/genetic group	<i>N</i>	<i>H_o</i>	<i>H_E</i>	<i>Ar</i>	<i>Ap</i>	Site_ID
Cultivated						
European/Irano-Caucasian cultivated apricots	87	0.585 (0.196)	0.663 (0.209)	1.661	0.283	Cult02_EUR, Cult11_TUR, Cult01_AZE
South Central Asian cultivated apricots	47	0.648 (0.181)	0.724 (0.165)	1.729	0.269	Cult12_SCA
North Central Asian cultivated apricots	57	0.650 (0.166)	0.754 (0.165)	1.756	0.270	Cult(04, 05, 06, 07, 08, 09)_KAZ, Cult10_UZB, Cult13_KGZ
Chinese cultivated apricots (Xinjiang)	14	0.674 (0.210)	0.740 (0.127)	1.764	0.297	ch_113, ch_114, ch_117, ch_119, ch_120, ch_121, ch_136, ch_137, CH211, CH213, CH214, CH215, CH216, CH217
Chinese apricot landraces	57	0.653 (0.188)	0.738 (0.192)	1.739	0.412	Chinese_landraces
<i>Prunus mume</i>	9	0.549 (0.229)	0.652 (0.232)	1.714	0.739	<i>Prunus mume</i>
Wild						
Central Asian wild apricots (Sary Chelek)	22	0.656 (0.197)	0.700 (0.160)	1.716	0.311	Wild15_KGZ
North Central Asian wild apricots	190	0.661 (0.161)	0.724 (0.169)	1.724	0.296	Wild(02, 03, 04, 05, 07, 08)_KAZ, Wild(08, 10, 11, 12, 13, 14)_KGZ, Wild01_CHN
<i>Prunus sibirica</i> from northwest China	34	0.708 (0.182)	0.783 (0.175)	1.791	0.416	<i>Prunus sibirica</i> pop 1, 2, 3
<i>Prunus sibirica</i> from northeast China	50	0.637 (0.248)	0.736 (0.255)	1.737	0.620	<i>Prunus sibirica</i> pop 4, 5, 6, 7
<i>Prunus mandshurica</i>	8	0.500 (0.282)	0.565 (0.220)	1.605	0.669	<i>Prunus mandshurica</i>
<i>Prunus brigantina</i>	2	0.280 (0.352)	0.291 (0.261)	1.399	0.669	<i>Prunus brigantina</i>

Footnote: *N*: number of individuals, *H_E* and *H_o*: expected and observed heterozygosities, *Ar* and *Ap*: allelic richness and private allelic richness, Site_ID: Identification name of the site included in each eco-geographical groups. In brackets, standard deviations.

Table I-S11. Summary of the Wilcoxon signed rank tests to investigate differences in genetic diversity in the eco-geographical groups of wild and cultivated apricots presented in Table S10 ($n=577$ individuals, 34 SSR markers). Significantly lower and higher diversities in the cultivated groups are highlighted in light blue and red respectively ($FDR_Q < 0.05$).

Eco-geographical group	Value 2	Compare with :															
		Central Asian wild apricots (Sary Chelek)				North Central Asian wild apricots				<i>Prunus sibirica</i> from northwest China				<i>Prunus sibirica</i> from northeast China			
Value 1		HO	HE	AR	AP	HO	HE	AR	AP	HO	HE	AR	AP	HO	HE	AR	AP
European/Irano-Caucasian cultivated apricots	value 1	0.585	0.663	1.661	0.283	0.585	0.663	1.661	0.283	0.585	0.663	1.661	0.283	0.585	0.663	1.661	0.283
	value1 std	0.196	0.209	0.211	0.221	0.196	0.209	0.211	0.221	0.196	0.209	0.211	0.221	0.196	0.209	0.211	0.221
	value 2	0.656	0.700	1.716	0.311	0.661	0.724	1.724	0.296	0.708	0.783	1.791	0.416	0.637	0.735	1.737	0.620
	value2 std	0.197	0.160	0.166	0.235	0.161	0.169	0.172	0.192	0.182	0.175	0.179	0.241	0.248	0.225	0.259	0.364
	p-value	0.038	0.234	0.078	0.322	0.008	0.035	0.027	0.342	0.000	0.000	0.000	0.000	0.012	0.004	0.005	0.000
	FDR_Q	0.073	0.277	0.049	0.425	0.026	0.057	0.016	0.434	0.001	0.000	0.000	0.001	0.031	0.009	0.002	0.001
South Central Asian cultivated apricots	value 1	0.648	0.723	1.729	0.269	0.648	0.723	1.729	0.269	0.648	0.723	1.729	0.269	0.648	0.723	1.729	0.269
	value1 std	0.181	0.165	0.169	0.180	0.181	0.165	0.169	0.180	0.181	0.165	0.169	0.180	0.181	0.165	0.169	0.180
	value 2	0.656	0.700	1.716	0.311	0.661	0.724	1.724	0.296	0.708	0.783	1.791	0.416	0.637	0.735	1.737	0.620
	value2 std	0.197	0.160	0.166	0.235	0.161	0.169	0.172	0.192	0.182	0.175	0.179	0.241	0.248	0.225	0.259	0.364
	p-value	0.976	0.177	0.638	0.285	0.704	0.803	0.976	0.478	0.027	0.000	0.001	0.000	0.749	0.110	0.144	0.000
	FDR_Q	0.991	0.221	0.590	0.383	0.877	0.815	0.976	0.584	0.056	0.000	0.000	0.000	0.883	0.152	0.098	0.000
North Central Asian cultivated apricots	value 1	0.650	0.753	1.756	0.270	0.650	0.753	1.756	0.270	0.650	0.753	1.756	0.270	0.650	0.753	1.756	0.270
	value1 std	0.166	0.165	0.167	0.176	0.166	0.165	0.167	0.176	0.166	0.165	0.167	0.176	0.166	0.165	0.167	0.176
	value 2	0.656	0.700	1.716	0.311	0.661	0.724	1.724	0.296	0.708	0.783	1.791	0.416	0.637	0.735	1.737	0.620
	value2 std	0.197	0.160	0.166	0.235	0.161	0.169	0.172	0.192	0.182	0.175	0.179	0.241	0.248	0.225	0.259	0.364
	p-value	0.865	0.003	0.032	0.211	0.373	0.039	0.008	0.038	0.043	0.003	0.001	0.000	0.881	0.497	0.535	0.000
	FDR_Q	0.936	0.007	0.019	0.297	0.560	0.062	0.001	0.058	0.082	0.006	0.000	0.000	0.908	0.556	0.462	0.000
Chinese cultivated apricots (Xinjiang)	value 1	0.674	0.739	1.764	0.297	0.674	0.739	1.764	0.297	0.674	0.739	1.764	0.297	0.674	0.739	1.764	0.297
	value1 std	0.210	0.127	0.133	0.187	0.210	0.127	0.133	0.187	0.210	0.127	0.133	0.187	0.210	0.127	0.133	0.187
	value 2	0.656	0.700	1.716	0.311	0.661	0.724	1.724	0.296	0.708	0.783	1.791	0.416	0.637	0.735	1.737	0.620
	value2 std	0.197	0.160	0.166	0.235	0.161	0.169	0.172	0.192	0.182	0.175	0.179	0.241	0.248	0.225	0.259	0.364
	p-value	0.472	0.035	0.007	0.719	0.653	0.549	0.018	0.779	0.190	0.001	0.014	0.001	0.589	0.131	0.516	0.000
	FDR_Q	0.677	0.058	0.003	0.791	0.828	0.594	0.010	0.843	0.292	0.002	0.007	0.003	0.762	0.174	0.430	0.000
Chinese apricot landraces	value 1	0.653	0.737	1.739	0.412	0.653	0.737	1.739	0.412	0.653	0.737	1.739	0.412	0.653	0.737	1.739	0.412
	value1 std	0.188	0.192	0.195	0.245	0.188	0.192	0.195	0.245	0.188	0.192	0.195	0.245	0.188	0.192	0.195	0.245
	value 2	0.656	0.700	1.716	0.311	0.661	0.724	1.724	0.296	0.708	0.783	1.791	0.416	0.637	0.735	1.737	0.620
	value2 std	0.197	0.160	0.166	0.235	0.161	0.169	0.172	0.192	0.182	0.175	0.179	0.241	0.248	0.225	0.259	0.364
	p-value	0.741	0.059	0.215	0.003	0.992	0.263	0.263	0.001	0.021	0.000	0.001	0.968	0.810	0.254	0.294	0.000
	FDR_Q	0.890	0.087	0.153	0.004	0.992	0.300	0.191	0.002	0.047	0.001	0.000	0.968	0.938	0.295	0.223	0.000

Footnote: value1 is for cluster in the line; value 2 is for cluster in the column; significance level is chosen at the 0.05 level by two-tailed test; FDR_Q is the adjusted p-value calculated using Benjamini-Hochberg method; red: diversity estimate significantly higher in the cultivated group than in the wild group; blue: diversity estimate significantly lower in the cultivated group than in the wild group.

ANNEXE 1

Table I-S12. Inferences of recent effective population size reductions with the BOTTLENECK software for each of the three cultivated populations (PaEUR_C1, PaSCA_C2, and China_C4).

	mean_N	mean_k	mean_H _E	one tail P-value for heterozygosity excess (TPM)	one tail P-value for heterozygosity excess (SMM)
PaEUR_C1	123.44	6.64	0.666	0.511	0.955
PaSCA_C2	48.64	6.32	0.690	0.356	0.879
China_C4	55.44	8.32	0.671	1.000	1.000

Footnote: TPM: Two Phase Model; SMM Stepwise Mutation Model; N: Sample size; k: number of alleles; H_E: expected heterozygosity

Table I-S13. Random forest approximate Bayesian computation (ABC-RF) parameter estimates for the two most likely models of domestication scF_GF_5_GF and scF_GF_8_GF.

	scF_GF_5_GF				scF_GF_8_GF				Average over scF_GF_5_GF and scF_GF_8_GF		
	Mean	q5%	q95%	NMAE	Mean	q5%	q95%	NMAE	Mean	q5%	q95%
μ	7.08E-05	2.70E-05	9.89E-05	0.4177837186	6.88E-05	2.04E-05	9.85E-05	0.3918606288	0.00006978944685	0.0000237058	0.00009866427836
T_{W1-ANC}	12304096.490	8030540	15966430	0.161	11972355.150	8208090	15763250	0.162	12138225.820	8119315	15864840
T_{W4-ANC}	11028116.560	8040150	15633340	0.160	11664698.140	8029390	15975290	0.163	11346407.350	8034770	15804315
T_{W2-W1}	127156.869	5053.318	442960	1.259	119541.761	7470	315470	1.207	246698.629	6261.659	379215
$T_{W3-W1-W4}$	201527.523	11950	524230	0.752	137526.954	13280	424300	0.699	339054.477	12615	474265
T_{C4-W2} or T_{C4-W3}	65263.939	1770	272190	2.485	92971.739	2730	281730	2.245	79117.839	2250	276960
T_{C2-W1}	9632.387	420	28520	1.148	9454.492	420	27100	1.179	9543.440	420	27810
T_{C1-W2}	4391.467	200	21470	0.683	3570.959	190	17410	0.680	3981.213	195	19440
N_{C1}	4491.978	596	9573	0.960	4927.104	552	9758	0.976	4709.541	574	2836231
N_{C2}	4584.016	582	9665.725	1.116	5449.880	652	9923	1.068	5016.948	617	9794
N_{C4}	5187.881	585	9838	1.200	4823.837	544	9817	1.160	5005.859	565	2895399
N_{W1}	3942.531	554	9497	0.953	3441.565	544	9099	0.934	3692.048	549	9298
N_{W2}	4482.450	578	9856	0.989	4457.393	606	9533	0.992	4469.921	592	2846823
N_{W3}	4441.847	602	9704	0.703	4644.776	574	9796.720	0.675	4543.312	588	9750.349
N_{W4}	5737.955	582	9911	1.062	5993.127	593	9741	1.084	5865.541	588	9826
M_{C1-C4}	0.024	0.002	0.048	0.750	0.025	4.515E-04	0.049	0.829	0.025	0.001	0.048
M_{C2-C4}	0.023	3.334E-04	0.046	48.773	0.022	0.002	0.048	1.234	0.023	0.001	0.047
M_{C4-C1}	0.027	0.002	0.050	2.696	0.029	0.001	0.050	8.975	0.028	0.002	0.050
M_{C4-C2}	0.020	0.001	0.045	1.871	0.019	0.001	0.047	1.835	0.019	0.001	0.046

Footnote: Mean: the average estimate value for each parameter; CI: 95% credible intervals deduced from posterior quantile estimate of order 2.5% and 97.5%; NMAE: normalized mean absolute error used to evaluate the predictive accuracy measures for each parameter; μ : mutation; T_{Wx-ANC} : divergence time between the wild *Prunus armeniaca* or *Prunus sibirica* from the unknown ancestral population; T_{Cx-Wx} : divergence time between the wild group x and cultivated group x; N_{Cx} : Effective population size of the cultivated group x; N_{Wx} : effective population size of the wild apricot group x; M_{Cx-Cy} : migration rate between the cultivated groups x and y; GF: gene flow.

