



HAL
open science

Reconstruction des scénarios évolutifs d'introduction et d'expansion en Europe des psylles du complexe *Cacopsylla pruni*

Margaux Darnis

► **To cite this version:**

Margaux Darnis. Reconstruction des scénarios évolutifs d'introduction et d'expansion en Europe des psylles du complexe *Cacopsylla pruni*. *Systématique, phylogénie et taxonomie*. 2022. hal-03772001

HAL Id: hal-03772001

<https://hal.inrae.fr/hal-03772001>

Submitted on 7 Sep 2022

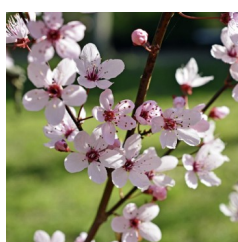
HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

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.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Reconstruction des scénarios évolutifs d'introduction et d'expansion en Europe des psylles du complexe *Cacopsylla pruni*



Plant
Health
Institute
Montpellier

INRAE

Margaux Darnis

M2 Biodiversité, Ecologie, Evolution (BEE)

Parcours Ecologie Evolutive et Fonctionnelle (EEF)

Menu Ecologie Evolutive (Evol)

Encadrant : Nicolas Sauvion

(Ingénieur de Recherche INRAE – Entomologiste – Epidémiologiste)

Structure d'accueil : UMR PHIM (Plant Health Institute of Montpellier)

Pôle PRISM, équipe Forisk

Projet ANR BEYOND

Avenue du campus d'Agropolis – 34980 Montferrier sur lez - France

6 mois : du 17/01/2022 au 17/07/2022

date de soutenance : 23/06/2022

Année 2021-2022

Remerciements :

Je tiens tout d'abord à remercier mon maître de stage Nicolas SAUVION pour avoir suivi mon travail et m'avoir apporté son soutien et conseillé tout en restant bienveillant et disponible tout au long de ce stage. Je souhaiterais également le remercier pour m'avoir fait découvrir les joies de la chasse aux psylles et du travail dans le cadre d'un ANR, dont les plénières et les réunions étaient toujours particulièrement intéressantes.

Je remercie également Virginie RAVIGNE pour ces nombreux et précieux conseils regardant les aspects les plus techniques de génétique des populations, qui m'ont été d'une grande utilité tout au long du stage.

J'aimerais également remercier tout les membres de l'UMR PHIM et plus particulièrement ceux du pôle PRISM pour leur accueil chaleureux et dont les réunions et discussions ont rythmé mes semaines depuis janvier.

Je remercie aussi les doctorants et les autres stagiaires du laboratoire pour avoir rendu le bureau des stagiaires plus convivial ainsi que pour leur soutien et les rires partagés pendant ces quelques mois.

Et pour finir, je tiens à remercier mes camarades de promotion pour tout nos échanges durant nos stages respectifs, liés à nos inquiétudes vis à vis de nos différents plannings et à nos recherches de thèses, qui ont été d'un grand soutien et m'ont apporté beaucoup de légèreté durant les périodes de travail parfois intense.

Résumé : Cette étude avait pour but de décrire la structure des populations et d'inférer des scénarios phylogéographiques des psylles du complexe *Cacopsylla pruni* à l'échelle du paléarctique ouest. Dans ce but, nous avons conduit diverses analyses de génétique des populations (reconstruction de réseaux d'haplotypes, analyses multivariées, clustering bayésien, analyses phylogénétiques, calcul d'indicateurs statistiques, etc) ainsi que des analyses ABC (Approximate Bayesian computation). Plusieurs types de marqueurs ont été utilisés : des séquences d'un gène mitochondrial (COI-t) et d'un gène nucléaire (ITS2), et 8 marqueurs microsatellites. Nous avons confirmé que le complexe était composé de deux et uniquement deux espèces cryptiques (A & B) et nous avons décrit la structuration génétique et spatiale des populations de *C. pruni*. Ces deux espèces auraient divergé d'un ancêtre commun il y a 40-50 Ma en France. L'espèce A aurait ensuite divergé en Espagne avant que l'espèce B ne diverge vers l'Est. Les populations françaises de ces deux groupes ont également vécu des expansions démographiques après leur divergence avant que les populations ne rétrécissent à des temps plus récent (entre 3-10 Ma et 16.000 ans selon l'espèce). Notre étude a permis de poser de nouvelles hypothèses sur l'histoire évolutive des deux espèces du complexe *C. pruni*. Nous proposons des perspectives de travail pour préciser nos résultats.

Abstract : The aim of this study was to describe populations structure and investigate the phylogeography of the psyllid cryptic species complex, *Cacopsylla pruni*, at the west paleartic scale. Therefore, we conducted several population genetic analysis (haplotype networks, discriminant analysis of principal components, bayesian clustering, phylogenetic analysis, statistics computations, etc) and an Approximate Bayesian Computation(ABC) analysis. We used several types of markers such as sequences of mitochondrial (COI-t) and nuclear (ITS2) genes, and 8 microsatellite markers. We confirmed that *C. pruni* is a complex of two, and only two, cryptic species (A&B), and we unveiled the spatial and genetic structure of their populations. The two species would have diverged from a common ancestor 40-50 My ago in what we call France today. Then, the A species would have diverged in Spain before the B species diverged toward the East. The french populations of both species have also experienced demographic expansions after their divergence before the populations shrank at more recent times (between 3-10 My and 16.000 years ago depending on the species). Our study allowed to propose new hypothesis on the evolutionary history of the two species of the *C. pruni* complex. In the discussion, we propose workprospects to confirm our results.

Sommaire :

I – Introduction	2
II - Matériel et méthodes	4
1 – <i>Echantillonnage et marqueurs génétiques</i>	4
2 - <i>Analyses de la structuration génétique et spatiale de C. pruni</i>	5
1.A – <i>Analyses des séquences COI-t et ITS2 et réseaux d'haplotypes</i>	5
1.B – <i>Analyses des marqueurs microsatellites</i>	6
1 – <i>Analyses multivariée</i>	6
2 – <i>Clustering Bayésien</i>	7
3 – <i>Calculs d'indicateurs statistiques et tests</i>	8
3 – <i>Reconstruction des scénarios évolutifs : analyses ABC (Approximate Bayesian Computation)</i>	8
4 - <i>Analyses phylogénétiques</i>	10
III – Résultats	10
1 - <i>Analyses de la structuration génétique et spatiale de C. pruni</i>	10
1.A – <i>Analyses des séquences COI-t et ITS2 et réseaux d'haplotypes</i>	10
1.B – <i>Analyses des marqueurs microsatellites</i>	13
1 – <i>Analyses multivariée</i>	13
2 – <i>Clustering Bayésien</i>	14
3 – <i>Calculs d'indicateurs statistiques et tests</i>	17
2 – <i>Reconstruction des scénarios évolutifs : analyses ABC (Approximate Bayesian Computation)</i>	18
3 - <i>Analyses phylogénétiques</i>	20
IV - Discussion	23
Bibliographie	28
Résumé / Abstract	31
Annexes	32 à 44

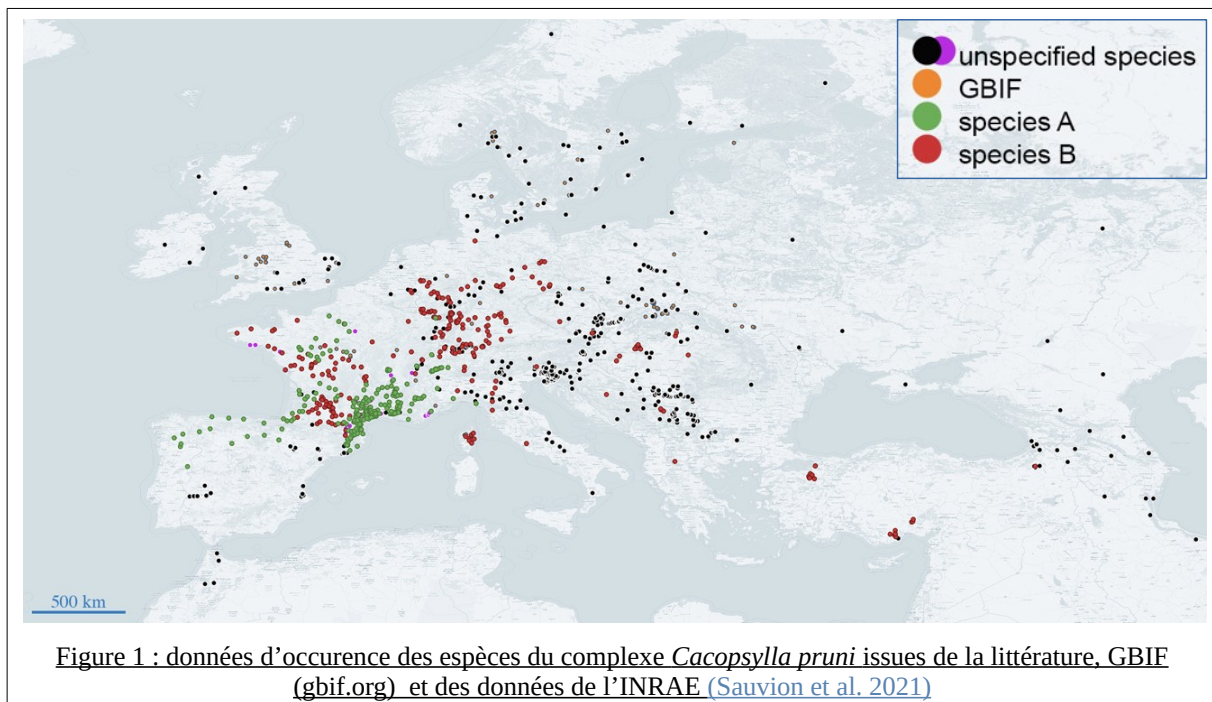
I – Introduction

La phylogéographie est une discipline grandissante qui, à l'origine, a été développée pour explorer les processus sous-jacents à la distribution spatiale de la diversité et des lignées génétiques au sein et entre les espèces (Avisé 2009; Avisé et al., 1987). Depuis l'invention du mot en 1987 (Avisé et al., 1987), il a été mis en évidence que les grands événements biogéographiques tels que les oscillations climatiques ou la modification de l'habitat au cours du temps sont des facteurs clefs jouant sur les variations géographiques au sein des espèces (Hewitt 2000; 2001). Mais pour les insectes ravageurs ou vecteurs, l'activité humaine, en particulier l'agriculture, et les déplacements humains influencent également la distribution et la démographie des populations et ont grandement accéléré leur expansion (Liebhold et al. 2008). Ces facteurs peuvent en effet remodeler la structure génétique des populations et ajouter de nouvelles signatures génétiques sur des patterns phylogéographiques naturels existants (Stone et al. 2007). Aussi, une gestion efficace des insectes ravageurs ou vecteurs implique une bonne connaissance de leur structure phylogéographique et de la structure génétique des populations, notamment pour caractériser les échelles spatiales auxquelles la lutte doit être raisonnée (Roderick 1996).