Annex 2: Chapter II supporting information

Chapter II supplemental Figures

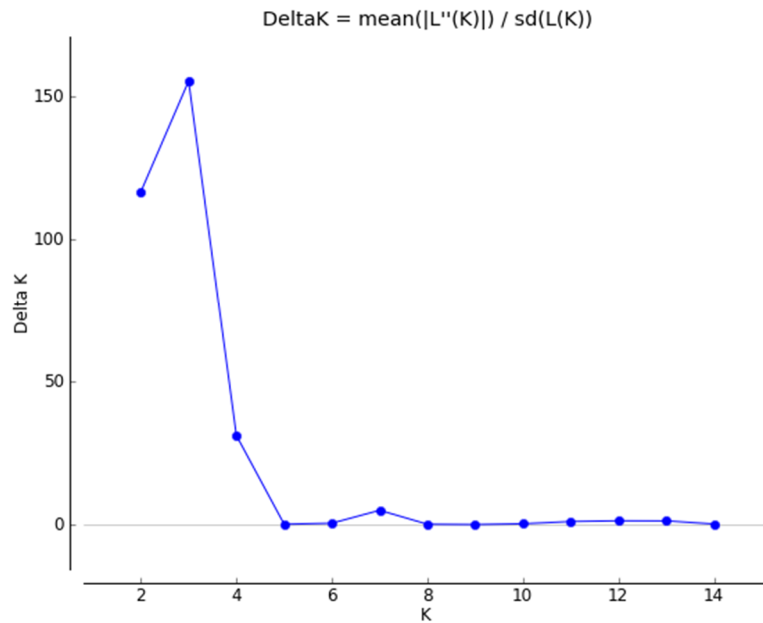


Figure II-S1. Detect the appropriate number of model components in 71 *P. brigantina* accessions.

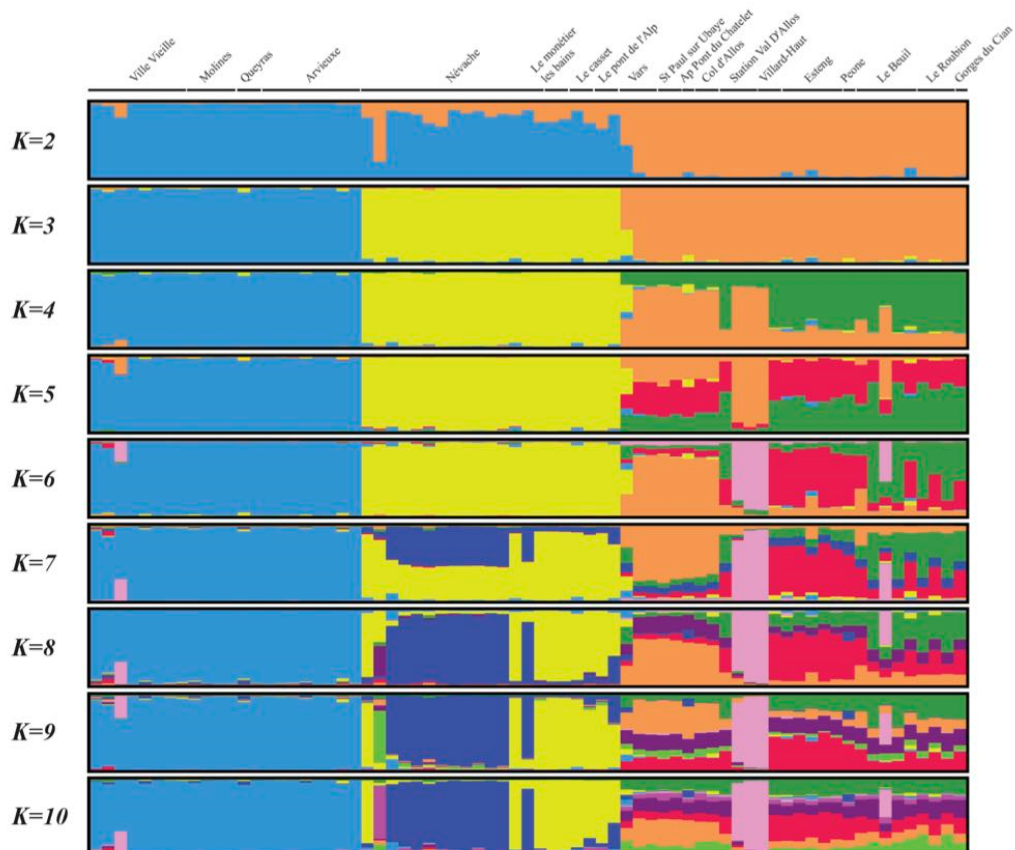


Figure II-S2. Bayesian clustering on 71 *P. brigantina* samples in French Alpen Mountains.

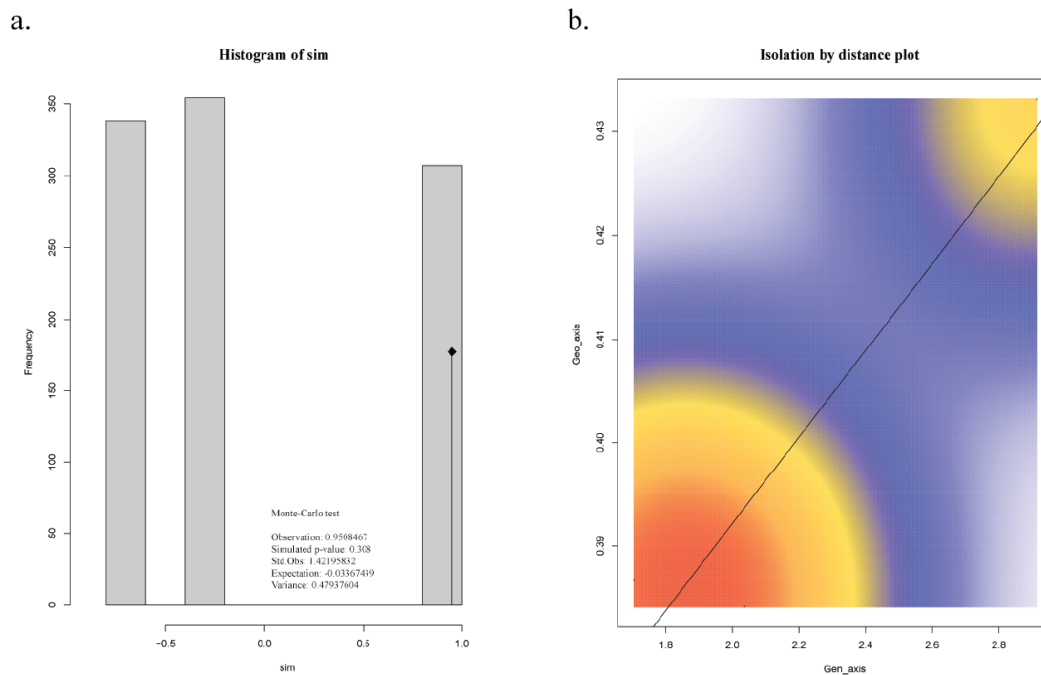


Figure II-S3. Isolation by distance (IBD) test for the three *P. brigantina* populations. a. Histogram of simulation, black square inside of the reference distribution indicated no significant spatial structure. b. IBD plot, the scatter plot clearly shows two clouds of points with discontinuities. Density estimation was displayed in colour from white to red. The solid line linked two high density clouds.

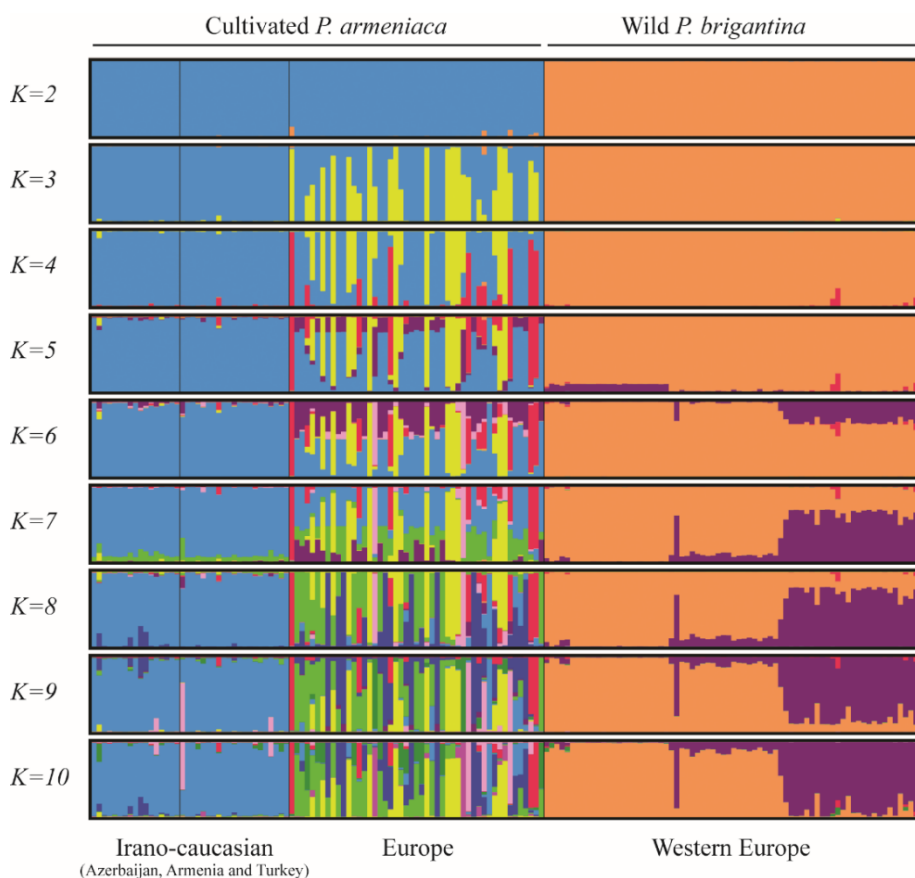


Figure II-S4. Bayesian analysis on European/Irano-caucasian cultivated *P. armeniaca* and wild *P. brigantina* accessions.

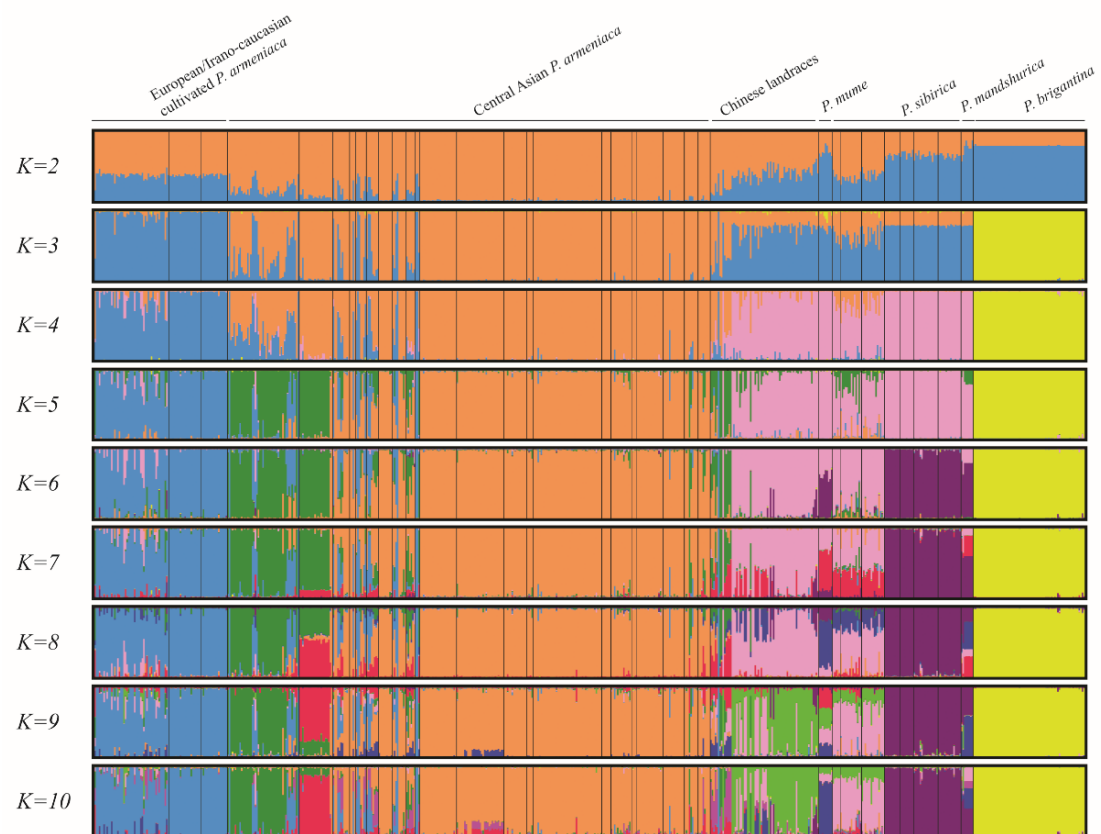


Figure II-S5. Bayesian analysis of current *P. brigantina* and large apricot dataset from the previous study.



Figure II-S6. Classification between *P. brigantina* and other *Prunus* species.