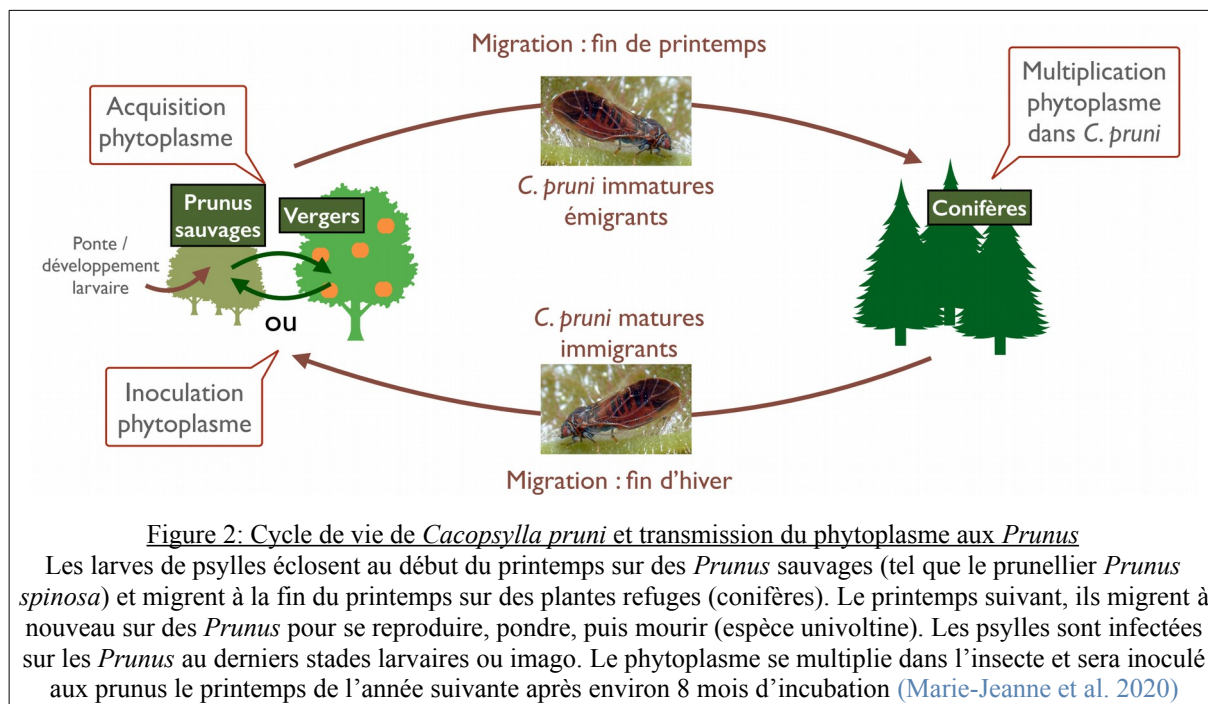
Pour résoudre des scénarios aux patterns phylogéographiques complexes, la méthode ABC (Approximate Bayesian Computation) (Sunnåker et al. 2013; Csilléry et al. 2010), liée à une bonne connaissance de la génétique des populations étudiées, est une méthode de plus en plus utilisée. Ainsi, elle a déjà été mise en œuvre pour la reconstruction de scénarios d'ordre et de date de divergence et de migrations (Fresia et al. 2013; Ali et al. 2014), de scénarios d'invasions (Ascunce et al. 2011; Barrès et al. 2012; Mallez et al., 2021; Zepeda-Paulo et al. 2010), de scénarios de refuge durant les glaciations du quaternaire et de recolonisations post-glaciations (Song et al. 2018; Bidegaray-Batista et al. 2016; Sim et al. 2016), ou encore de scénarios de colonisation de divers habitats (Estoup et al. 2001; 2004; Thornton et al. 2006; Pascual et al. 2007).

Dans le cadre de ce stage, je me suis intéressée aux psylles du complexe *Cacopsylla* (*Thamnopsylla*) *pruni* (Scopoli, 1763) [Hemiptera : Psylloidea]. Ces insectes sont connus pour transmettre une maladie (ESFY : European stone fruit yellows) aux arbres fruitiers du genre *Prunus*, due à une phytoplasme (bactérie sans paroi) nommé '*Candidatus* Phytoplasma prunorum' (Danet et al. 2011; Lee et al. 1998; Carraro et al. 1998; Marie-Jeanne et al. 2020).

En 1998, l'espèce *C. pruni* avait été décrite comme vectrice du phytoplasme. Depuis, des travaux réalisés au sein de mon équipe d'accueil ont montré que *C. pruni* serait en fait un complexe de deux espèces cryptiques, des espèces indistinguables ou quasi indistinguables sur la base de critères morphologiques mais qui présentent une grande différenciation génétique et qui sont sur des trajectoires évolutives différentes. Il a été démontré que (Peccoud et al. 2013; Peccoud et al. 2018; Marie-Jeanne et al. 2020; Sauvion et al. 2021) : (1) Les deux espèces sont présentes en sympatrie ou en allopatrie stricte selon les zones géographiques et endémique au milieu sauvage du paléarctique ouest (Fig. 1); (2) L'isolement reproducteur entre les deux groupes est très fort : *in natura*, des accouplements inter-espèces ont été observés mais les hybrides sont très rares (<1/1000); (3) Les deux espèces, que nous appellerons ici A et B, ont le même cycle biologique, en particulier dans les zones où elles sont en sympatrie, et les deux espèces sont vectrices du phytoplasme (Fig. 2).



Le but de mon stage était, dans un premier temps, de confirmer la séparation de *C. pruni* en deux espèces cryptiques, et uniquement deux, via une approche de taxonomie intégrative (Dayrat 2005). Pour cela, différents types de marqueurs génétiques et différentes approches d'analyses génétiques ont été utilisées à l'aide d'échantillons collectés dans le paléarctique ouest pour décrire la diversité génétique et la structure génétique et spatiale des populations de *C. pruni*. Ces résultats acquis, il a été possible de reconstruire les scénarios évolutifs d'introduction et d'expansion des deux espèces à l'échelle européenne.



II - Matériel et méthodes

1 – Echantillons et marqueurs génétiques

Au total, 1650 psylles venant de 97 sites d'échantillonnage différents répartis sur presque toute l'aire de répartition connue de *C. pruni* ont été analysées (voir Annexe 1 & 2). Trois marqueurs génétiques distincts ont été utilisés :

- 544 séquences d'ADN mitochondrial de *C. pruni* (fragment de 621 bp de la région mitochondriale COI-t-tRNA-COII (= COI-t-t)) ont été obtenues avec les amorces UEA9 et C2-N-3389 comme décrit dans (Percy 2003). Cette région mitochondriale a été choisie car elle est polymorphe et donc adaptée à des analyses de niveau inférieur comme la phylogénétique d'espèces étroitement apparentées. De plus, il était ainsi envisageable de réutiliser les séquences de Percy (2003) disponibles dans GeneBank pour nos analyses.

- 783 séquences d'ADN nucléaire de *C. pruni*, (fragment de 753 pb de l'espaceur interne transcrit 2 (ITS2) - une portion de l'ADN ribosomique située entre des parties des gènes 28S et 5.8S- ont été amplifiées avec les amorces CAS5p8sFcm et CAS28sB1d (Peccoud et al. 2013).

Les séquençages (type Sanger) ont été réalisés par le centre de séquençage français Genoscope. Les séquences ont été analysées avec Geneious Pro (10.2.6) (<https://www.geneious.com>) antérieurement à mon stage en suivant la procédure suivante :

pour chaque individu, les séquences forward et reverse ont été alignées, comparées, et assemblées pour obtenir une séquence consensus. Les séquences d'ADN mitochondrial ont fait l'objet d'une attention particulière. En effet, les copies nucléaires des gènes mitochondriaux (numts) peuvent générer des résultats douteux (Hurst et al. 2005; Song et al. 2008). Ainsi, tous les doubles pics ont été visualisés pour s'assurer que les haplotypes étaient des gènes codants fonctionnels (absence d'indels ou de codons stop).

• 758 échantillons ont été séquencés avec huit des neuf loci microsatellites décrits dans Sauvion et al. (2009) ont été utilisés (Cp6_115, Cp5_45, Cp6_144, Cp4_127, Cp5_43, Cp6_15, Cp6_129, Cp4_108). Les microsatellites étant très polymorphiques, ils permettront des analyses au niveau infra-spécifique. Les produits PCR ont été calibrés sur une machine ABI Prism 3130XL à 16 capillaires (Applied Biosystems, Foster City, CA, USA) de la plateforme Génomique Environnementale du LabEx CeMEB (<http://www.labex-cemeb.org/en/genotyping-sequencing-genseq>). Le logiciel GeneMapper™ 5 (Applied Biosystems) a été utilisé pour l'assignation des allèles.

Enfin, les individus ont été assignés aux groupes A et B par PCR diagnostique selon le protocole décrit par Peccoud, et al. (2013).

2 - Analyses de la structuration génétique et spatiale de *C. pruni*

Nous avons choisis quatre approches complémentaires pour les analyses de génétique des populations de *C. pruni*. Nous avons ainsi utilisé des méthodes s'appliquant sur différents types de marqueurs, des méthodes de clustering avec et sans a priori sur le nombre de clusters, des méthodes avec et sans prise en compte de la spatialisation des données et enfin des méthodes prenant en compte l'admixture. Notre démarche s'apparente à une de validation croisée par comparaison des résultats obtenus par les différentes approches et s'inscrit parfaitement dans l'approche de taxonomie intégrative que nous avons cherché à suivre dans cette étude (Dayrat 2005).

2.A – Analyses des séquences COI-t et ITS2 et réseaux d'haplotypes

Dans un premier temps, des analyses des séquences COI-t et ITS2 ont été réalisées avec différents logiciels. Le nombre de séquences, de sites, de sites polymorphiques (sites de ségrégation), de mutations totales, d'haplotypes, ainsi que la diversité haplotypique, nucléotidique et le nombre de sites informatifs, et le D de Tajima (statistique résumée permettant de détecter les excès d'allèle en faible/forte fréquence ou en fréquence

intermédiaire et qui permet de détecter des effets démographiques) ont été calculés avec le logiciel DnaSP (v6.12.3) (Rozas et al. 2017). Des réseaux d'haplotypes des deux gènes ont été recréés avec la méthode de Median Joining Network du logiciel PopART (Leigh et al. 2015), une méthode appropriée pour la reconstruction de réseaux de données intraspécifiques (Bandelt et al. 1999; Woolley et al. 2008). PopArt a également permis de visualiser sous forme de diagramme circulaire les proportions d'individus des différentes origines géographiques pour chaque haplotype, et de calculer la longueur des séquences, le nombre et pourcentage de sites identiques et le pourcentage d'identité par paire des deux gènes. Toutes ces analyses ont été réalisées sur la population totale (A+B), puis indépendamment sur les groupes A et B, pour les gènes COI-t et ITS2.

Enfin, nous avons calculé le nombre d'haplotypes privés (haplotypes ne se trouvant que dans une zone géographique donnée) de A et de B pour chaque gène en utilisant une approche de raréfaction avec le logiciel *HP-rare* (Kalinowski 2005). La raréfaction est une approche utilisée initialement pour calculer des richesses alléliques (simples et privées) pour des échantillons de tailles inégales. En effet, il a été montré que la taille d'un échantillon affecte la probabilité d'échantillonner des allèles dans celui-ci : on a plus de chance de trouver un grand nombre d'allèles dans un grand échantillon que dans un petit. Ainsi, la raréfaction permet de résoudre ce problème et d'obtenir des statistiques comparables non biaisées par la taille des échantillons (Kalinowski 2004; Sjöstrand et al. 2014).

2.B – Analyses des marqueurs microsatellites

Une analyse des marqueurs microsatellites a ensuite été réalisée en suivant trois approches distinctes et complémentaires.

1 – Analyses multivariées

Dans un premier temps, pour explorer la structure des populations et identifier les clusters génétiques ainsi que leur nombre optimal sans à priori, une analyse DAPC (Discriminant Analysis of Principal Components) a été réalisée à l'aide de la librairie *adegenet* (procédure *k.means*, fonction *find.clusters*) dans R (version 4.1.2) (Jombart 2008; Jombart et al., 2015). La DAPC est une approche de k-means clustering (approche qui permet de déterminer le nombre k de groupes qui maximisent la variation entre les groupes) permettant d'explorer la structure des populations en se basant sur les analyses discriminantes et les analyses en composantes principales sans avoir à poser l'hypothèse de la panmixie mais surtout sans à priori sur le nombre de clusters dans les populations. C'est donc une très bonne

alternative aux méthodes de clustering bayésien que nous détaillerons ci-après (Pritchard et al. 2000). Pour éviter la perte d'information, toutes les composantes principales possibles (ici 16) ont été retenues et les différentes solutions de clusters ont été comparées en utilisant le critère BIC (Bayesian Information Criterion) (Jombart et al., 2015). Ces analyses ont été réalisées pour le groupe A, B et la population totale (A+B).