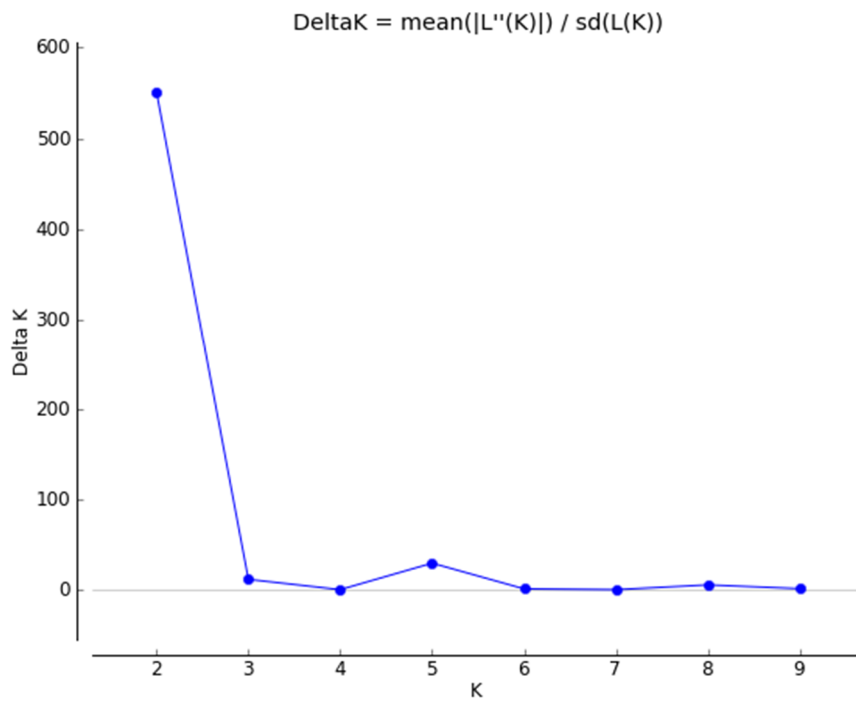


Figure II-S7. Detect the relevant *K* clusters from *P. brigantina* and other *Prunus* species dataset.

Chapter II supplemental Tables

Table II-S1a. Sampling locations and geographic regions of *P. brigantina* and *P. cerasifera* in French Alps Mountains.

Sample code	Sampling locality	Genetic cluster	Sampling location in GPS coordinates
FR-001	Ville Vieille	Queyras	44°44'29N - 6°45'47E
FR-002	Ville Vieille	Queyras	44°44'27N - 6°45'43E
FR-003	Ville Vieille	Queyras	44°44'27N - 6°45'43E
FR-004	Ville Vieille	Queyras	44°44'27N - 6°45'43E
FR-005	Ville Vieille	Queyras	44°46'8N - 6°49'56E
FR-006	Ville Vieille	Queyras	44°46'8N - 6°49'56E
FR-007	Ville Vieille	Queyras	44°46'8N - 6°49'56E
FR-008	Ville Vieille	Queyras	44°46'8N - 6°49'56E
FR-009	Molines	Queyras	44°42'21N - 6°54'00E
FR-010	Molines	Queyras	44°42'21N - 6°54'04E
FR-011	Molines	Queyras	44°42'21N - 6°54'04E
FR-012	Molines	Queyras	44°42'21N - 6°54'047E
FR-013	Arvieux	Queyras	44°44'38N - 6°46'11E
FR-014	Arvieux	Queyras	44°44'27N - 6°45'43E
FR-015	Arvieux	Queyras	44°47'54N - 6°44'28E
FR-016	Arvieux	Queyras	44°47'54N - 6°44'28E
FR-017	Arvieux	Queyras	44°44'29N - 6°45'42E
FR-018	Arvieux	Queyras	44°44'27N - 6°45'43E
FR-019	Arvieux	Queyras	44°47'54N - 6°44'28E
FR-020	Arvieux	Queyras	44°47'49N - 6°44'29E
FR-021-A	Arvieux	Queyras	44°44'27N - 6°45'43E
FR-021-B	Arvieux	Queyras	44°44'27N - 6°45'43E
FR-023	Névache	Ecrins	44°54'3N - 6°39'4E
FR-024	Névache	Ecrins	44°54'3N - 6°39'4E
FR-025	Névache	Ecrins	44°57'34N - 6°40'28E
FR-026	Névache	Ecrins	44°57'36N - 6°40'53E
FR-027	Névache	Ecrins	44°57'37N - 6°40'30E
FR-028	Névache	Ecrins	44°57'37N - 6°40'30E
FR-029	Névache	Ecrins	44°58'30N - 6°40'32E
FR-030-1	Névache	Ecrins	45°01'48N - 6°34'46E
FR-030-2	Névache	Ecrins	44°59'11N - 6°38'48E
FR-031	Névache	Ecrins	45°01'48N - 6°34'46E
FR-032	Névache	Ecrins	45°01'48N - 6°34'46E
FR-033	Névache	Ecrins	44°57'40N - 6°40'26E
FR-034-1	Névache	Ecrins	44°56'16N - 6°34'44E
FR-034-2	Névache	Ecrins	44°01'15N - 6°35'17E
FR-035	Névache	Ecrins	44°58'5N - 6°31'37
FR-036	Le monétier les bains	Ecrins	44°58'34N - 6°30'51E
FR-037	Le monétier les bains	Ecrins	44°58'52N - 6°29'58E
FR-038	Le casset	Ecrins	44°59'17N - 6°29'26E
FR-039-A	Le casset	Ecrins	45°0'39N - 6°28'24E
FR-039-B	Le pont de l'Alp	Ecrins	45°1'7N - 6°27'54E
FR-040-A	Le pont de l'Alp	Ecrins	45°1'24N - 6°27'21E
FR-041	Vars	Mercantour	44°26'22N - 6°41'19E
FR-042	Vars	Mercantour	44°35'57N - 6°41'23E
FR-043	Vars	Mercantour	44°31'45N - 6°43'2E
FR-044	St Paul sur Ubaye	Mercantour	44°31'2N - 6°45'25E
FR-045	St Paul sur Ubaye	Mercantour	44°31'28N - 6°47'47E
FR-046	Ap Pont du Chatelet	Mercantour	44°31'45N - 6°47'18E
FR-047	Col d'Allos	Mercantour	44°19'18N - 6°36'6E
FR-048	Col d'Allos	Mercantour	44°19'12N - 6°35'57E
FR-049	Station Val D'Allos	Mercantour	44°16'54N - 6°34'23E
FR-050	Station Val D'Allos	Mercantour	44°16'3N - 6°34'52E
FR-051	Station Val D'Allos	Mercantour	44°14'36N - 6°38'46E
FR-052	Villard-Haut	Mercantour	44°14'43N - 6°38'58E
FR-053	Villard-Haut	Mercantour	44°9'56N - 6°42'40E
FR-054	Esteng	Mercantour	44°9'13N - 6°43'53E
FR-055	Esteng	Mercantour	44°14'5N - 6°45'8E
FR-056	Esteng	Mercantour	44°14'39N - 6°45'18E
FR-057	Esteng	Mercantour	44°14'28N - 6°45'9E
FR-058	Esteng	Mercantour	44°13'10N - 6°44'45E
FR-059	Peone	Mercantour	44°7'38N - 6°55'2E
FR-060	Le Beuil	Mercantour	44°6'7N - 6°59'11E
FR-061	Le Beuil	Mercantour	44°6'4N - 6°59'36E
FR-062	Le Beuil	Mercantour	44°6'4N - 6°59'36E
FR-063	Le Beuil	Mercantour	44°6'7N - 6°59'36E
FR-064	Le Beuil	Mercantour	44°6'9N - 6°59'45E
FR-065	Le Roubion	Mercantour	44°5'15N - 7°2'18E
FR-066	Le Roubion	Mercantour	44°5'13N - 7°2'18E
FR-067	Le Roubion	Mercantour	44°5'24N - 7°2'21E
FR-068	Gorges du Cian	Mercantour	44°1'50N - 6°58'37E
FR-070*	Entraume		44°11'21N - 6°45'3E

Footnotes: * individual is *P. cerasifera* (Myrobalan)

Table II-S1b: List of individuals included in the different datasets of Table 1 together with the *P. brigantina* samples.INRA data portal: <https://doi.org/10.15454/FFKGUP>.**Table II-S2. Analysis of genetic variability from microsatellites for *P. brigantina* population.**

Locus	% missing	<i>N_a</i>	<i>N_e</i>	<i>I</i>	<i>H_o</i>	<i>H_e</i>
AMPA100	7.04	1	1.00	0.00	0.00	0.00
AMPA101	1.41	5	2.57	1.09	0.21	0.61
AMPA103	1.41	8	3.67	1.50	0.56	0.73
aprigms18	4.23	5	2.34	1.00	0.49	0.57
BPPCT004	4.23	2	1.14	0.24	0.10	0.12
BPPCT025	2.82	3	1.52	0.64	0.25	0.34
BPPCT040	5.63	8	5.25	1.75	0.60	0.81
CPPCT030	1.41	1	1.00	0.00	0.00	0.00
CPPCT006	0.00	6	3.40	1.32	0.49	0.71
CPPCT033	1.41	11	3.94	1.78	0.69	0.75
EPPCU0532	4.23	1	1.00	0.00	0.00	0.00
G22SSR	1.41	2	1.01	0.04	0.01	0.01
SSRLg1_11m52a	0.00	13	4.50	1.89	0.66	0.78
N86B11ssr3	2.82	3	1.92	0.83	0.43	0.48
pchgms03	2.82	1	1.00	0.00	0.00	0.00
PGS-1.21	1.41	3	1.24	0.37	0.19	0.19
SSR04AG51	0.00	9	4.06	1.72	0.55	0.75
SSR5piso4E	2.82	1	1.00	0.00	0.00	0.00
SSR5piso4Ga	1.41	3	2.35	0.94	0.37	0.57
UDA-021	2.82	2	1.33	0.41	0.17	0.25
UDAp-414	0.00	8	5.20	1.79	0.48	0.81
UDAp-480	2.82	8	2.61	1.25	0.43	0.62
UDP98-409	2.82	11	4.28	1.77	0.58	0.77
UDP96-018	4.23	6	2.19	0.95	0.31	0.54
Mean	2.46	5.042	2.48	0.89	0.32	0.43

Footnote: *N_a*: number of different alleles, and *N_e*: number of effective alleles. *I*: Shannon diversity index. *H_e* and *H_o*: expected and observed heterozygosities.

Table II-S3. The description of representatives from the core collection of *P. brigantina*.

Percentage of genetic diversity of the core collection	100%	90%	80%	70%	60%	50%	40%	30%	<= 20%
	FR_003	FR_003	FR_010	FR_025	FR_003	FR_032	FR_003	FR_003	FR_039_B
	FR_009	FR_010	FR_014	FR_030_1	FR_037	FR_036	FR_037	FR_058	
	FR_010	FR_014	FR_032	FR_051	FR_043	FR_051	FR_057		
	FR_012	FR_029	FR_037	FR_052	FR_052	FR_065			
	FR_014	FR_032	FR_043	FR_057	FR_059				
	FR_015	FR_037	FR_045	FR_063	FR_065				
	FR_019	FR_043	FR_051	FR_065					
	FR_023	FR_045	FR_052	FR_066					
	FR_025	FR_050	FR_057						
	FR_028	FR_051	FR_063						
	FR_029	FR_052	FR_065						
	FR_030_1	FR_057	FR_066						
	FR_032	FR_059							
	FR_036	FR_063							
	FR_037	FR_065							
	FR_039_A	FR_066							
	FR_043	FR_067							
Individuals in the core collection	FR_045	FR_068							
	FR_046								
	FR_047								
	FR_048								
	FR_049								
	FR_050								
	FR_051								
	FR_052								
	FR_053								
	FR_057								
	FR_058								
	FR_059								
	FR_061								
	FR_063								
	FR_064								
	FR_065								
	FR_066								
	FR_067								
	FR_068								
Core collection size	36	18	12	8	6	4	3	2	1

ANNEXE 2

Table II-S4. Genetic variability from microsatellites for *P. brigantina* core collection.

Locus	% miss	<i>N_a</i>	<i>N_e</i>	<i>I</i>	<i>H_o</i>	<i>H_e</i>
AMPA100	3.1	1	1.00	0.00	0.00	0.00
AMPA101	0.0	5	2.83	1.21	0.31	0.65
ampa103	0.0	8	3.39	1.51	0.47	0.71
aprigms18	3.1	5	2.45	1.07	0.55	0.59
BPPCT004	6.3	2	1.14	0.25	0.07	0.12
BPPCT025	0.0	3	1.47	0.60	0.13	0.32
BPPCT040	6.3	8	5.83	1.84	0.67	0.83
CCPCT030	0.0	1	1.00	0.00	0.00	0.00
CPPCT006	0.0	6	3.75	1.43	0.53	0.73
CPPCT033	0.0	11	5.69	2.03	0.75	0.82
eppco0532	0.0	1	1.00	0.00	0.00	0.00
g22ssr	3.1	2	1.03	0.08	0.03	0.03
L11m52a	0.0	13	5.36	2.10	0.69	0.81
N86B11SSR3	0.0	3	1.96	0.84	0.41	0.49
pchms03	0.0	1	1.00	0.00	0.00	0.00
PGS-21a	3.1	3	1.34	0.47	0.23	0.25
ssr04ag51	0.0	9	4.46	1.78	0.56	0.78
ssr5piso4e	3.1	1	1.00	0.00	0.00	0.00
ssr5piso4ga	0.0	3	2.00	0.80	0.22	0.50
uda021	0.0	2	1.44	0.48	0.25	0.31
udap-414	0.0	8	4.84	1.75	0.41	0.79
udap-480	0.0	8	3.14	1.46	0.38	0.68
UDP409	6.3	11	6.48	2.06	0.70	0.85
udp96-018	3.1	6	2.30	1.07	0.39	0.57
Mean	1.6	5.042	2.75	0.95	0.32	0.45

Footnote: *N_a*: number of different alleles, and *N_e*: number of effective alleles. *I*: shannon diversity index. *H_e* and *H_o*: expected and observed heterozygosities.