2 – Clustering Bayésien

Deux approches dites de clustering bayésien ont été mises en œuvre. Ces approches estiment les paramètres du modèle par un algorithme MCMC (Monte Carlo Markov Chain). Elles impliquent qu'un nombre de clusters maximum K_{\max} soit renseigné à priori.

Le Logiciel STRUCTURE (v2.3.3) (Pritchard et al. 2000) est extrêmement populaire en génétique des populations et bien adapté pour des populations éloignées géographiquement comme dans notre cas (Hausdorf et al. 2010). Dans un premier temps, nous l'avons donc utilisé pour réaliser des analyses sur la population totale (A+B), afin de voir comment le logiciel assigne les individus aux groupes A et B, et comparer avec les assignations par PCR diagnostiques déjà connues. Nous avons testés plusieurs options : avec ou sans admixture, et avec ou sans corrélation des fréquences alléliques. Une admixture n'était pas exclue puisque nous savions que les deux groupes A et B sont retrouvés en sympatrie dans certaines localités. Les analyses ont ensuite été effectuées pour les groupes A et B séparément. Pour celles-ci, nous avons choisi un modèle avec admixture car ce type de modèle est robuste à l'absence d'admixture dans l'échantillon, alors que les modèles sans admixture, à l'inverse, ne sont pas robustes à l'inclusion d'individus admixés dans l'échantillon (François et al. 2010). Des simulations LOCPRIOR ont également été utilisées, pour prendre en compte la population dans laquelle ont été échantillonnés les individus. Afin de déterminer le nombre k_{\max} de clusters, nous avons utilisé la méthode ΔK de Evanno (Evanno et al. 2005) implémentée dans *Structure harvester* (Earl et al. 2012).

La version récente de STRUCTURE a une option pour prendre en compte des informations spatiales sur les échantillons. Cependant, nous avons préféré un autre logiciel, TESS (v2.3) (Chen et al. 2007; Durand et al. 2009), qui a été spécifiquement développé pour prendre en compte des données spatialisées. Ces analyses ont été répétées avec des modèles avec et sans admixture ainsi qu'avec les deux modèles avec admixture proposés par TESS, le modèle CAR (Conditionnal Auto Regressive), inhérent à TESS, et le modèle BYM, décrit dans Durand et al. (2009). Pour chaque run, le critère DIC (Deviance Information Criterion)

(Durand et al. 2009; Durand et al. 2009), une mesure statistique des capacités de prédiction du modèle, a été estimée pour choisir le meilleur des modèles testés ainsi que le meilleur run associé au modèle retenu. Les paramètres d'entrée des différentes analyses (burnin, nombre d'itérations, de K_{max} , runs...) sont explicités dans l'Annexe 3.

3 – Calculs d'indicateurs statistiques et tests

Pour compléter les analyses précédentes, plusieurs tests statistiques ont été conduits avec le logiciel Genepop (version 4.7.5) (Rousset 2008, 2020). Ainsi, des tests exacts de l'équilibre de Hardy-Weinberg (permettant d'obtenir la probabilité de respecter l'équilibre de HW à chaque locus pour chaque population) (Hardy 1908; Weinberg 1908), des tests de Mantel d'isolement par la distance (IBD) entre les populations (Wright 1938, 1940, 1942), et le calcul des Rho-statistiques ont été réalisés. Les Rho-statistiques sont des statistiques analogues aux F-statistiques et s'interprètent de la même manière (Wright 1930; Nei et al. 1983) mais sont plus adaptées pour les séquences microsatellites car elles sont basées sur la taille des allèles (le nombre de répétitions) (Rousset 2020). Elles estiment la déviation des proportions d'homozygotes et d'hétérozygotes attendues sous l'équilibre de HW dans les sous-populations (Rhois), dans la population totale (Rhoit) et la différenciation génétique entre les sous-populations (Rhost). Une valeur de Rhois négative indique un excès d'hétérozygotes alors qu'une valeur positive correspond à un déficit d'hétérozygotes ($Rhois = 0 \Leftrightarrow$ panmixie). Ces analyses ont été réalisées sur la population totale (A+B) puis sur les groupes A et B séparément.

Enfin, la diversité génétique des groupes A et B a été calculée via le calcul de la richesse allélique privée de chacun des groupes par une approche de raréfaction, méthode décrite précédemment (Kalinowski 2004), à l'aide du logiciel *HP-rare* (Kalinowski 2005; 2004).

3 – Reconstruction des scénarios évolutifs : analyses ABC

Pour explorer les scénarios évolutifs possibles à l'origine du complexe d'espèces *C. pruni*, une analyse ABC (Approximate Bayesian Computation) a été réalisée sur les données microsatellites et ITS2 pour les données de France, et COI-t pour la comparaison des données de France et hors France (les séquences COI-t étant les seules disponibles pour les populations hors de France les plus éloignées). Le logiciel DIYABC (version 2.1) (Cornuet et al. 2008; 2014) a été utilisé pour comparer différentes hypothèses en compétition vis à vis de la démographie et de la divergence des populations de *C. pruni* à l'échelle du paléarctique-

ouest. Le logiciel simule un grand nombre de jeux de données en tirant différentes valeurs de paramètres des distributions des priors (paramètres à priori entrés par l'utilisateur) sous différents scénarios entrés dans le logiciel par l'utilisateur. Le logiciel estime ensuite la probabilité à posteriori des différents scénarios en comparant les jeux de données simulés et observés sur les statistiques résumées choisies (Cornuet et al. 2008).

Dans notre cas, les différents scénarios comparés dans l'analyse ABC ont été construits sur la base des résultats des analyses de génétique des populations décrites plus haut. Les scénarios ont été élaborés pour tester la date et le mode de divergence de *C. pruni* en deux espèces (A et B), tester les scénarios démographiques expliquant cette divergence et tester le sens de l'expansion des deux espèces sur le continent européen. Nos analyses ont portées sur 39 populations pour lesquelles nous avons suffisamment de données pour les différents marqueurs utilisées (chaque population (sympatrie) ou couple de populations (allopatrie) a été considéré(e) comme un réplica pour les différentes analyses).

Les scénarios ont été testés de façon hiérarchique : dans un premier temps, 7 scénarios ne faisant varier que la date de divergence de *C. pruni* en deux espèces A et B ont été comparés afin de déterminer la plus probable (1). Puis, 3 scénarios sur le mode de divergence ont été testés en utilisant la date inférée dans l'étape 1 pour déterminer si les espèces A et B venaient d'un ancêtre commun ou si l'une venait de l'autre (2). En parallèle, 10 scénarios démographiques post-divergence ont été testés sur les différentes populations A et B françaises disponibles (population seule) en utilisant la date inférée dans la première analyse (3). Puis, 6 autres scénarios démographiques (sur population seule) ont été testés pour préciser les scénarios démographiques plus récents en se basant sur les informations obtenus en 3 (4). Enfin, 4 scénarios finaux ont été testés sur les différentes populations en se basant sur les informations inférées par tous les tests précédents afin d'évaluer la zone géographique d'origine et le sens d'expansion de *C. pruni* dans le paléarctique-ouest (5) (voir Annexe 5).

Pour toutes ces analyses, un total de 12 à 62 statistiques résumées ont été générées (Annexe 4) pour comparer les jeux de données observés et inférés ainsi que les scénarios entre eux, selon le nombre de populations et de marqueurs utilisés dans l'analyse. Les priors utilisés dans les scénarios ainsi que leurs distributions sont résumés dans l'Annexe 6.

4 – Analyses phylogénétiques

Des phylogénies ont été obtenues à partir des 544 séquences du gène COI-t et des 783 séquences du gène ITS2 (une pour les séquences A et une pour les séquences B des deux gènes). Les arbres phylogénétiques ont été réalisés en utilisant la méthode PhyML (Guindon et al. 2003; 2010), basée sur une approche de maximum de vraisemblance, implémentée dans le service web NGPhylogeny.fr (<https://ngphylogeny.fr/>). Cette plateforme de services est en libre accès. Elle permet, à partir d'un fichier FASTA, l'exécution en ligne sur des serveurs distants d'analyses phylogénétiques plus ou moins sophistiquées. NGPhylogeny intègre un grand nombre d'outils dans leurs versions les plus récentes (TNT, FastME, MrBayes, PhyML, SMS, FasTree, etc) et couvre une large gamme d'utilisations (alignement de séquences multiples, détection de modèle, inférence d'arbre et dessin d'arbres phylogénétiques) sur un large panel de méthodes (distance, parcimonie, maximum de vraisemblance, bayésien) (Lemoine et al. 2019). Enfin, les différents arbres phylogénétiques ont été formatés en utilisant le programme Interactive Tree Of Life (iTOL) v5 (Letunic et al. 2021), accessible via NGPhylogeny.

III – Résultats

1 - Analyses de la structuration génétique et spatiale de *C.pruni*

1.A – Analyses des séquences COI-t et ITS2 et réseaux d'haplotypes

Gene	COI-t - AB	COI-t - A	COI-t - B	ITS2 - AB	ITS2 - A	ITS2 - B
Nb séquences	544	212	332	783	339	444
Longueur (pb)	621	621	621	753	735	744
Ntot sites	581	581	621	726	735	744
Nb sites identiques	524	570	590	691	730	734
Nb sites poly (S)	57	11	31	35	5	10
Ntot mutations	61	11	33	37	5	11
Nb haplo (h)	44	13	34	17	8	12
Div haplo (Hd)	0,651	0,395	0,347	0,772	0,622	0,521
Div nucl (π)	0,02683	0,00086	0,00065	0,01898	0,00184	0,00369
sites informatifs	42	4	10	34	4	10
D de Tajima	2,11911	-1,72889	-2,48958	4,41588	1,31188	1,4567
P-value	*, P < 0,05	P > 0,05	***, P < 0,001	***, P < 0,001	P > 0,10	P > 0,10

Tableau 1 : Analyses descriptives et statistiques des séquences des gènes COI-t et ITS2. Nb = nombre, Ntot = nombre total, poly = polymorphique, haplo = haplotype/haplotypique, div = diversité, nucl = nucléotidique