Table II-S5. Mann-Whitney U tests (two-tailed) between the whole dataset *P. brigantina* population and its core collection.

Mann-Whitney U Test (two-tailed)	<i>I</i>	<i>H_o</i>	<i>H_e</i>	μH_e
<i>Z-value</i>	0.330	-0.052	-0.340	-0.412
<i>p-value</i>	0.741	0.960	0.728	0.682

Footnotes: significant at 0.05 level.

Annex 3: Chapter III supporting information.

Supplementary notes on the de novo assembled apricot genome

In the frame of the ANR CHEX ABRIWG project, 93 apricot cultivars from the INRA's Centre National de Genetic Resources of Avignon (France) were sequenced through the ILLUMINA technology, at 15X coverage (Mariette et al. 2016). All accessions belong to the European-Irano-Caucasian genetic cluster as described in Decroocq et al (2016). 100 bp paired-end sequencing was performed by GATC Inc. and David H. Murdock Institute, Kannapolis, NC utilizing the Illumina HiSeq 2000/2500 platforms. After quality filtering, the remaining reads were mapped onto peach genome v.1 (Verde et al., 2013), using only the eight longest scaffolds which represent ~99% of the genome (for details see Decroocq et al, 2016). The aligned reads were indexed with SAMtools v0.1.18 (Li et al., 2009) (<http://samtools.sourceforge.net/>) for fastest random access keeping only those that aligned on the annotated peach gene space. All SNPs predicted by samtools for each accession were used to calculate the homozygosity and residual heterozygosity for each chromosome. Homozygosity varied depending on the accession and on the chromosome (Table SN-1). For example, homozygosity ranged for chromosome 1 from 54.85% (accession IS 13-91) to 95.04% (Marouch #14). The most homozygous apricot genomes were: Marouch #14 (A2067 in the INRA's collection) and Bakour, which both originate from North-Africa, Morocco and Tunisia respectively. They are both self-compatible and were propagated by seedlings, mainly. We selected the cultivar Marouch #14 for long-range sequencing and optical mapping to produce the final apricot (*Prunus armeniaca*) assembled the genome. It will be later called 'Marouch' for the rest of the PhD manuscript.

To perform *de novo* assembly of the Marouch genome, we combined three different technologies: short-read sequencing, long-read sequencing and optical mapping (Figure SN-1).

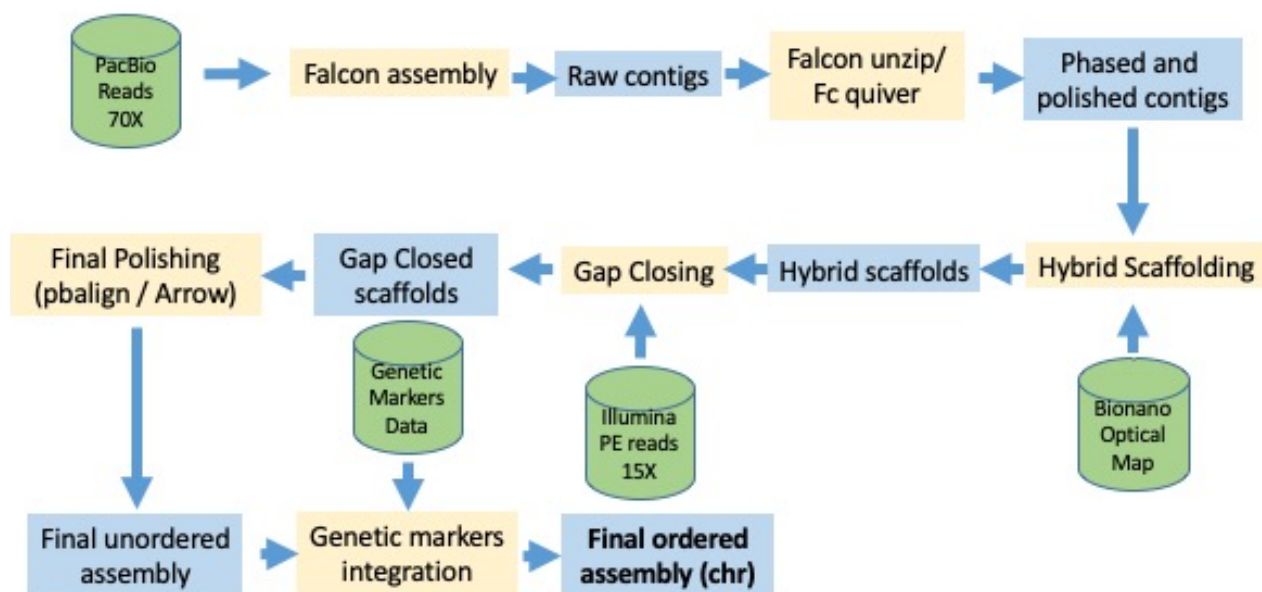


Figure SN-1. Overview of the processing pipeline used for the *de novo* assembly of the Marouch genome.

The final Bionano assembly produced 61 scaffolds which were then mapped onto the apricot consensus genetic map. All 61 scaffolds resulting from the hybrid assembly of the BioNano optical map and Illumina/PacBio first assembly were mined for SSR (Single Sequence Repeat) markers that were genetically mapped on a high density multi-family genetic linkage map, by aligning their primer sequences (with 10 bp of undefined bases (N) added between the forward and reverse primer sequences). Based on this information 59 scaffolds were anchored, oriented and assembled into 8 pseudomolecules corresponding to the **8 chromosomes**. Only two scaffolds (of 6.85 Mb in total) could not be assigned to a chromosome (Table SN-2). BUSCO assessment showed that 95.62% of the coding sequences were complete with 93.61% in single-copy. Only 3.05% of the predicted coding sequences were missing in the Marouch v2.0 assembled genome.

Genome annotation was implemented following the NCBI Eukaryotic genome annotation pipeline https://www.ncbi.nlm.nih.gov/genome/annotation_euk/process/. Messenger RNA was predicted with Braker2 (Li et al., 2009), repeat elements and transposons with REPET (Flutre et al., 2011) while contigs of Marouch young leaf RNA were assembled with TRINITY (Grabherr et al., 2011). BlastP and blastX were performed through the NCBI Blast platform. The apricot Gbrowse is accessible under the link: <https://services.cbib.u-bordeaux.fr/apollo/annotator>.

Table SN-1. Homozygosity over the coding sequence of Marouch #14.

Total Marouch #14 SNPs 1078669	Scaffold 1		Scaffold 2		Scaffold 3		Scaffold 4		Scaffold 5		Scaffold 6		Scaffold 7		Scaffold 8	
	Hmz	Htz	Hmz	Htz	Hmz	Htz	Hmz	Htz	Hmz	Htz	Hmz	Htz	Hmz	Htz	Hmz	Htz
	95.04 %	4.96 %	92.46 %	7.54 %	93.24 %	6.76 %	92.39 %	7.61 %	89.67 %	10.33 %	94.22 %	5.78 %	94.77 %	5.23 %	91.99 %	8.01 %

Footnote: Hmz refers to the homozygous; Htz refers to heterozygous.

Table SN-2. Brief information on Marouch v2.0 *de novo* assembly.

	Final assembly v2.0
Total number of Scaffolds	10
Chromosomes	8
length	195.85 Mb
%	99.20%
Unplaced Scaffolds	2
length	6.85 Mb
%	0.80%

Reference

Decroocq, S., Cornille, A., Tricon, D., Babayeva, S., Chague, A., Eyquard, J. P., ... & Liu, W. (2016). New insights into the history of domesticated and wild apricots and its contribution to Plum pox virus resistance. *Molecular Ecology*, 25(19), 4712-4729.

Flutre, T., Duprat, E., Feuillet, C., & Quesneville, H. (2011). Considering transposable element diversification in *de novo* annotation approaches. *PLoS ONE*, 6(1), e16526.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., ... & Chen, Z. (2011). Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nature Biotechnology*, 29(7), 644.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... & Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16), 2078-2079.

Mariette, S., Wong Jun Tai, F., Roch, G., Barre, A., Chague, A., Decroocq, S., ... & Nikolski, M. (2016). Genome-wide association links candidate genes to resistance to Plum Pox Virus in apricot (*Prunus armeniaca*). *New Phytologist*, 209(2), 773-784.

Verde, I., Abbott, A. G., Scalabrin, S., Jung, S., Shu, S., Marroni, F., ... & Zuccolo, A. (2013). The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. *Nature Genetics*, 45(5), 487.

Chapter III supplemental Figures

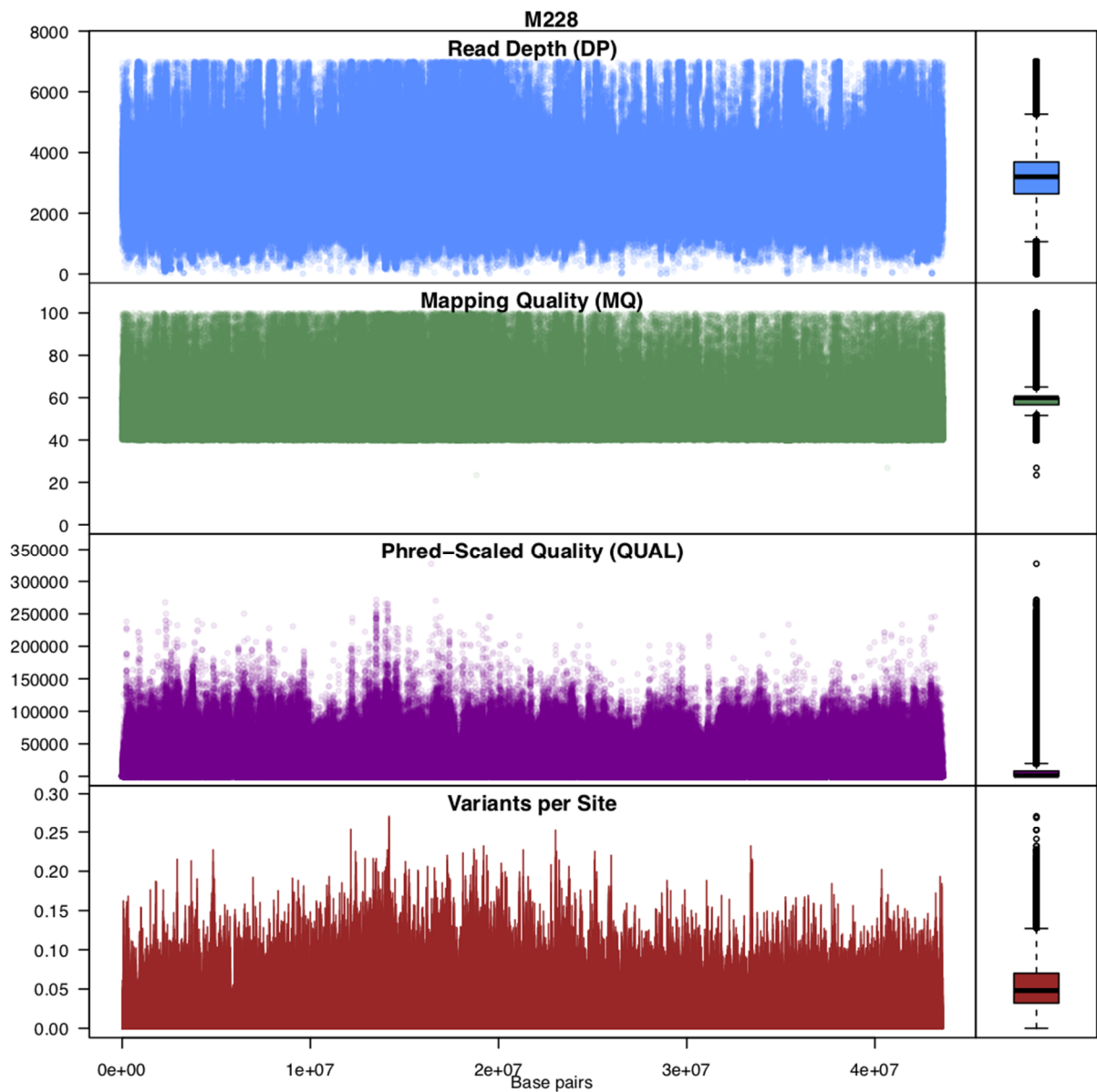


Figure III-S1. Scatter plots of read depth, mapping quality, Phred-scaled quality and SNPs variants on Chromosome 1.

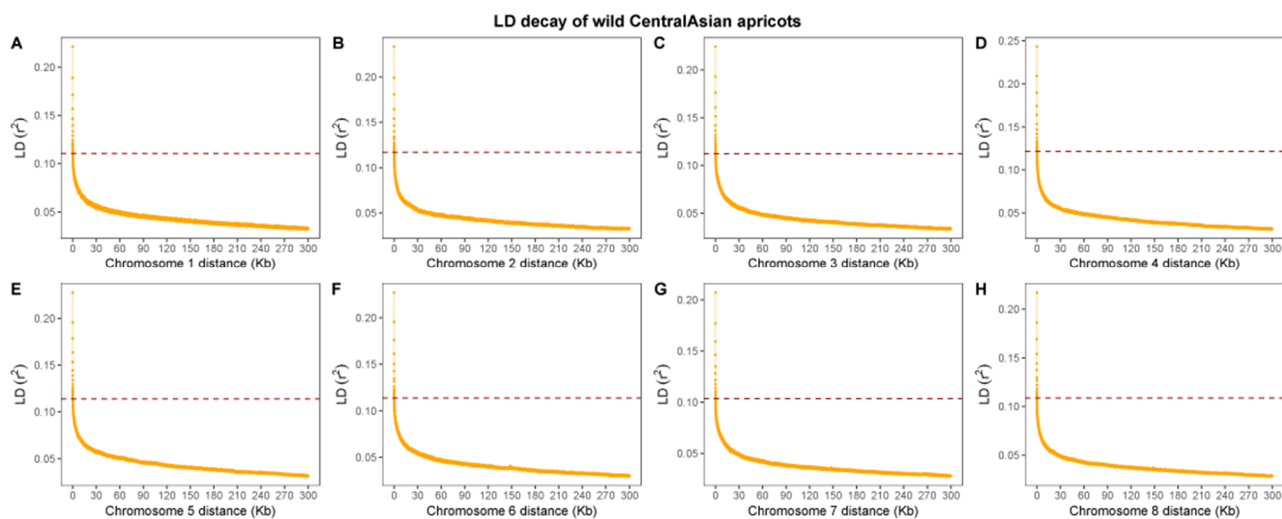


Figure III-S2. Linkage disequilibrium decay in wild Central Asian apricot genomes on eight chromosomes.

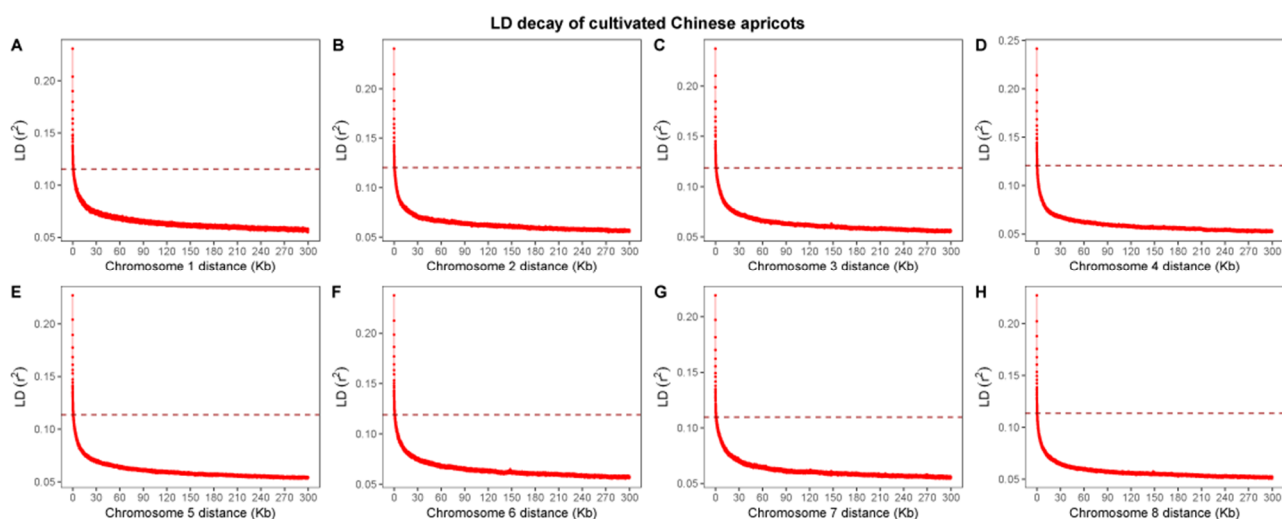


Figure III-S3. Linkage disequilibrium decay in cultivated Chinese apricot genomes on eight chromosomes.

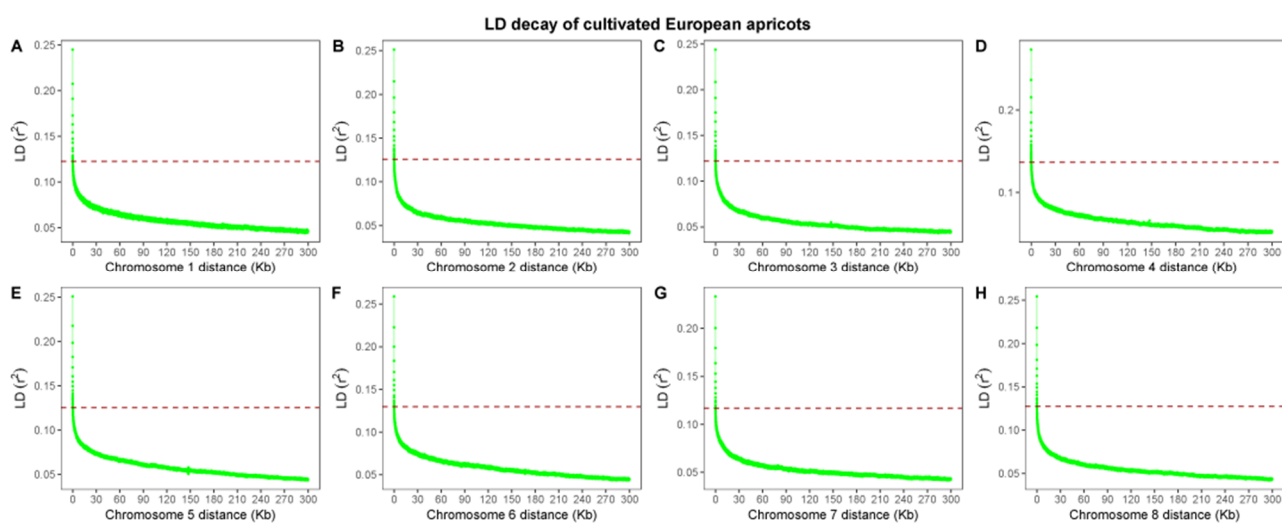


Figure III-S4. Linkage disequilibrium decay in cultivated European apricot genomes on eight chromosomes.

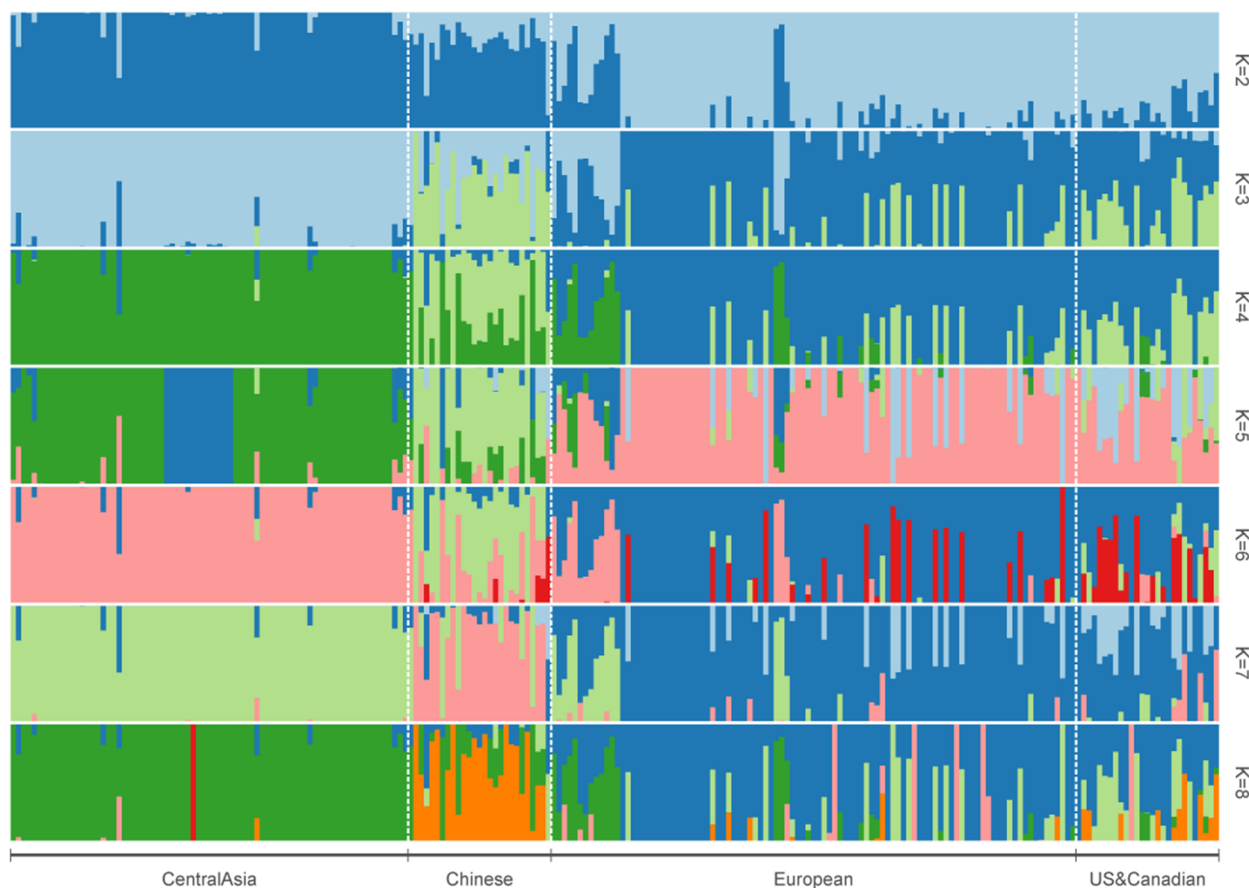


Figure III-S5. Structure plot for Central Asian, Chinese, European and North-American apricots (from left to right) based on the entire SNPs dataset. Abbreviations under the structure bar plot indicated apricot group.

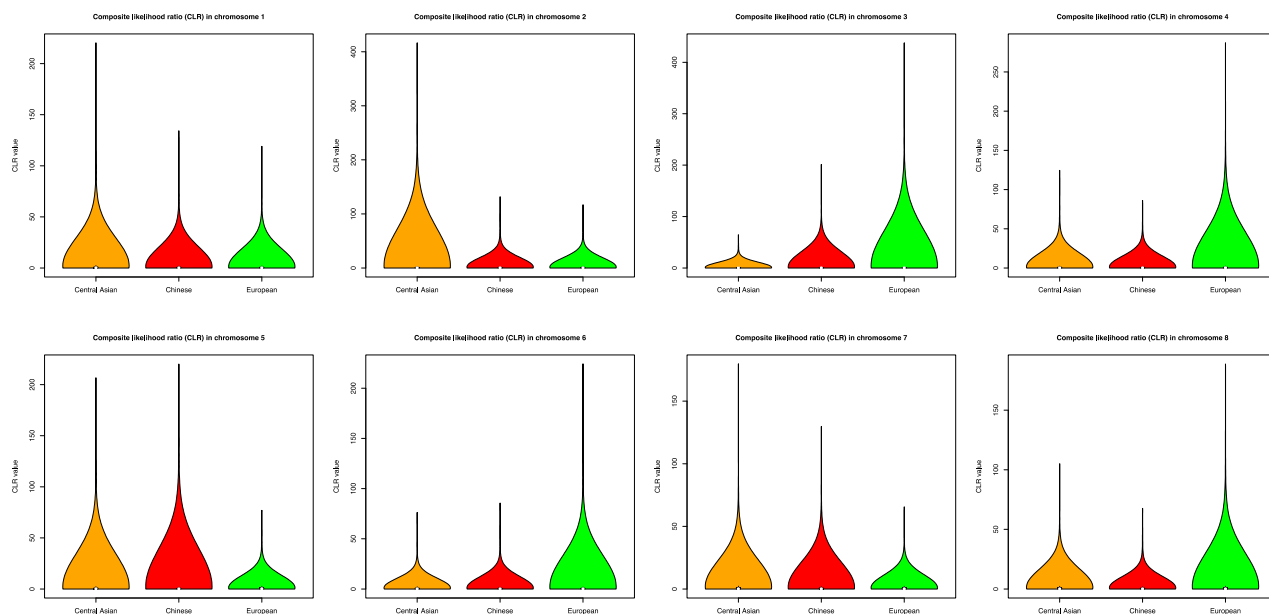


Figure III-S6. Distribution of CLR statistics on eight chromosomes among the wild Central Asian, cultivated European and Chinese apricots. Color area in violin charts indicates the density of values (width=frequency); the black vertical line was separated into four parts, they were upper adjacent value, median value (small white circle), interquartile range, and lower adjacent value.