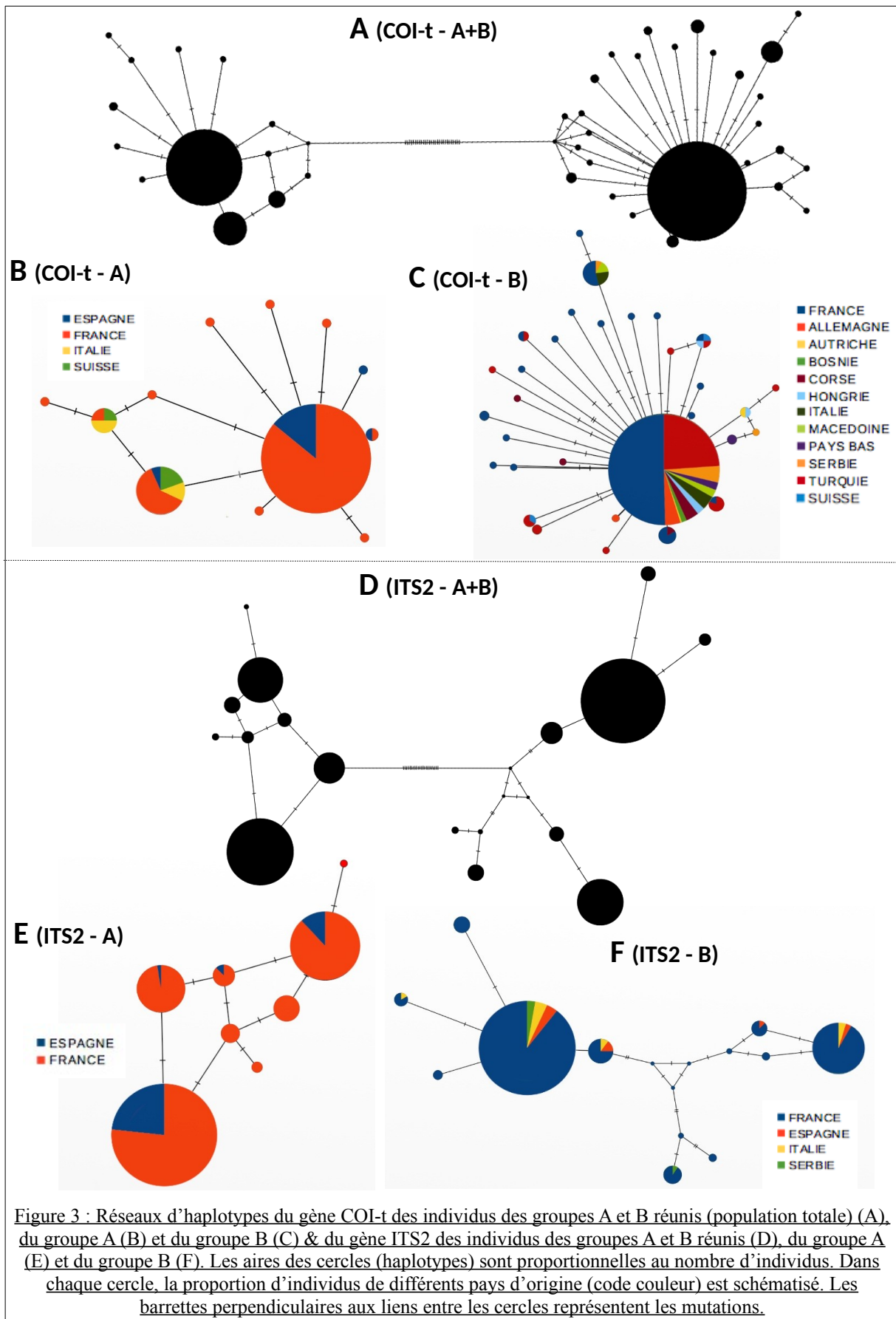
Le réseau d'haplotypes du gène COI-t pour la population totale (Fig. 3A) montre deux groupes principaux très éloignés génétiquement (33 mutations les séparent). La valeur du D de Tajima est positive et significative (Tab. 1), ce qui indique un excès de mutations à fréquences intermédiaires (=manque d'allèles rares). Cela peut s'expliquer par une sélection balancée, une contraction de population soudaine ou un effet Wahlund (Wahlund 1928).

En considérant uniquement les individus génotypés A par PCR diagnostique, on observe que le réseaux d'haplotypes pour le gène COI-t est très connecté et présente globalement très peu de mutations (Fig. 3B) : il présente 2 groupes principaux séparés par seulement une mutation ainsi que quelques individus isolés très apparentés (1 voire 2 mutations de différence). La valeur du D de Tajima est négative et non significative (Tab. 1) ce qui indique que la population A évolue de manière neutre (sans pression de sélection) pour le gène COI-t.

Le réseaux d'haplotypes des individus B pour le gène COI-t (Fig. 3C) présente une plus grande diversité d'haplotypes que pour A (31 vs 11), mais les individus B sont génétiquement très proches entre eux : le réseau se caractérise par un grand groupe principal avec beaucoup de très petits groupes séparés que d'une (ou parfois 2) mutation du groupe principal. La valeur du D de Tajima est négative et significative (Tab. 1) ce qui indique un excès d'allèles rares (= excès de mutations à faible fréquences) qui peut s'interpréter par une expansion rapide de la population des psylles du groupe B après un goulot d'étranglement.

Le réseau d'haplotypes du gène ITS2 pour la population totale (Fig. 3D) se caractérise aussi par 2 groupes très nettement séparés sur le plan génétique (21 mutations) et une valeur du D de Tajima positive et très significative (Tab. 1). Le réseaux d'haplotypes des individus A pour le gène ITS2 est très connecté (Fig. 3E) : tous les groupes sont séparés par un maximum de 4 mutations les uns des autres, et uniquement 1 à 3 mutations séparent les trois groupes principaux. La population A semble évoluer de manière neutre pour le gène ITS2 (D de Tajima = 1,3118 et p-value > 0,10).

Le réseaux d'haplotypes des individus B pour le gène ITS2 présentent 2 groupes principaux (Fig. 3F). Il semble plus différencié que les groupes du réseaux A : 7 mutations séparent les groupes. La distance génétique semble donc plus importante entre les individus du groupe B qu'entre ceux du groupe A. Cependant, la population B semble évoluer de manière neutre pour le gène ITS2 (D de Tajima = 1,4567 et p-value >0,10).



La figure 3 permet également d'analyser le nombre d'haplotypes privés en fonction du pays d'origine des échantillons. Pour le groupe A (Fig. 3B & 3E), les populations françaises regroupent la très grande majorité des haplotypes privés (7/8 pour COI-t et 4/4 pour ITS2). De même, les populations françaises du groupe B regroupent la majorité des haplotypes privés (13/23 pour COI-t et 9/9 pour ITS2) (Fig. 3C & 3F). Ces observations se vérifient par les analyses de raréfaction que nous avons réalisées sur les deux groupes en fonction de la zone géographique des échantillons collectés (tableau 2).

	Groupe	Pays	Gènes COI-t	Gène ITS2
nombre d'haplotypes privés	A	France	4,27	2,6747
		Espagne	3,47	0,0806
		Italie+Suisse	0,22	-
	B	France	9,13	3
		Europe centrale	3,15	0,26
		Balkans	5,36	0,07
		Corse	2,03	-
		Espagne	-	0,15

Tableau 2 : Pourcentages d'haplotypes privés de chaque groupe pour les gènes COI-t et ITS2 obtenus par analyse de raréfaction pour les différentes zones géographiques. Les - marquent une absence d'échantillon.

1.B – Analyses des marqueurs microsatellites

1 – Analyses multivariées

Pour la population totale (A+B), 2 groupes principaux sont discriminés selon l'axe 1 de la DAPC ($\approx 50\%$ de l'information apportée par cet axe) (Fig. 4). Un troisième voire quatrième groupe semblent se distinguer sur l'axe 2, mais celui-ci est moins informatif ($\approx 10\%$ de l'information) (Fig. 4).

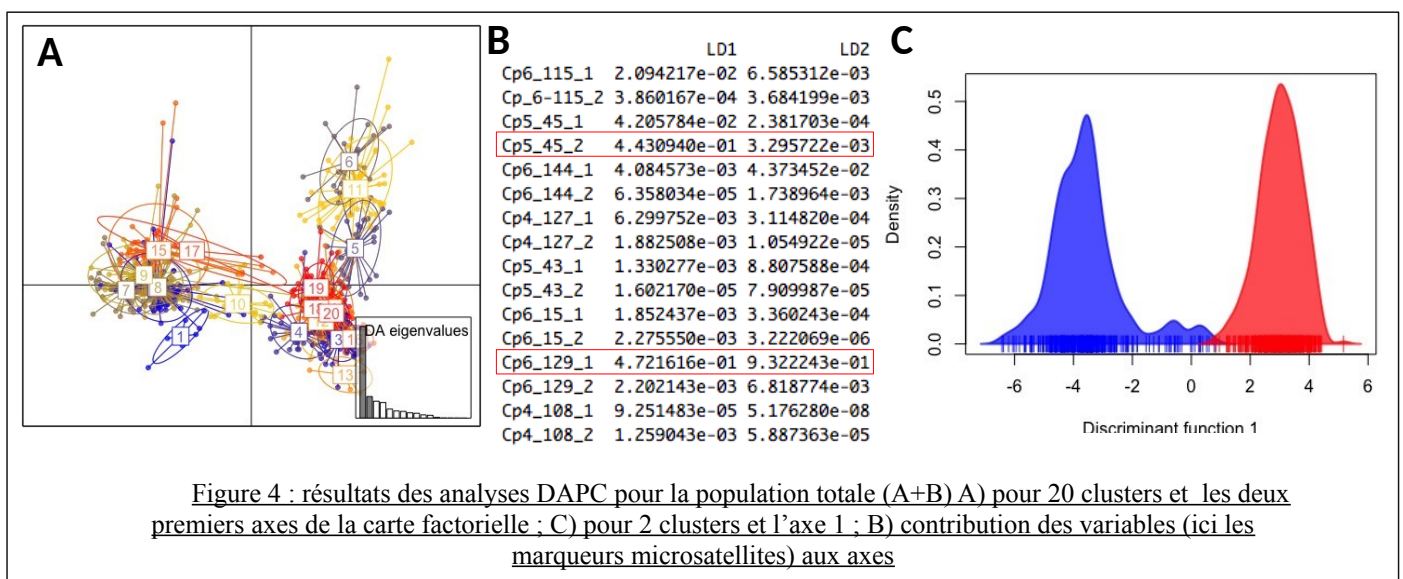
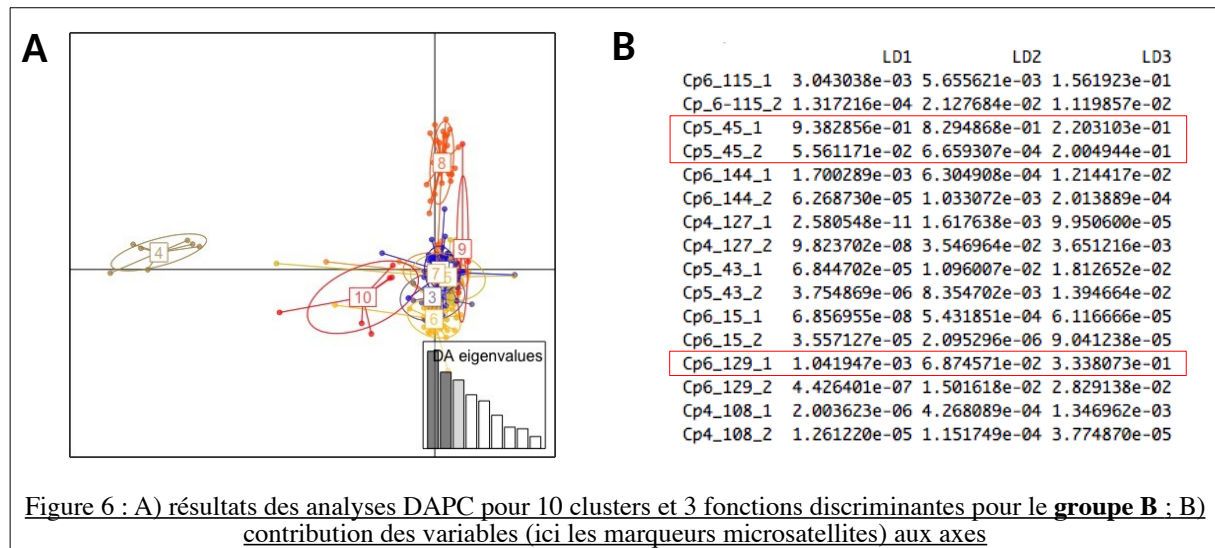
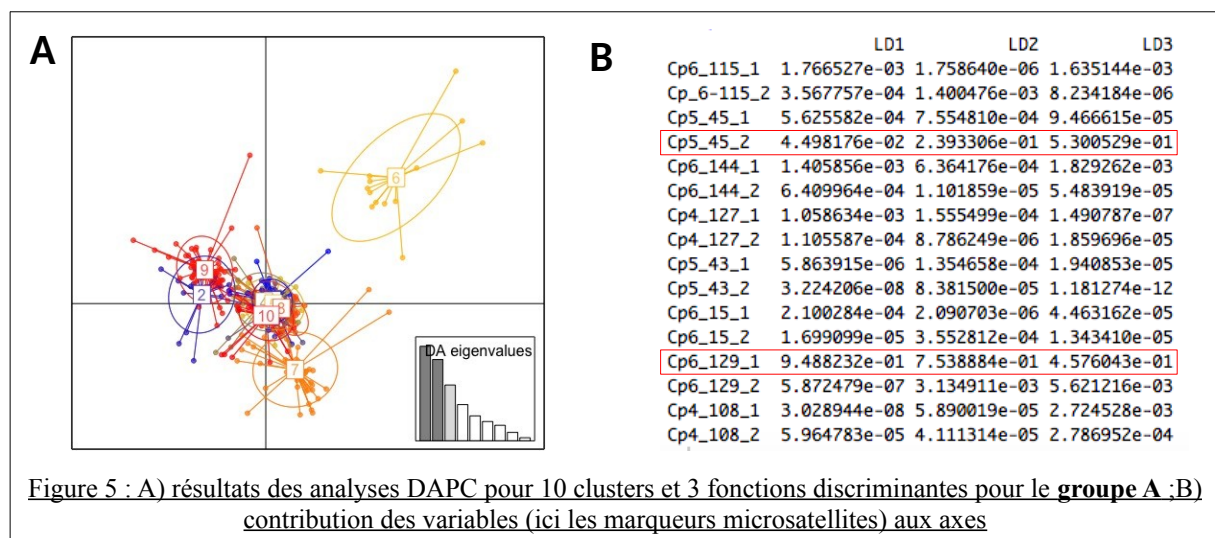


Figure 4 : résultats des analyses DAPC pour la population totale (A+B) A) pour 20 clusters et les deux premiers axes de la carte factorielle ; C) pour 2 clusters et l'axe 1 ; B) contribution des variables (ici les marqueurs microsatellites) aux axes

Analyse sur les individus des groupes A et B séparément : la majorité des clusters du groupe A sont regroupés (Fig. 5) et seul un cluster (n°6) semble se distinguer des autres selon l'axe 1 et 2. La majorité des clusters du groupe B sont également regroupés selon l'axe 1, le plus informatif (Fig. 6). Seul un cluster (n°4) semble clairement se distinguer des autres. Ces résultats indiquent une grande homogénéité génétique des individus au sein des groupes A et B respectivement, à l'exception de quelques individus, qui semblent un peu plus éloignés génétiquement des autres dans chaque groupe.