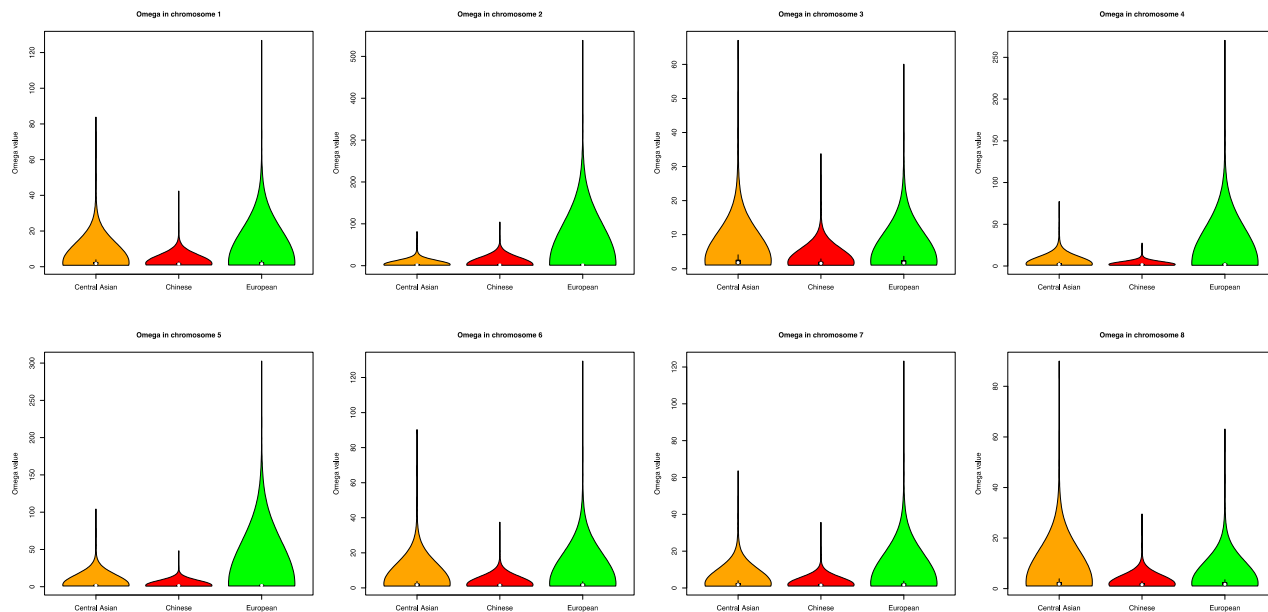


Figure III-S7. Distribution of Omega statistics on eight chromosomes among the wild Central Asian and the cultivated European and Chinese apricots. Description of violin charts similar to Figure III-S6.

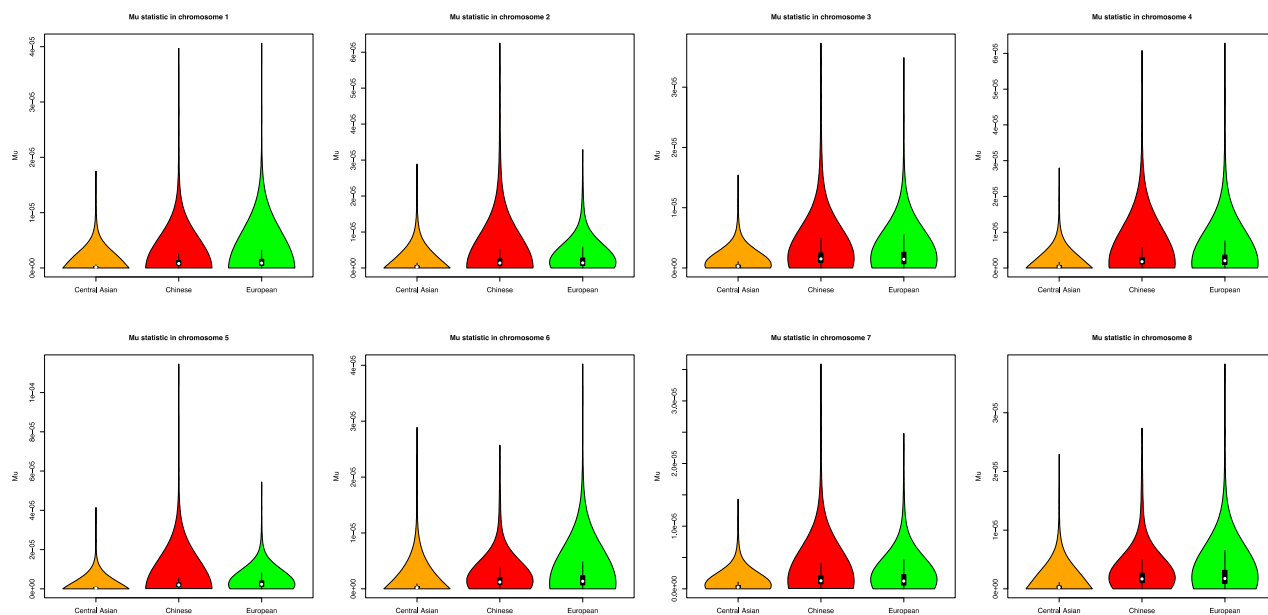


Figure III-S8. Distribution of μ statistics on eight chromosomes among the wild Central Asian, and the cultivated European and Chinese apricots. Description of violin charts similar to Figure III-S6.

Chapter III supplemental Tables

Table III-S1. Information for each accession of the study. Accession number, accession registered name, origin and available genotyping data. Footnote: EUR, N-A, CHN and CentralAsia correspond to the 4 apricot groups, they are related to their geographical origin or sampling sites in Europe, North-America, China and Central Asia, respectively. INRA data portal: <https://doi.org/10.15454/FFKGUP>.

Table III-S2. Nucleotide diversity (P_i , π) values for wild and cultivated apricot groups.

Type	Chr.	N	Chr1	Chr2	Chr3	Chr4	Chr5	Chr6	Chr7	Chr8	Mean	Std
Apricots groups	Wild Central Asian	75	5.88E-03	5.76E-03	6.49E-03	5.80E-03	5.84E-03	5.97E-03	6.81E-03	6.58E-03	6.14E-03	4.17E-04
	Cultivated Chinese	27	5.73E-03	5.58E-03	6.17E-03	5.85E-03	5.83E-03	5.87E-03	6.59E-03	6.25E-03	5.98E-03	3.28E-04
	Cultivated European	99	4.97E-03	4.63E-03	5.53E-03	4.63E-03	4.98E-03	5.07E-03	5.55E-03	4.94E-03	5.04E-03	3.48E-04
	Mean		5.53E-03	5.32E-03	6.06E-03	5.43E-03	5.55E-03	5.64E-03	6.32E-03	5.93E-03		
	Std		4.87E-04	6.08E-04	4.92E-04	6.89E-04	4.92E-04	4.96E-04	6.74E-04	8.69E-04		

Table III-S3. Tajima's D values for wild and cultivated apricot groups.

Type	Chr.	N	Chr1	Chr2	Chr3	Chr4	Chr5	Chr6	Chr7	Chr8	Mean	Std
Apricots groups	Wild Central Asian	75	5.37E-01	4.39E-01	6.11E-01	6.39E-01	6.22E-01	4.93E-01	5.67E-01	5.69E-01	5.60E-01	6.80E-02
	Cultivated Chinese	27	-2.80E-02	-6.99E-02	6.14E-02	8.17E-02	7.32E-02	2.26E-02	1.05E-02	-5.41E-02	1.22E-02	5.83E-02
	Cultivated European	99	2.92E-01	3.05E-02	3.92E-01	2.03E-01	3.76E-01	4.10E-01	1.37E-01	3.13E-02	2.34E-01	1.57E-01
	Mean		2.67E-01	1.33E-01	3.55E-01	3.08E-01	3.57E-01	3.08E-01	2.38E-01	1.82E-01		
	Std		2.83E-01	2.70E-01	2.77E-01	2.93E-01	2.75E-01	2.51E-01	2.92E-01	3.38E-01		

Table III-S4. Genetic difference F_{st} estimates for wild and cultivated apricot groups.

Groups	Cultivated Chinese	Cultivated European
Wild Central Asian	0.049	0.086
Cultivated Chinese		0.082

Table III-S5. Summary statistics of nucleotide diversity (Pi , π) test among the 8 chromosomes of wild and cultivated apricots.

Group	Chr.	Total signals	Max	Min	Mean	Std	Bottom 0.5%		Bottom 1%		Bottom 5%	
							Rank	Value	Rank	Value	Rank	Value
Three groups	All	58552	0.034	0.000	0.006	0.003	292	6.47E-05	585	1.70E-04	2927	1.24E-03
Wild Central Asian	Chr1	4350	0.029	0.000	0.006	0.003	21	1.25E-04	43	2.30E-04	217	1.48E-03
	Chr2	2546	0.026	0.000	0.006	0.004	12	2.88E-05	25	9.07E-05	127	7.95E-04
	Chr3	2187	0.029	0.000	0.006	0.004	10	7.53E-05	21	2.29E-04	109	1.62E-03
	Chr4	1901	0.026	0.000	0.006	0.003	9	4.55E-05	19	1.45E-04	95	1.19E-03
	Chr5	1723	0.034	0.000	0.006	0.003	8	3.10E-05	17	1.18E-04	86	1.49E-03
	Chr6	2568	0.027	0.000	0.006	0.003	12	7.31E-05	25	2.12E-04	128	1.53E-03
	Chr7	2358	0.029	0.000	0.007	0.004	11	7.63E-05	23	2.60E-04	117	1.88E-03
	Chr8	1885	0.032	0.000	0.007	0.004	9	9.16E-05	18	1.73E-04	94	1.80E-03
	Total	19518	0.034	0.000	0.006	0.004	97		195		975	
Cultivated Chinese	Chr1	4349	0.027	0.000	0.006	0.003	21	1.03E-04	43	1.93E-04	217	1.43E-03
	Chr2	2544	0.024	0.000	0.006	0.003	12	4.21E-05	25	9.97E-05	127	7.72E-04
	Chr3	2188	0.028	0.000	0.006	0.003	10	3.75E-05	21	1.59E-04	109	1.44E-03
	Chr4	1900	0.023	0.000	0.006	0.003	9	4.65E-05	19	1.46E-04	95	1.14E-03
	Chr5	1722	0.033	0.000	0.006	0.003	8	3.59E-05	17	1.31E-04	86	1.41E-03
	Chr6	2568	0.026	0.000	0.006	0.003	12	6.27E-05	25	1.90E-04	128	1.52E-03
	Chr7	2358	0.026	0.000	0.007	0.004	11	5.98E-05	23	2.31E-04	117	1.78E-03
	Chr8	1885	0.032	0.000	0.006	0.003	9	8.83E-05	18	2.11E-04	94	1.80E-03
	Total	19514	0.033	0.000	0.006	0.003	97		195		975	
Cultivated European	Chr1	4352	0.029	0.000	0.005	0.003	21	8.66E-05	43	2.16E-04	217	1.20E-03
	Chr2	2545	0.024	0.000	0.005	0.003	12	2.72E-05	25	7.33E-05	127	6.71E-04
	Chr3	2188	0.027	0.000	0.006	0.003	10	5.78E-05	21	1.62E-04	109	1.37E-03
	Chr4	1901	0.022	0.000	0.005	0.003	9	3.37E-05	19	1.43E-04	95	8.37E-04
	Chr5	1723	0.034	0.000	0.005	0.003	8	3.28E-05	17	1.11E-04	86	1.08E-03
	Chr6	2569	0.027	0.000	0.005	0.003	12	4.10E-05	25	1.41E-04	128	1.23E-03
	Chr7	2358	0.023	0.000	0.006	0.004	11	5.19E-05	23	2.40E-04	117	1.29E-03
	Chr8	1884	0.029	0.000	0.005	0.003	9	8.27E-05	18	2.60E-04	94	1.20E-03
	Total	19520	0.034	0.000	0.005	0.003	97		195		975	

ANNEXE 3

Table III-S6. Summary statistics of Tajima's *D* test among the 8 chromosomes of wild and cultivated apricots.