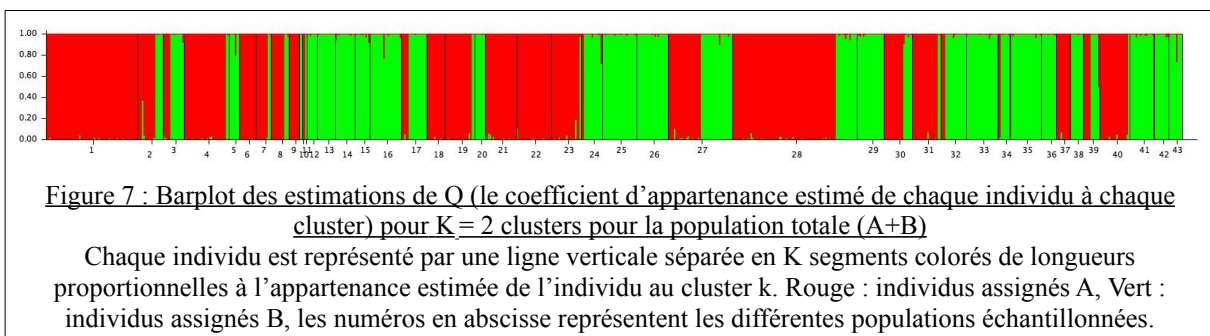
Deux variables (Cp6_129_1 et Cp5_45_2) contribuent pour l'essentiel (95%) à la définition de l'axe 1 (Fig. 5B, 6B et 7B) pour les groupes A, B et la population totale.



2 – Clustering Bayésien

Sur la population totale : lorsque que le kmax est fixé à 2, STRUCTURE assigne les individus aux groupes attendus d'après les PCR diagnostiques (Fig. 7).

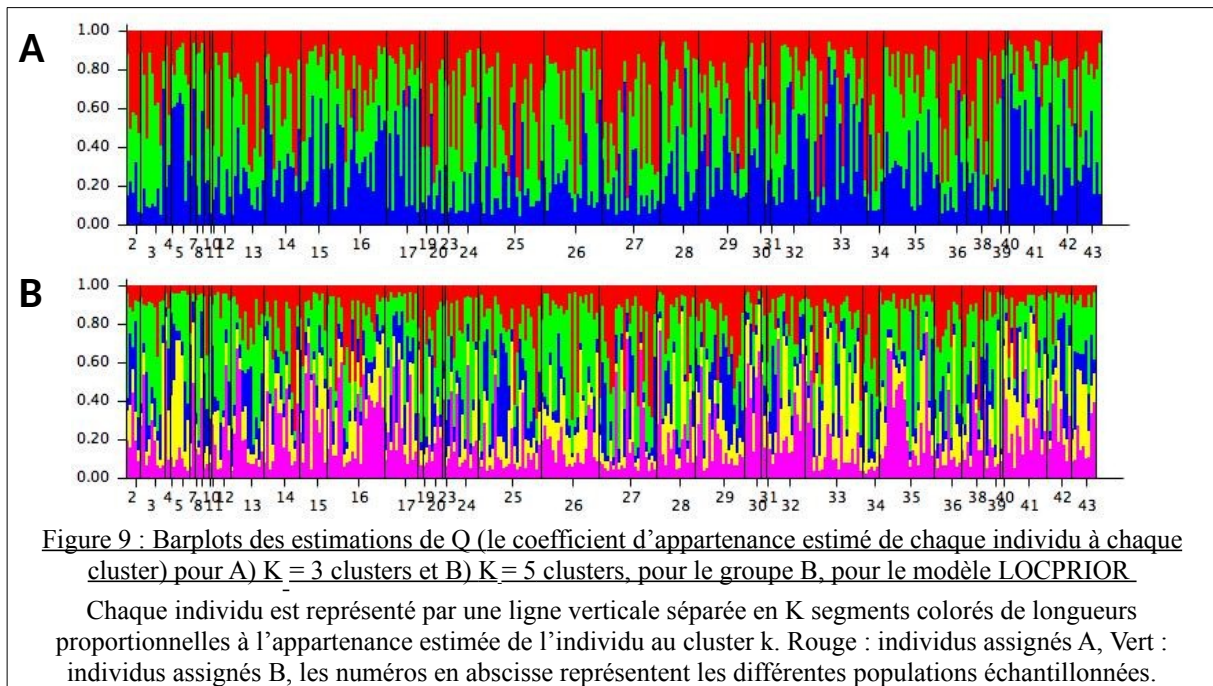
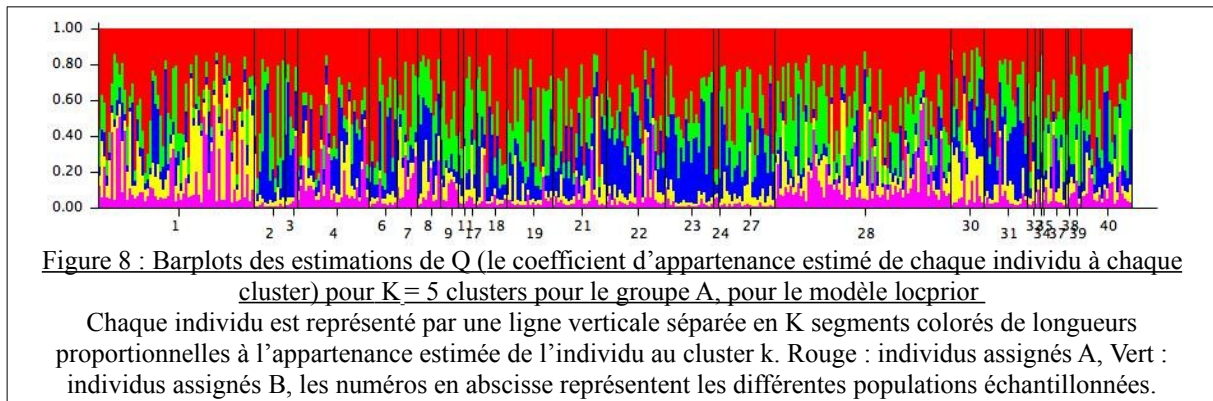
Les analyses STRUCTURE montrent que, quel que soit le modèle utilisé (avec ou sans admixture, $k=2$ ou plus, avec et sans fréquences alléliques corrélées), la moyenne des estimations des Ln probabilités ($\ln(P(X|K))$) est la plus grande pour $k = 2$ clusters. Ces analyses montrent aussi l'absence d'admixture entre les populations A et B et que les fréquences alléliques sont indépendantes entre ces deux groupes. TESS donne un résultat similaire. Les valeurs du DIC de tous les runs pour tous les modèles testés (avec et sans admixture) et tous les k_{max} considérés, sont négatives et extrêmement proches de 0 (la valeur du DIC pour $k_{max}=2$). Ces valeurs très proches entre elles et proches de la valeur du DIC pour $k_{max} = 2$ semble indiquer que l'on est déjà dans le « plateau » de la courbe pour $k_{max} = 2$. Les analyses TESS confirme aussi l'absence d'admixture.



Pour le groupe A : STRUCTURE montre que pour le modèle LOCPRIOR, la moyenne des estimations des Ln probabilités ($L(K)$) est la plus haute pour $k = 5$ clusters (Fig. 8). La méthode d'Evanno sur le même modèle ainsi que sur un modèle différent non LOCPRIOR donnent des ΔK de Evanno maximisés pour $k = 5$ clusters. Les analyses TESS, quand à elles, montrent que c'est le modèle avec admixture de type CAR qui a le DIC moyen, pour tout k, le plus faible. Les modèles avec admixture ont des DIC plus faibles que ceux sans admixture dans TESS, ce qui indiquerait de l'admixture au sein du groupe A. Les courbes de DIC pour le modèle CAR montrent que l'on a le début du plateau de la courbe des DIC en fonction des K_{max} pour une valeur de K entre 5 et 7. Cependant, sur le modèle BYM, le DIC le plus faible est pour $k_{max} = 5$. Le nombre de cluster maximum K_{max} pour TESS est donc de 5. Cependant, le modèle avec le DIC le plus bas pour $k_{max} = 5$ ne présente que 3 (voire 2) clusters sur la projection Hard Clustering sur les tessellations de Voronoï et sur le barplot des estimations de Q : aucun individu n'est attribué aux deux derniers clusters. Le groupe A présente donc probablement entre 2 et 5 clusters.

Pour le groupe B : STRUCTURE montre que pour le modèle LOCPRIOR, la moyenne des estimations des Ln probabilités ($L(K)$) est la plus haute pour $k = 3$ clusters. Le ΔK de

Evanno est maximisé pour $k = 3$ clusters pour le modèle LOCPRIOR et $k = 5$ pour l'autre modèle (Fig. 9). Les analyses TESS ont montrées que les DIC de tous les runs pour tous les modèles et tous les k_{max} considérés sont négatifs et extrêmement proche de 0. Ces valeurs très proches entre elles et de la valeurs du DIC pour $k_{max} = 2$ semblent indiquer que l'on est déjà dans le « plateau » de la courbe pour $k_{max} = 2$, d'où le très peu de variations passé cette valeur de k_{max} . Le nombre de cluster K du groupe B est donc probablement compris entre 2 et 5.



Pour aucun des groupes inférés par STRUCTURE ou TESS nous n'avons pu faire le lien entre la structuration génétique et une éventuelle structuration géographique. Autrement dit, les différents clusters génétiques semblent être répartis uniformément dans tous les sites d'échantillonnages. Nous avons estimés les valeurs de r , un paramètre permettant d'estimer la quantité d'information portée par les localisations introduites dans le modèle LOCPRIOR.

Elles ont toujours été bien supérieures à 1 pour toutes les itérations du modèle, pour les groupes A et B et pour tous les k testés. Ce résultat indique que la structure des populations observée est indépendante de la localisation. Au sein des deux groupes, une proportion non négligeable d'individus présente une proportion d'assignation assez importante à plusieurs clusters, ce qui indique que l'on aurait un niveau d'admixture assez haut au sein des groupes.

3 – Calculs d'indicateurs statistiques et tests

Les tests exacts de l'équilibre d'Hardy-Weinberg pour les différents groupes étudiés (Tab. 3) ont révélés que ni la population totale (A+B), ni les groupes A et B pris séparément ne suivent l'équilibre de Hardy-Weinberg. En effet, Les probabilités obtenues par le test sont toutes très faibles voir égales à 0 et sont hautement significatives pour tous les marqueurs et pour toutes les populations.

Globalement, les Rhos sont positifs pour tous les locus et toutes les populations (Tab. 3), ce qui confirme que les écarts à l'équilibre de HW décrits plus haut seraient dus à un déficit d'hétérozygotes dans la population totale (A+B), le groupe A et le groupe B.

La population totale (A+B) se caractérise aussi par une différenciation entre les populations assez forte (Rhost = 39,47%) mais surtout par une très forte différenciation génétique entre individus au sein des sous-populations (Rhois = 56,25%) et au sein de la population totale prise dans son ensemble (Rhoit = 73,52%).

Au sein du groupe A, la différenciation entre les populations est très faible (Rhost = 1,32%), bien que la différenciation entre les individus au sein des sous-populations et au sein de la population totale soient assez hautes et très proches (Rhois = 35,11% et Rhoit = 35,97%). Pour le groupe B, la différenciation entre les populations est très faible (Rhost = 0,79%), bien que la différenciation entre les individus au sein des sous-populations et au sein de la population totale soient assez hautes et proches (Rhois = 45,90% et Rhoit = 46,33%). Les résultats des tests exact de Hardy-Weinberg et les Rho-statistiques calculées par locus pour toutes les populations sont décrits dans l'Annexes 7.

Groupe	Test exact de Hardy - Weinberg			Rho-statistiques		
	Chi2	Df	Proba	Rhois	Rhost	Rhoit
AB	> 3582.8118	670	High. Sign (<0,01)	0.5625	0.3947	0.7352
A	> 1403.5052	380	High. Sign (<0,01)	0.3511	0.0132	0.3597
B	> 1771.3262	460	High. Sign (<0,01)	0.4590	0.0079	0.4633

Tableau 3 : Tableau des résultats pour tout les locus et toutes les populations du test exact de Hardy-Weinberg (test probabilité) et des Rho-statistiques pour la population totale AB (noir), A (rouge) et B (vert).

Df = degré de liberté ; Proba = probabilité

Les résultats des tests de Mantel (tests d'isolement par la distance) entre populations pour les différents groupes étudiés sont détaillés dans l'Annexe 8. Ils montrent que l'on a une pente et donc une corrélation positive entre la matrice des distances génétiques et celle des distances géographiques pour la population totale AB et le groupe B. Il y aurait donc un isolement par la distance des différentes populations de ces deux groupes. Cependant, cet isolement par la distance disparaît pour le groupe B quand les échantillons hors France sont retirés. À l'échelle française, il n'y a pas non plus d'IBD au sein du groupe A.

La diversité génétique des groupes A et B, reflétée par le calcul de la richesse allélique et la richesse allélique privée de chacun des groupes est présentée dans le tableau 4. Le groupe A possède plus de loci (5/8) avec une plus grande richesse allélique que B. Mais en moyenne le groupe B a une plus grande richesse allélique que le groupe A (22,63 vs 20,42). La même tendance est observable pour la richesse allélique privée (8,82 pour B vs 6,62 pour A).

	Groupe	Average over loci	Cp6_115	Cp5_45	Cp6_144	Cp4_127	Cp5_43	Cp6_15	Cp6_129	Cp4_108
R A	A	20.42	10.7269	10.6128	35.777	20.5961	29.1785	29.9544	8.5964	17.9570
	B	22.63	9.00	7.00	25.00	39.00	19.00	44.00	8.00	30.00
R A privée	A	6.62	3.7432	4.6148	11.8039	2.8716	16.58051	6.6148	2.7411	3.9816
	B	8.82	2.0163	1.0021	1.0269	21.2755	6.4016	20.6604	2.1447	16.0246

Tableau 4 : Tableau de la richesse allélique (RA) et de la richesse allélique privée (RA privée) des groupes A et B calculée par raréfaction

2 – Reconstruction des scénarios évolutifs : analyses ABC

Le résumé des résultats de l'analyse ABC est présenté dans le tableau 5 et le détail des différents tests et différentes étapes sont présentés en Annexe 9.

Etape 1 : les résultats montrent une date de divergence entre A et B très ancienne, estimée entre 40 et 50 Ma (Eocène). Ce résultat correspond au scénario 6 testé dans l'ABC (Annexe 5), inféré par 55% des répliques avec une probabilité à posteriori de 0,6439. Cependant, il est possible que la date de divergence soit un peu plus ancienne : 35% des tests ont inféré une date >50Ma. Néanmoins, avec 95% des répliques inférant une date de divergence >40Ma, il est très probable que la date de divergence de A et B est plus ancienne que 40 Ma.

Etape 2 : 55% des répliques montrent que A et B partagent un ancêtre commun (scénario 1) avec une probabilité à posteriori de 0,7039. Mais 30% des répliques indiquent que A serait l'espèce ancestrale et que B aurait dérivé de A (scénario 2). Il est donc difficile de trancher entre l'une et l'autre des hypothèses.

Etape 3 : les résultats indiquent que les populations françaises ont subi une expansion après la date de divergence avant de rétrécir à des dates plus récentes (81,82% des répliques avec une probabilité à posteriori de 0,5554 pour A et 100% des répliques avec une probabilité à posteriori de 0,6967 pour B), ce qui correspond dans ces tests au scénario 7 (Annexe 5). Le postérieur de la date de rétrécissement étant $< 5\text{Ma}$ dans 100 % des cas, nous avons essayé de caractériser le type de rétrécissement (rétrécissement simple ou bottleneck) et sa date plus précise dans l'étape 4.

Etape 4 : les résultats sont différents entre les populations A et B. Les tests réalisés sur les individus B ont inféré un rétrécissement simple (et non pas un bottleneck) il y a 16.000 ans ou plus récemment (scénario 4) dans 85,72% des répliques avec une probabilité à posteriori de 0,4998. Les résultats obtenus avec les populations A sont plus difficiles à interpréter : le scénario 4 et 3 (bottleneck pré glaciation (entre 3 et 10 Ma) puis expansion) sont inférés à égalité par 33% des répliques.

Etape 5 : les résultats des analyses sur le sens d'expansion des populations en Europe indique que la zone d'origine de *C. pruni* serait ce que l'on appelle aujourd'hui la France. Les populations A et B auraient ensuite divergé indépendamment à l'ouest (en Espagne) pour les A et à l'est (en Europe centrale, de l'ouest et dans les Balkans) pour les B (95,82% des répliques inférant le scénario 3 ou 4). Par contre, il n'est pas possible de distinguer plus précisément lequel du scénario 3 (divergence des A entre la France et Espagne avant celle des B entre la France et l'est de l'Europe) ou du scénario 4, inverse, est le bon. Le scénario 3 est celui le plus inféré avec 54,16% des répliques et une probabilité à posteriori de 0,8093 mais le scénario 4 est inféré dans 41,66% des cas.

Pour préciser nos analyses, nous avons réalisé une étape complémentaire, dans lequel nous avons refait les mêmes tests que dans l'étape 5, avec les même scénarios, mais écrits de manière différente. Cette fois-ci, nous n'avons pas précisé la divergence entre A et B dans le scénario. Ainsi, les populations sont écrites comme divergeant des populations françaises en mélange (A et B non dissociés) (5bis). Le scénario 3 est alors ressorti comme le vainqueur (75% des répliques le choisissent avec une probabilité à postérieure moyenne de 0,9182). Nous sommes ainsi plus confiants sur le fait que le scénario 3 serait le plus probable. Cependant, les tests 5 et 5bis ne permettent pas d'estimer les paramètres du modèles. Aussi, il n'a pas été possible de préciser les dates de divergences des populations d'Espagne et d'Europe de l'Est avec la population française d'origine.

étape	nombre de scénarios	scénario sélectionné	proba. Post. moyenne du scénario (régression logistique)	P.E.R (régression logistique)	% du nombre de fois où le scénario est gagnant	second scénario gagnant et pourcentage
1	7	6 : divergence A-B entre 40 et 50 Ma	0,6439	0,6413	55,00 %	7 : divergence entre 50 et 100 Ma (35%)
2	3	1 : A et B descendent d'un ancêtre commun	0,7039	0,5907	55,00 %	2 : B descend de A, l'espèce ancestrale (30%)
3 groupe A	10	7 : expansion après la divergence puis rétrécissement	0,5554	0,773	81,82 %	5 : expansion après la divergence puis constant (18,18%)
3 groupe B	10	7 : expansion après la divergence puis rétrécissement	0,6927	0,578	100,00 %	X (0%)
4 groupe A	6	4 : rétrécissement anthropique (16.000 ans à maintenant)	0,4367	0,756	33,00 %	3 : bottleneck pré-glaciation (3 à 10Ma) puis réexpansion (33%)
4 groupe B	6	4 : rétrécissement anthropique (16.000 ans à maintenant)	0,4998	0,591	85,72 %	3 : bottleneck pré-glaciation (3 à 10Ma) puis réexpansion (14,28 %)
5	4	3 : population française = population d'origine et divergence des populations espagnoles puis balkaniques indépendantes	0,8093	0,178	54,16 %	4 : population française = population d'origine et divergence des populations balkaniques puis espagnoles indépendantes (41,66%)
5bis	4	3 : population française = population d'origine et divergence des populations espagnoles puis balkaniques indépendantes	0,9182	0,299	75,00 %	2 : Population Espagnole = population d'origine et divergence des populations françaises puis balkaniques à partir des françaises (20,8%)

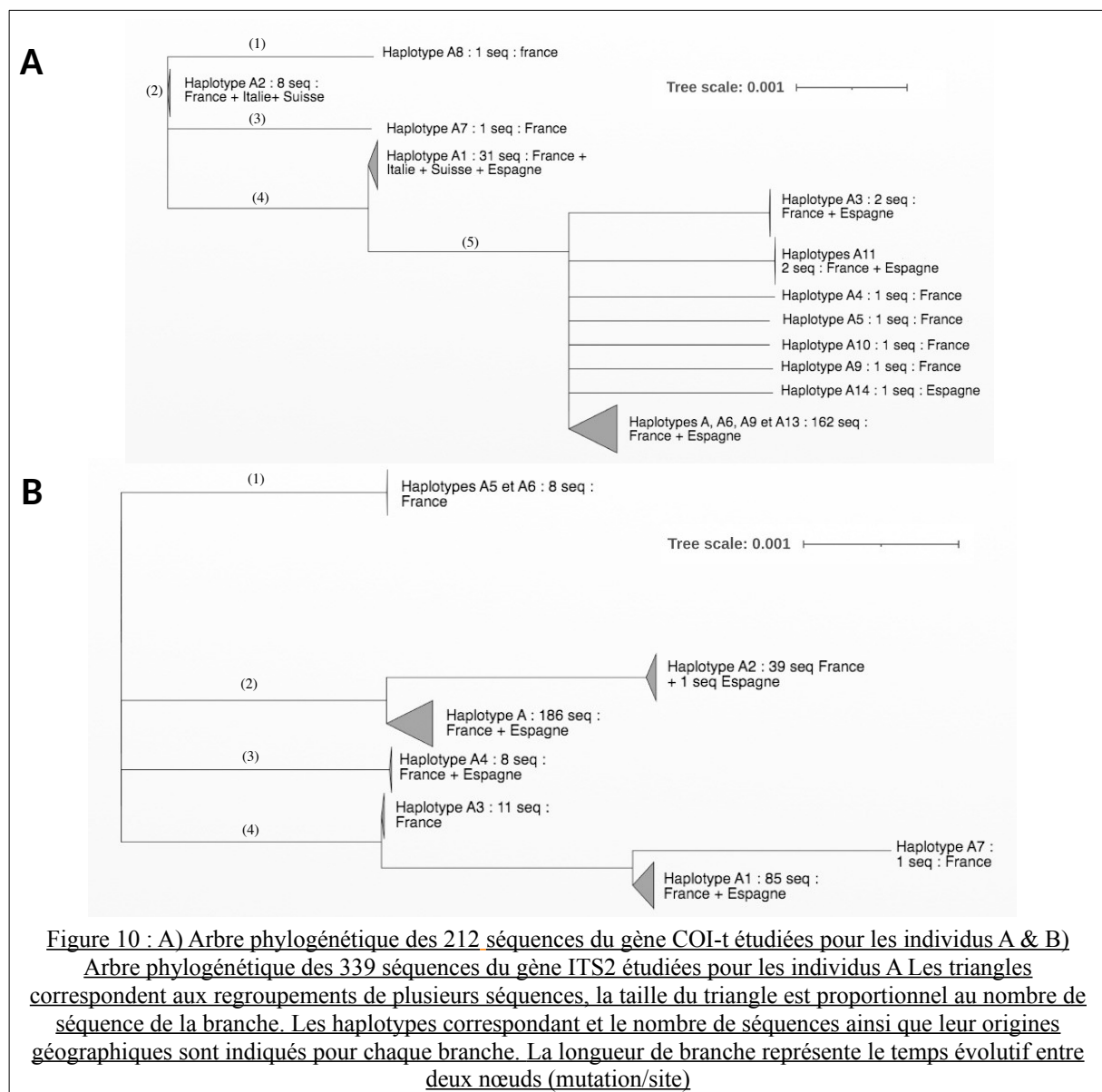
Tableau 5 : Tableau des résultats de l'analyse ABC. Proba. Post = probabilité à postériori ; P.E.R : Erreur Prédictive à Postériori moyenne

3 – Analyses phylogénétiques

Les arbres phylogénétiques des séquences des gènes COI-t et ITS2 utilisées dans les analyses précédentes (réseaux d'haplotypes, haplotypes privés, analyses ABC) sont présentés en figure 10 et 11.