Group	Chr.	Total signals	Max	Min	Mean	Std	Top 0.5%		Top 1%		Top 5%		Bottom 0.5%		Bottom 1%		Bottom 5%	
							Rank	Value	Rank	Value	Rank	Value	Rank	Value	Rank	Value	Rank	Value
Three groups	All	58552	4.778	-2.683	0.265	0.810	292	2.590	585	2.300	2927	1.626	292	-1.788	585	-1.589	2927	-1.041
Wild Central Asian	Chr1	4350	3.369	-1.978	0.537	0.740	21	2.662	43	2.364	217	1.717	21	-1.393	43	-1.256	217	-0.691
	Chr2	2546	4.121	-2.383	0.439	0.803	12	2.625	25	2.323	127	1.736	12	-1.689	25	-1.445	127	-0.882
	Chr3	2187	4.296	-1.895	0.611	0.739	10	2.936	21	2.692	109	1.823	10	-1.404	21	-1.160	109	-0.612
	Chr4	1901	4.427	-1.879	0.639	0.799	9	3.150	19	2.635	95	1.912	9	-1.350	19	-1.254	95	-0.682
	Chr5	1723	3.354	-1.912	0.622	0.765	8	2.716	17	2.399	86	1.861	8	-1.551	17	-1.343	86	-0.689
	Chr6	2568	3.324	-1.782	0.493	0.722	12	2.573	25	2.253	128	1.657	12	-1.409	25	-1.169	128	-0.698
	Chr7	2358	3.495	-1.795	0.567	0.757	11	2.699	23	2.430	117	1.819	11	-1.443	23	-1.196	117	-0.686
	Chr8	1885	3.791	-2.079	0.569	0.701	9	2.770	18	2.453	94	1.732	9	-1.383	18	-1.112	94	-0.548
	Total	19518	4.427	-2.383	0.551	0.755	97	2.709	195	2.420	975	1.763	97	-1.437	195	-1.269	975	-0.695
Cultivated Chinese	Chr1	4349	2.674	-2.248	-0.028	0.621	21	1.852	43	1.569	217	0.974	21	-1.675	43	-1.519	217	-1.071
	Chr2	2544	2.420	-1.983	-0.070	0.619	12	1.776	25	1.444	127	0.969	12	-1.667	25	-1.481	127	-1.071
	Chr3	2188	2.986	-2.218	0.061	0.605	10	1.952	21	1.770	109	1.051	10	-1.552	21	-1.360	109	-0.921
	Chr4	1900	3.241	-1.852	0.082	0.598	9	1.868	19	1.550	95	1.041	9	-1.543	19	-1.390	95	-0.892
	Chr5	1722	2.844	-2.099	0.073	0.599	8	1.810	17	1.577	86	1.082	8	-1.682	17	-1.444	86	-0.899
	Chr6	2568	2.410	-1.782	0.023	0.589	12	1.660	25	1.557	128	1.022	12	-1.468	25	-1.315	128	-0.960
	Chr7	2358	2.633	-1.882	0.010	0.616	11	1.868	23	1.512	117	1.051	11	-1.560	23	-1.430	117	-1.040
	Chr8	1885	2.448	-2.107	-0.054	0.565	9	1.643	18	1.499	94	0.897	9	-1.679	18	-1.446	94	-0.972
	Total	19514	3.241	-2.248	0.005	0.607	97	1.799	195	1.542	975	1.008	97	-1.585	195	-1.434	975	-1.001
Cultivated European	Chr1	4352	4.488	-2.359	0.292	0.919	21	2.922	43	2.559	217	1.823	21	-1.824	43	-1.630	217	-1.157
	Chr2	2545	3.126	-2.683	0.030	0.940	12	2.490	25	2.235	127	1.580	12	-2.204	25	-2.091	127	-1.529
	Chr3	2188	4.778	-2.109	0.392	0.851	10	3.071	21	2.504	109	1.764	10	-1.699	21	-1.522	109	-1.027
	Chr4	1901	3.893	-2.537	0.203	1.047	9	2.984	19	2.674	95	1.907	9	2.984	19	2.674	95	1.907
	Chr5	1723	3.932	-2.448	0.376	0.929	8	3.117	17	2.561	86	1.821	8	-1.983	17	-1.748	86	-1.208
	Chr6	2569	3.528	-2.212	0.410	0.960	12	2.919	25	2.680	128	1.949	12	-1.925	25	-1.792	128	-1.198
	Chr7	2358	3.865	-2.371	0.137	0.903	11	2.674	23	2.374	117	1.698	11	-1.892	23	-1.734	117	-1.304
	Chr8	1884	3.505	-2.498	0.031	0.871	9	2.366	18	2.106	94	1.453	9	-2.155	18	-1.994	94	-1.428
	Total	19520	4.778	-2.683	0.240	0.938	97	2.767	195	2.502	976	1.780	97	-2.046	195	-1.842	975	-1.291

Table III-S7. Summary statistics of CLR test among the 8 chromosomes of wild and cultivated apricots.

Group	Chr.	Total signals	Max	Min	Mean	Std	Top 0.5%		Top 1%		Top 5%	
							Rank	Value	Rank	Value	Rank	Value
Three groups	All	58746	437.700	0.000	1.377	6.487	293	31.420	587	19.604	2937	5.362
Wild Central Asian	Chr1	4362	220.277	0.000	1.280	5.419	21	28.502	43	17.894	218	5.092
	Chr2	2554	416.418	0.000	2.273	12.928	12	52.281	25	34.491	127	8.691
	Chr3	2196	64.778	0.000	1.255	4.105	10	28.766	21	18.368	109	5.516
	Chr4	1906	124.478	0.000	1.473	5.844	9	47.987	19	20.630	95	5.977
	Chr5	1729	206.510	0.000	1.289	6.075	8	27.752	17	16.092	86	5.328
	Chr6	2581	76.222	0.000	1.166	3.648	12	24.530	25	17.839	129	5.182
	Chr7	2361	179.664	0.000	1.393	5.889	11	38.997	23	22.926	118	5.154
	Chr8	1893	105.151	0.000	1.303	4.747	9	24.691	18	19.766	94	5.853
	Total	19582	416.418	0.000	1.427	6.710	97	30.652	195	20.339	979	5.752
Cultivated Chinese	Chr1	4362	119.002	0.000	0.976	3.870	21	26.549	43	17.607	218	5.202
	Chr2	2554	131.973	0.000	1.772	7.021	12	53.649	25	32.032	127	7.210
	Chr3	2196	201.255	0.000	1.312	5.943	10	28.206	21	17.409	109	5.709
	Chr4	1906	86.106	0.000	1.415	5.668	9	34.801	19	18.370	95	6.647
	Chr5	1729	220.225	0.000	1.672	9.643	8	51.447	17	34.025	86	4.608
	Chr6	2581	85.544	0.000	1.062	3.905	12	26.649	25	17.305	129	4.510
	Chr7	2361	129.777	0.000	1.164	5.203	11	26.894	23	16.709	118	4.545
	Chr8	1893	67.639	0.000	1.219	4.108	9	30.219	18	19.182	94	5.492
	Total	19582	220.225	0.000	1.331	5.865	97	34.801	195	19.977	979	5.364
Cultivated European	Chr1	4362	119.002	0.000	0.976	3.870	21	18.896	43	11.995	218	3.794
	Chr2	2554	117.123	0.000	1.504	5.305	12	30.745	25	23.332	127	5.721
	Chr3	2196	437.700	0.000	1.802	11.378	10	42.953	21	23.610	109	6.215
	Chr4	1906	287.085	0.000	2.031	10.807	9	70.713	19	44.311	95	6.547
	Chr5	1729	76.962	0.000	1.360	4.444	8	30.871	17	17.198	86	5.442
	Chr6	2581	224.179	0.000	1.131	5.786	12	26.028	25	15.329	129	4.397
	Chr7	2361	65.631	0.000	1.093	3.516	11	20.798	23	15.553	118	4.518
	Chr8	1893	188.762	0.000	1.638	8.001	9	49.773	18	25.795	94	5.455
	Total	19582	437.700	0.000	1.373	6.843	97	29.979	195	18.699	979	5.050

ANNEXE 3

Table III-S8. Summary statistics of Omega test among the 8 chromosomes of wild and cultivated apricots.

Group	Chr.	Total signals	Max	Min	Mean	Std	Top 0.5%		Top 1%		Top 5%	
							Rank	Value	Rank	Value	Rank	Value
Three groups	All	57990	538.402	0.828	2.514	4.580	289	23.642	579	16.448	2899	5.891
Wild Central Asian	Chr1	4320	83.764	0.828	2.755	4.068	21	31.137	43	20.464	216	7.434
	Chr2	2504	81.107	1.038	3.278	4.975	12	35.147	25	28.629	125	9.718
	Chr3	2173	67.080	1.056	2.793	3.812	10	25.714	21	22.123	108	7.585
	Chr4	1880	77.099	1.008	3.199	4.425	9	33.258	18	24.283	94	9.780
	Chr5	1707	104.066	1.022	2.610	4.251	8	28.211	17	16.127	85	5.982
	Chr6	2550	90.192	1.061	2.706	3.987	12	27.355	25	17.418	127	6.539
	Chr7	2336	63.658	1.041	2.718	3.821	11	28.974	23	20.419	116	6.978
	Chr8	1868	89.955	1.063	2.715	3.973	9	22.881	18	17.228	93	6.815
	Total	19338	104.066	0.828	2.843	4.178	96	28.974	193	20.810	966	7.646
Cultivated Chinese	Chr1	4315	42.335	1.021	1.898	1.415	21	9.836	43	7.876	215	3.684
	Chr2	2505	104.508	1.007	2.233	3.440	12	17.418	25	10.076	125	4.927
	Chr3	2171	33.730	1.013	1.977	1.583	10	11.696	21	8.722	108	4.152
	Chr4	1879	27.589	0.945	2.072	1.697	9	14.202	18	10.010	93	4.235
	Chr5	1705	48.351	0.884	1.857	1.529	8	8.795	17	6.143	85	3.437
	Chr6	2551	37.414	1.033	1.942	1.648	12	11.064	25	8.219	127	3.700
	Chr7	2334	35.563	1.074	1.913	1.419	11	9.562	23	7.961	116	3.680
	Chr8	1868	29.463	1.018	1.934	1.347	9	9.085	18	7.788	93	3.843
	Total	19328	104.508	0.884	1.975	1.881	96	10.738	193	8.304	966	3.916
Cultivated European	Chr1	4309	126.830	0.960	2.489	3.818	21	22.113	43	16.225	215	6.070
	Chr2	2502	538.402	1.031	3.413	12.042	12	38.865	25	29.452	125	8.959
	Chr3	2173	60.081	1.039	2.634	3.750	10	26.687	21	18.938	108	6.912
	Chr4	1879	270.336	0.954	3.020	7.531	9	31.986	18	21.955	93	7.889
	Chr5	1708	302.516	1.041	2.694	8.476	8	27.965	17	16.371	85	5.853
	Chr6	2550	129.317	1.057	2.537	4.284	12	27.231	25	16.385	127	5.930
	Chr7	2335	123.162	1.056	2.668	4.412	11	36.754	23	20.288	116	6.398
	Chr8	1868	63.153	1.056	2.505	3.483	9	25.302	18	16.494	93	5.685
	Total	19324	538.402	0.954	2.724	6.442	96	28.700	193	19.304	966	6.555

Table III-S9. Summary statistics of Mu (μ) test among the 8 chromosomes of wild and cultivated apricots.

Group	Chr.	Total signals	Max	Min	Mean	Std	Top 0.5%		Top 1%		Top 5%	
							Rank	Value	Rank	Value	Rank	Value
Three groups	All	58324	1.146E-04	0.000E+00	1.690E-06	3.015E-06	291	1.958E-05	583	1.478E-05	2916	5.705E-06
Wild Central Asian	Chr1	4347	1.749E-05	3.306E-14	3.540E-07	9.642E-07	21	7.051E-06	43	4.770E-06	217	1.197E-06
	Chr2	2535	2.891E-05	4.135E-14	8.866E-07	2.023E-06	12	1.429E-05	25	9.939E-06	126	4.147E-06
	Chr3	2190	1.543E-05	7.664E-14	5.573E-07	1.038E-06	10	7.232E-06	21	5.635E-06	109	2.040E-06
	Chr4	1900	2.797E-05	6.332E-14	8.102E-07	1.711E-06	9	1.072E-05	19	8.833E-06	95	3.358E-06
	Chr5	1719	4.153E-05	8.252E-14	7.438E-07	1.958E-06	8	1.229E-05	17	8.606E-06	85	2.034E-06
	Chr6	2566	2.891E-05	0.000E+00	5.002E-07	1.359E-06	12	1.063E-05	25	5.664E-06	128	1.691E-06
	Chr7	2360	1.434E-05	4.687E-14	5.350E-07	9.910E-07	11	6.872E-06	23	5.317E-06	118	1.854E-06
	Chr8	1888	2.293E-05	6.016E-14	5.529E-07	1.215E-06	9	7.502E-06	18	6.206E-06	94	1.921E-06
	Total	19505	4.153E-05	0.000E+00	5.852E-07	1.421E-06	97	9.608E-06	195	7.033E-06	975	2.108E-06
Cultivated Chinese	Chr1	4323	3.968E-05	2.272E-09	1.343E-06	2.251E-06	21	1.806E-05	43	1.169E-05	216	3.881E-06
	Chr2	2522	6.254E-05	2.820E-13	3.019E-06	5.416E-06	12	4.045E-05	25	3.050E-05	126	1.222E-05
	Chr3	2183	3.732E-05	7.021E-09	2.480E-06	3.425E-06	10	2.620E-05	21	1.859E-05	109	8.036E-06
	Chr4	1889	6.081E-05	4.560E-13	3.121E-06	5.038E-06	9	4.101E-05	18	2.462E-05	94	1.109E-05
	Chr5	1711	1.146E-04	1.365E-07	3.010E-06	5.151E-06	8	3.448E-05	17	2.700E-05	85	8.162E-06
	Chr6	2550	2.572E-05	0.000E+00	1.933E-06	2.585E-06	12	1.893E-05	25	1.625E-05	127	6.055E-06
	Chr7	2349	3.590E-05	5.242E-08	2.000E-06	2.689E-06	11	2.040E-05	23	1.552E-05	117	5.823E-06
	Chr8	1889	2.738E-05	0.000E+00	2.534E-06	2.913E-06	9	2.027E-05	18	1.769E-05	94	7.776E-06
	Total	19416	1.146E-04	0.000E+00	2.281E-06	3.749E-06	97	2.507E-05	194	1.886E-05	970	7.259E-06
Cultivated European	Chr1	4323	4.065E-05	9.670E-14	1.391E-06	2.181E-06	21	1.372E-05	43	9.331E-06	216	3.922E-06
	Chr2	2512	3.291E-05	2.463E-13	2.445E-06	3.296E-06	12	2.463E-05	25	1.773E-05	125	8.160E-06
	Chr3	2178	3.493E-05	1.751E-13	2.125E-06	2.470E-06	10	1.490E-05	21	1.223E-05	108	6.260E-06
	Chr4	1894	6.289E-05	3.499E-13	3.207E-06	4.656E-06	9	3.150E-05	18	2.512E-05	94	9.278E-06
	Chr5	1717	5.451E-05	2.509E-13	3.293E-06	3.843E-06	8	2.688E-05	17	2.109E-05	85	8.584E-06
	Chr6	2550	4.026E-05	0.000E+00	2.051E-06	2.830E-06	12	1.934E-05	25	1.382E-05	127	6.101E-06
	Chr7	2344	2.481E-05	1.289E-13	1.856E-06	2.264E-06	11	1.570E-05	23	1.278E-05	117	5.635E-06
	Chr8	1885	3.833E-05	0.000E+00	2.540E-06	2.791E-06	9	1.801E-05	18	1.551E-05	94	7.295E-06
	Total	19403	6.289E-05	0.000E+00	2.210E-06	3.064E-06	97	2.061E-05	194	1.507E-05	970	6.658E-06

ANNEXE 3

Table III-S10. Summary statistics of F_{st} test among the 8 chromosomes of wild and cultivated apricots.