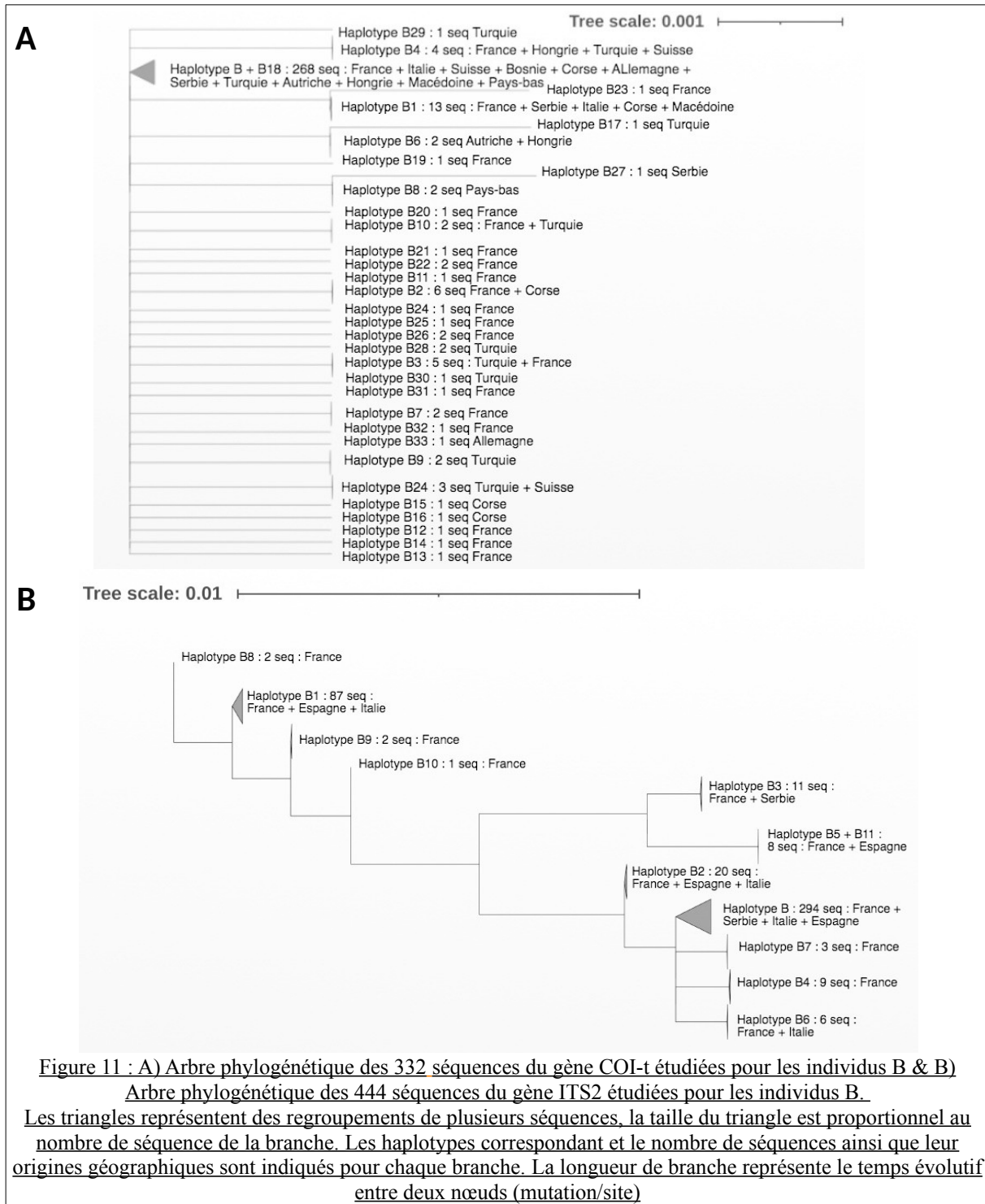
Pour le gène COI-t (Fig. 10A), les haplotypes des individus A d'Espagne sont tous regroupés sur la branche (4) de l'arbre, celle ayant accumulée le plus de mutations par site. Ces haplotypes espagnols sont apparentés à des haplotypes français. Les échantillons trouvés en Italie et en Suisse sont associés à deux des haplotypes (A1 et A2) le plus souvent observés, notamment en France et plus rarement en Espagne. Ces haplotypes (A1 et A2) avec les haplotypes français A7 et A8 semblent être les haplotypes ancestraux alors que les haplotypes trouvés majoritairement en Espagne et en France (branche 5) sont plus dérivés pour le gène COI-t et les individus A. Pour le gène ITS2 chez les individus A (Fig. 10B), on retrouve encore une fois que les haplotypes retrouvés en Espagne (A, A1, A2, A4 & A7) sont plus

souvent dérivés que les haplotypes français (A3, A5, A6). Cependant pour le gène ITS2 l'algorithme a eu du mal à dissocier les différents haplotypes par manque de variabilité des séquences. Il est donc difficile ici d'interpréter cet arbre.



Pour le groupe B, les haplotypes du gène COI-t retrouvés sur toute l'aire de répartition semblent tous très proches génétiquement entre eux, la très grande majorité des haplotypes présentant la même quantité de mutations par site (Fig. 11A). Cependant l'haplotype B, de loin le plus répandu, semble présenter moins de mutations par site que les autres. Cela indique que les autres haplotypes sont dérivés de l'haplotype B, qui serait l'haplotype ancestral. Cet haplotype se retrouvant sur toute l'aire de répartition de l'espèce B, il est cependant impossible de conclure sur la zone géographique ancestrale pour le gène COI-t de l'espèce B.

Pour le gène ITS2 des individus B (Fig. 11B), les haplotypes se trouvant à gauche de l'arbre et donc les moins dérivés sont des haplotypes échantillonnés en France, Italie et Espagne, et donc à l'extrémité ouest de l'aire de répartition du groupe B. Les haplotypes échantillonnés en Serbie, eux, présentent un plus grand nombre de mutations par site et semblent donc être les haplotypes les plus dérivés.



IV - Discussion

C. pruni : deux groupes génétiques très différenciés à l'échelle de l'Europe

L'objectif premier de cette étude était de confirmer la séparation de *C. pruni* en deux espèces cryptiques à l'échelle de l'aire de distribution connue (Sauvion et al., 2021), et ainsi confirmer les résultats antérieurs produits à plus petite échelle (Marie-Jeanne et al. 2020).

Deux groupes (A et B) sont bien distingués génétiquement quelles que soient les approches utilisées (réseaux d'haplotypes, analyse multivariée, clustering bayésien) et le type de marqueurs analysés (gènes COI-t, ITS2, microsatellites).

Les calculs de divers indicateurs statistiques appuient aussi ce point. Par exemple, les valeurs du D de Tajima calculées pour la population totale pour les gènes COI-t et ITS2 sont positives et significatives, ce qui indique un excès de mutations à fréquences intermédiaires (=manque d'allèles rares). Cela peut résulter soit d'une sélection balancée, soit d'une contraction de population soudaine ou encore d'un Effet Wahlund (Wahlund 1928), hypothèse que l'on privilégiera ici (nous savons que nous avons deux populations en sympatrie dans plusieurs sites d'échantillonnage (Annexe2)). Cette dernière est appuyée par le fait que la population totale ne suit pas l'équilibre de Hardy-Weinberg (déficit d'hétérozygotes).

La distance génétique entre les groupes A et B explique l'essentiel de la différenciation génétique observée dans la population totale et le fait que l'on ai pas, ou peu, de sous structuration géographique des groupes A et B. En effet, les Rho-statistiques montrent une assez forte différenciation entre les populations lorsque que l'on considère la population totale, mais une différenciation entre les individus au sein des sous-populations plus importante, et une différenciation entre les individus au sein de la population totale encore plus grande ($R_{hoit} > R_{hois} > R_{host}$). De plus, le déficit d'hétérozygotes s'estompe et la différenciation entre les population (R_{host}) disparaît presque totalement lorsqu'on sépare les groupes A et B.

De plus, la population totale ne présente pas d'admixture entre les groupes et les fréquences alléliques ne sont pas corrélées. En effet, les simulations utilisant des modèles avec et sans admixture de STRUCTURE et TESS et avec et sans fréquences alléliques corrélées de STRUCTURE donnent des résultats similaires. Cela confirme le fort isolement reproducteur et l'absence de flux de gènes entre A et B qui avait été étudié à l'échelle du sud de la France (Peccoud et al. 2018), mais cette fois-ci à plus grande échelle.

Ainsi, au vu des résultats des différentes analyses, nous pouvons confirmer que le complexe *C. pruni* est divisé en deux et uniquement deux groupes très différents génétiquement à l'échelle du paléarctique ouest. Ces deux groupes ont le même cycle biologique à ceci près que les individus A semblent préférés les pins et les B les épicéas ou sapins comme plante refuge. Un seul critère morphologique très tenu permet de distinguer les mâles A et B et les femelles des deux groupes sont indistinguables (Sauvion, Con. Pers). Tous ces éléments indiquent que nous avons à faire à deux espèces cryptiques (Struck et al. 2018; Jörger et al. 2013; Fišer et al. 2018).

Structuration intra-spécifique ?

Les analyses de structuration des populations au sein des groupes A et B permettent de préciser le processus de spéciation hypothétique en cours au sein de ces groupes.

Plusieurs clusters se distinguent au sein des groupes A et B quelles que soient les approches utilisées (réseaux d'haplotypes, analyses multivariées, clustering bayésien) et le type de marqueurs analysés (gènes COI-t, ITS2, microsatellites). Mais, les individus au sein des groupes semblent malgré tout génétiquement assez proches entre eux. L'analyse DAPC révèle l'existence de plusieurs groupes d'individus se différenciant de la grande majorité dans les groupes A et B. Mais ce résultat est très lié au poids de deux marqueurs microsatellites (Cp129 & Cp45) qui influencent fortement la définition des axes de la DAPC. Nous éviterons donc de surinterpréter ces résultats. Les analyses de clustering bayésien sur STRUCTURE et TESS distinguent également des clusters au sein des groupes A et B, mais pas le même nombre (STRUCTURE : 5 clusters pour le groupe A et 3 à 5 pour B vs TESS : 2 ou 3 clusters au sein des groupes A et B). Les groupes A et B ne suivent pas non plus l'équilibre de Hardy-Weinberg (déficit d'hétérozygotes), l'hypothèse d'un effet Wahlund (Wahlund 1928) confirmerait donc une structuration des groupes A et B. Les différenciation entre les individus au sein des sous-populations (Rhois) et au sein de la population totale (Rhoit) très fortes pour les groupes A et B vont également dans le sens de l'existence de clusters génétiques au sein de ces groupes.