Group	Chr.	Total signals	Max	Min	Mean	Std	Top 0.5%		Top 1%		Top 5%	
							Rank	Value	Rank	Value	Rank	Value
Three pairwise	All	58573	0.463	0.000	0.072	0.040	292	0.224	585	0.201	2928	0.148
Wild Central Asian and cultivated Chinese	Chr1	4352	0.293	0.000	0.058	0.030	21	0.173	43	0.156	217	0.115
	Chr2	2546	0.229	0.000	0.048	0.026	12	0.165	25	0.132	127	0.097
	Chr3	2188	0.208	0.000	0.043	0.024	10	0.141	21	0.129	109	0.088
	Chr4	1901	0.257	0.000	0.051	0.025	9	0.156	19	0.135	95	0.096
	Chr5	1723	0.206	0.000	0.053	0.026	8	0.146	17	0.138	86	0.099
	Chr6	2569	0.236	0.000	0.047	0.027	12	0.158	25	0.142	128	0.098
	Chr7	2358	0.196	0.000	0.042	0.021	11	0.140	23	0.111	117	0.081
	Chr8	1885	0.193	0.000	0.041	0.022	9	0.136	18	0.121	94	0.082
	total	19522	0.293	0.000	0.049	0.027	97	0.156	195	0.138	976	0.099
Cultivated Chinese and cultivated European	Chr1	4353	0.324	0.000	0.078	0.034	21	0.201	43	0.185	217	0.139
	Chr2	2546	0.424	0.000	0.080	0.038	12	0.246	25	0.199	127	0.146
	Chr3	2189	0.255	0.000	0.081	0.036	10	0.227	21	0.202	109	0.149
	Chr4	1901	0.444	0.000	0.093	0.043	9	0.255	19	0.229	95	0.167
	Chr5	1723	0.306	0.000	0.080	0.040	8	0.239	17	0.209	86	0.151
	Chr6	2570	0.301	0.000	0.079	0.037	12	0.211	25	0.190	128	0.145
	Chr7	2358	0.314	0.000	0.085	0.039	11	0.232	23	0.207	117	0.160
	Chr8	1885	0.270	0.000	0.084	0.036	9	0.209	18	0.197	94	0.155
	total	19525	0.444	0.000	0.082	0.038	97	0.224	195	0.199	976	0.151
Wild Central Asia and cultivated European	Chr1	4353	0.335	0.000	0.088	0.043	21	0.249	43	0.224	217	0.171
	Chr2	2547	0.463	0.000	0.083	0.044	12	0.260	25	0.235	127	0.164
	Chr3	2188	0.443	0.000	0.076	0.037	10	0.211	21	0.195	109	0.145
	Chr4	1901	0.311	0.000	0.090	0.046	9	0.273	19	0.238	95	0.181
	Chr5	1723	0.394	0.000	0.093	0.048	8	0.297	17	0.251	86	0.181
	Chr6	2571	0.326	0.000	0.083	0.042	12	0.233	25	0.212	128	0.164
	Chr7	2358	0.329	0.000	0.086	0.042	11	0.252	23	0.220	117	0.165
	Chr8	1885	0.317	0.000	0.093	0.041	9	0.241	18	0.223	94	0.169
	total	19526	0.463	0.000	0.086	0.043	97	0.250	195	0.223	976	0.169

Table III-S11. Summary statistics of hapFLK test among the 8 chromosomes of wild and cultivated apricots.

Group	Chr.	Total signals	Max	Min	Mean	Std	Top 0.5%		Top 1%		Top 5%	
							Rank	Value	Rank	Value	Rank	Value
p<0.05 (FLK test)	Chr1	41630	6.488	0.013	2.224	0.986	208	5.187	416	4.770	2081	3.910
	Chr2	28248	6.752	0.013	2.221	0.964	141	5.448	282	4.989	1412	4.119
	Chr3	26124	8.088	0.040	2.381	1.069	130	5.352	261	5.106	1306	4.168
	Chr4	25970	6.870	0.042	2.272	1.053	129	5.204	259	4.911	1298	4.105
	Chr5	15720	7.925	0.036	2.389	1.070	78	5.578	157	5.189	786	4.211
	Chr6	27784	7.341	0.025	2.406	1.092	138	5.703	277	5.277	1389	4.317
	Chr7	28994	7.867	0.007	2.381	1.063	144	5.294	289	4.966	1449	4.152
	Chr8	20119	6.415	0.036	2.196	1.009	100	5.436	201	4.999	1005	4.075
	total	214589	8.088	0.007	2.303	1.038	1072	5.346	2145	4.988	10729	4.089
p<0.01 (FLK test)	Chr1	8645	6.488	0.016	2.252	1.038	43	5.565	86	5.187	432	4.145
	Chr2	8853	6.752	0.041	2.240	0.988	44	5.206	88	4.820	442	3.966
	Chr3	5572	8.088	0.041	2.638	1.242	27	6.457	55	5.562	278	4.766
	Chr4	9855	6.871	0.088	2.506	1.120	49	5.680	98	5.191	492	4.505
	Chr5	3849	7.925	0.046	2.616	1.177	19	6.560	38	5.986	192	4.553
	Chr6	6506	7.265	0.054	2.518	1.221	32	6.203	65	5.645	325	4.794
	Chr7	10095	7.867	0.039	2.540	1.113	50	5.822	100	5.275	504	4.439
	Chr8	5869	6.197	0.039	2.282	1.069	29	5.641	58	5.182	293	4.354
	total	59244	8.088	0.016	2.434	1.121	296	5.780	592	5.338	2962	4.439

Table III-S12. Details of selective sweeps detected from multiple tests across the 8 apricot chromosomes (multiple sheets). INRA data portal: <https://doi.org/10.15454/FFKGUP>.**Table III-S13. List and position of the 1,753 selective sweeps identified over the 8 apricot chromosomes.** In green, the ones identified in European cultivated apricots; in red, the ones for Chinese cultivated apricots and in orange, the ones identified in Central Asian wild apricot genomes. INRA data portal: <https://doi.org/10.15454/FFKGUP>.

Titre : Histoire évolutive et impact des différents processus évolutifs sur la diversité génétique de l'abricotier (*Prunus armeniaca* L.)

Résumé:

L'abricotier cultivé (*Prunus armeniaca* L.) appartient au genre *Prunus*, de la sous-famille des *Prunoideae* qui comprend la totalité des arbres fruitiers à noyaux de la famille des *Rosaceae*. Il fait partie de la section taxonomique *Armeniaca* (Lam.) Koch. qui se présente comme un complexe d'espèces diploïdes, inter-fertiles avec un génome d'environ 200-220 Mbp ($n=8$). La section *Armeniaca* comprend deux espèces cultivées, *P. armeniaca* (fruitière) et *P. mume* (ornementale) ; mais également cinq espèces encore disponibles à l'état sauvage en Asie Centrale et en Asie du Nord-Est, le plus souvent en altitude. Dans ce contexte, mon travail de thèse vise à mieux comprendre les différents processus de l'histoire évolutive d'une espèce fruitière pérenne et comment ceux-ci influent sur la variabilité et la structuration génétique de l'espèce cultivée. Ceci inclut son adaptation à de multiples et changeantes conditions environnementales mais également à l'action de l'Homme au travers de la domestication, de la sélection et de l'amélioration génétique et son effet sur l'architecture génomique de l'abricotier.

Dans un premier temps, des études de diversité réalisées à l'aide de marqueurs moléculaires de type microsatellites ont été réalisées chez l'abricotier et ses espèces apparentées, sauvages, afin de clarifier les généalogies et révéler les processus évolutifs qui sont à l'origine de la forme cultivée, fruitière. Notre étude de phylogéographie nous a permis de détecter des groupes génétiques différenciés résultant de l'histoire climatique passée de la planète mais également d'hybridation interspécifique et de flux de gènes récurrent entre individus sauvages et domestiques. Plusieurs événements indépendants de domestication ont ainsi été mis en évidence, ils sont à l'origine de l'abricotier cultivé en Occident, en Chine et en Asie Centrale.

La même approche a été utilisée dans un second temps afin de décrire la diversité et la structuration génétique de *P. brigantina* Vill., la seule espèce européenne de la section *Armeniaca*, ce qui nous a conduit à préciser sa classification dans le genre *Prunus*.

Enfin dans la troisième partie de cette thèse, la diversité génétique a cette fois été étudiée à l'échelle du génome complet de l'abricotier. L'objectif ici était de rechercher les régions génomiques permettant de différencier les groupes domestiques, européens et chinois, des populations sauvages d'Asie Centrale. Ces zones de forte différenciation dans les génomes correspondent à des signatures de balayages sélectifs. Nous avons ainsi identifié plus de 1700 régions génomiques comme cibles probables de l'adaptation et de la domestication de l'abricotier, pour lesquelles 136 présentaient un fort degré de similarité pour tous les cultivars d'abricotiers indiquant 56 régions génomiques de domestication homologues, non-chevauchantes. Pour 48 de ces régions, nous disposons d'annotations fonctionnelles qui permettent de déterminer les gènes sous sélection et leur fonction. Il apparaît que la plupart de ces gènes sont connus pour affecter l'expression de phénotypes liés 1) à la réponse aux pathogènes et au stress abiotique, 2) à la qualité du fruit ainsi qu'au 3) contrôle moléculaire de la floraison et de la transition entre période végétative et reproductive. Ce résultat constitue un premier pas vers la compréhension des mécanismes responsables du processus de domestication chez une espèce fruitière, pérenne. Il montre que des événements de domestication indépendants ont impliqué des régions génomiques homologues. Les travaux à venir devront également permettre de préciser les cibles génétiques des processus adaptatifs chez cette espèce fruitière, pérenne, et de fournir des cibles pour les programmes d'amélioration génétique de l'abricotier dans un contexte de changements climatiques.

Mots clés : Abricotier, diversité, évolution, population, domestication, adaptation

Title: Evolutionary history of apricot (*Prunus armeniaca* L.) and impact of different processes of evolution on genetic diversity**Abstract:**

Nowadays, increasing attention is focused on perennial crop species and their wild relatives. The domestication of perennials is expected to follow different processes than annuals, and there is limited knowledge about how perennial plant species evolve in response to human intervention or changing environmental conditions. Indeed, the diversity of perennial species results from a series of mechanisms of evolution, which include natural and artificial selection, gene flow between wild and cultivated compartments, and dynamics of dispersion at large scales, often over long periods. Unraveling the evolutionary history and domestication processes of long-lived tree species is expected to provide insights into the potential differences and similarities between annual and perennial species, and furtherly to facilitate breeding efforts for traits of interest.

In the current PhD thesis, we focused on apricot species, *Prunus armeniaca* L., and its related species from the section *Armeniaca* (Lam.) Koch.. We characterized genetic diversity and variability and addressed a few important questions related to fruit tree origin, evolution and domestication, and further identified candidate genes and loci underlying important agronomic traits that have been under selection during domestication.

Our microsatellite data and approximate Bayesian computation revealed that the wild species *P. armeniaca* and *P. sibirica* diverged ca. 8 to 16 Mya ago, followed by interspecific hybridization leading to a new, isolated species, in Western China. We also showed that the European and Chinese apricots were domesticated independently either both from the Central Asian wild progenitor or from the hybrid species.

Following the same strategy, we studied the genetic diversity and structuration of the only European *Armeniaca* species, *P. brigantina* Vill. and thus questioned its classification among the genus *Prunus*.

Finally, taking advantage of the *de novo* assembly of a high-quality apricot reference genome and of extensive resequencing data, we focused on how selection has influenced genomic architecture in apricot (*P. armeniaca*). To test for common or distinct signatures of selection, we took advantage of the parallel history of domestication in the European and Chinese apricots and compared with their wild, Central Asian progenitor. We detected evidence for artificial selection at a genome-wide scale, both for European and Chinese apricots, with a significant number of homologous genomic signatures of domestication, thus indicating convergent yet independent selection of a common set of genes during two geographically and culturally distinct domestication processes. We also identified signatures of selection which could be associated with local adaptation in either wild or cultivated apricots.

Therefore, a better knowledge on apricot evolutionary history combined with comparative population genomics enables the identification and utilization of adaptive and domestication traits that are important for apricot cultivation. It is expected to provide an unprecedented opportunity to identify the genetic basis of long-lived perennials' adaptation and domestication.

Keywords : Apricot, diversity, evolution, population, domestication, adaptation