La structuration génétique au sein des groupe A et B ne se retrouve pas au niveau géographique. L'introduction d'informations spatiales n'influencent pas les analyses de clustering bayésien réalisées avec STRUCTURE et TESS, la structuration génétique interne aux groupes A et B est donc indépendante de la localisation des individus. De même, les différents haplotypes des groupes A et B ne sont pas structurés géographiquement. Les Rho-

statistiques montrent que des individus pourtant très éloignés géographiquement ne se différencient pas ou très peu au sein de chacun des groupes (Rhoist faible & Rhois \approx Rhoit). C'est par exemple le cas entre les psylles A du Nord et du Sud de la France.

Aussi, il n'y a pas d'isolement par la distance (IBD) des populations au sein du groupe A. En revanche, un isolement est observé au sein du groupe B à l'échelle du paléarctique ouest. À l'échelle de la France, cet IBD n'est plus observé au sein du groupe B.

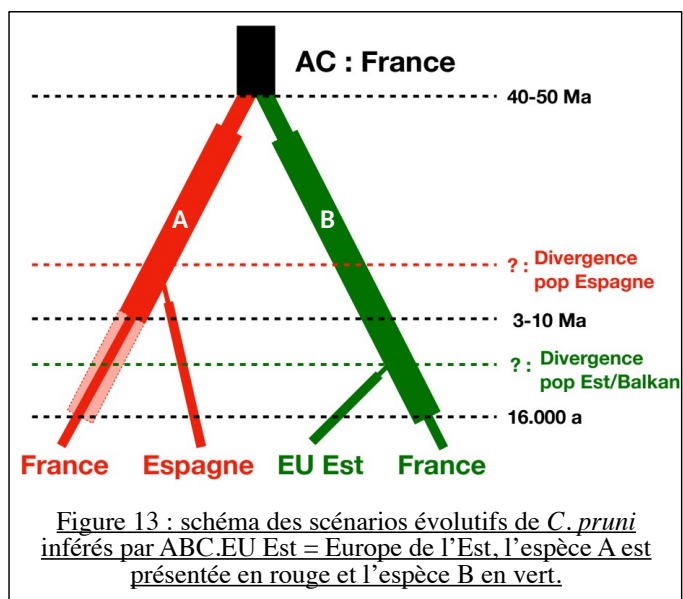
Les individus des groupes A et B sont donc génétiquement proches entre eux, pour autant qu'il existe des clusters génétiques au sein de ces groupes. On ne retrouve pas cette structuration à l'échelle géographique. Deux hypothèses peuvent expliquer ce phénomène : soit il existe un grand flux de gène homogénéisant les différentes populations de ce groupe à grande échelle spatiale (France), soit l'expansion de ces populations dans cette région a été soudaine et récente, la totalité de la diversité de la population ancestrale se retrouvant donc de manière homogène à l'échelle de la France. L'hypothèse d'une spéciation en cours au sein de l'espèce A semble peu probable du fait de la forte homogénéité génétique que nous observons à l'échelle du territoire français. L'espèce B est cependant répartie sur une aire géographique plus étendue. L'IBD observé traduirait un processus de spéciation allopatrique en cours sur un gradient Est-Ouest pour cette espèce.

Histoire phylogéographique de C.pruni

Les analyses (ABC, analyses phylogénétiques, calcul d'indicateurs statistiques) nous ont permis de reconstruire l'histoire évolutive du complexe *C. pruni* (Fig. 13).

Les analyses ABC ont tout d'abord permis d'estimer, avec une assez bonne certitude, la date de divergence des deux espèces A et B comme datant d'il y a 40-50 Ma. Cependant, nos résultats montrent qu'il est possible que celle-ci soit encore un peu plus ancienne.

Les espèces A et B ont probablement divergé d'un ancêtre commun mais un second scénario n'est pas à exclure : A pourrait être l'espèce ancestrale et B aurait dérivé de A à la date de divergence.



Ces résultats sont cohérents avec ceux des calculs de la richesse allélique privées (Tab. 4) qui montrent que l'espèce A présente plus de locus avec une plus grande richesse allélique privée que B, mais que l'espèce B présente plus de richesse allélique privée moyenne que A. Ici aussi, ces résultats ne permettent pas de privilégier un scénario plutôt qu'un autre.

Les analyses ABC montrent également que les populations françaises de A et de B ont subi une expansion après la date de divergence avant de rétrécir à des dates plus récentes. Pour les populations B, il semble que ce rétrécissement se soit opéré il y a 16.000 ans ou plus récemment. En revanche, pour les populations A, les calculs ne permettent pas de dissocier un scénario identique à B de celui d'un bottleneck pré-glaciation entre 3 et 10 Ma.

Enfin, les analyses montrent que la population ancestrale était localisée dans la zone géographique que nous appelons aujourd'hui la France. Celle-ci aurait alors divergé : des individus A auraient migré vers l'Espagne et, plus tard, des individus B auraient migré vers l'Est. Il n'a pas été possible d'inférer les dates de divergence des populations d'Espagne et des Balkans, le logiciel ayant échoué à estimer les paramètres du modèle pour les tests effectués. Ces résultats sont cohérents avec ceux obtenus avec les réseaux d'haplotypes, les calculs d'haplotypes privés, et les analyses phylogénétiques réalisées. En effet, les réseaux d'haplotypes pour les deux gènes (COI-t et ITS2) présentent plus d'haplotypes privés en France que dans toutes les autres populations, et ce pour les deux espèces (Fig. 3 & 4). Résultats confirmés par le calcul des haplotypes privés par raréfaction (Tab. 2). Les arbres phylogénétiques des séquences des gènes COI-t et ITS2 pour A et B (Fig. 10 & 11) vont également dans le sens de ces résultats. En effet, l'analyse des arbres obtenus pour l'espèce A et B montrent que les haplotypes trouvés en France, Italie et Suisse (et nord de l'Espagne pour l'espèce B) seraient les haplotypes ancestraux et que les haplotypes trouvés en Espagne pour l'espèce A, et dans les Balkans (Serbie) pour l'espèce B sont dérivés.

Nous avons noté que l'arbre phylogénétique obtenu avec les séquences ITS2 présente des résultats moins tranchés que l'arbre de COI-t pour l'espèce A. Ces résultats sont probablement dus au fait que les populations d'individus A échantillonnées en Espagne pour ITS2 étaient très proches de la frontière française (région de Barcelone), et donc probablement plus proches génétiquement des populations françaises que les populations échantillonnées au Nord de l'Espagne pour le gène COI-t. Aussi, le gène COI-t semble moins bien discriminer les populations que le gène ITS2 pour l'espèce B (visible à la fois sur l'arbre phylogénétique et le réseau d'haplotype). Nous pouvons poser l'hypothèse que le gène COI-t évolue plus lentement que le gène ITS2 chez l'espèce B par rapport à l'espèce A.

Nous sommes conscient que certains de nos résultats doivent être interprétés avec prudence. Par exemple, certaines de nos erreurs prédictives à postériori sont assez hautes. De plus, certains tests (5 et 5bis) n'ont pas été capables d'estimer les paramètres à postériori du scénario, possiblement à cause du trop petit nombre de marqueurs utilisés (un seul gène, ici COI-t), du trop grand nombre de population utilisées dans les scénarios, ou d'une combinaison de ces deux facteurs.

En effet, la clef d'une application fructueuse des méthodes ABC réside dans la capacité des statistiques résumées à capturer les propriétés pertinentes des données. Ainsi, bien que les analyses ABC ont le potentiel d'utiliser des modèles très complexes, augmenter la complexité du modèle n'apportera pas d'informations supplémentaires si les données ne peuvent pas supporter ce genre de complexité (Nielsen et al. 2009). De plus, l'estimation bayésienne de paramètres peut être compliquée pour des cas de plus de deux populations, ou à l'inverse, améliorée par la combinaison de jeux de données avec des types marqueurs différents (par exemple des microsatellites et des gènes) (Wegmann et al. 2010). Ainsi, notre jeu de données combinant des marqueurs variables (différentiation à l'échelle de l'espèce) et hypervariables (différentiation à l'échelle infra-spécifique), celui-ci était théoriquement très bien adapté à l'inférence d'histoires évolutives par des approches ABC. C'est d'ailleurs probablement cette combinaison de marqueurs qui nous a permis de reconstruire cette histoire évolutive du complexe *C. pruni*. Cependant, nous semblons tout de même avoir atteint la limite de résolution de nos données.

Pour compléter nos résultats dans le futur, l'idéal serait de compléter notre jeux de données actuel en séquençant tous les individus du jeu de données pour tous les marqueurs. Par exemple ici, les populations hors de France n'ont pas été séquencées pour les microsatellites. L'ajout de plus de gènes et plus de marqueurs microsatellites augmenteraient certainement beaucoup la résolution de nos analyses. Précisément dans ce but, 15 nouveaux marqueurs microsatellites sont en cours de développement dans mon équipe d'accueil.

Enfin, on peut se poser la question du sens biologique de l'histoire évolutive que nous avons mis en évidence pour le complexe cryptique de *C. pruni*. La date de divergence entre les deux espèces de *C. pruni* étant ancienne, nous nous trouverions face à un cas d'espèces cryptiques en stase (par opposition à une divergence récente ou une convergence, autres cas pouvant expliquer le manque de distinction morphologiques entre deux espèces) telle que décrite par Struck et al. (2018), également appelée hypothèse de conservatisme de niche

écologique ou de stase morphologique par Fišer et al. (2018). Sous cette hypothèse, les espèces cryptiques maintiennent un haut degré de similarité morphologique sur des périodes très longue (ici 40-50Ma), plus longues que ce à quoi on devrait s'attendre par dérive génétique. Ce phénomène peut résulter d'un manque de variation génétique, d'une forte sélection stabilisante, de flux de gènes forts submergeant les adaptations locales, ou encore d'effet pléiotropiques contraignant l'adaptation, ou d'une combinaison de ces facteurs (Struck et al. 2018; Fišer et al. 2018). Plusieurs espèces cryptiques présentant ce phénomène de stase ont déjà été décrites, comme chez les champignons *Niphargus* (Fiser et al., 2009.), ou des amphidodes *Leucothoie* (Richards et al. 2012).

Une hypothèse expliquant la cause de la divergence entre ces deux espèces serait celle d'un saut d'hôte. En effet, lors de la divergence des deux espèces de *C. pruni* il y a 40-50 Ma, les Rosacés ont connu une forte diversification, à l'origine notamment des Prunus (dont certains hôtes de *C. pruni*) (Shi et al. 2013; Liu et al. 2019; Li et al. 2021; Liu, 2019.). La date de divergence de *C. pruni* est donc cohérente avec la date de diversification des Rosacés. Pour vérifier l'hypothèse d'un saut d'hôte, il serait intéressant de reconstruire un arbre phylogénétique daté des Psyllinae, la famille de psylles à laquelle appartient *C. pruni* et inféodée essentiellement aux Saules et Rosacées.

Bibliographie

- Ali, S. et al. 2014. « Origin, Migration Routes and Worldwide Population Genetic Structure of the Wheat Yellow Rust Pathogen *Puccinia Striiformis* f.Sp. *Tritici* ». Édité par Bruce A. McDonald. *PLoS Pathogens* 10 (1): e1003903.
- Ascunce, M S. et al. 2011. « Global Invasion History of the Fire Ant *Solenopsis Invicta* ». *Science* 331 (6020): 1066–68.
- Avise, J C, et al., 1987. « INTRASPECIFIC PHYLOGEOGRAPHY: The Mitochondrial DNA Bridge Between Population Genetics and Systematics », *Ann. Rev. Ecol. Syst.* 1987. 18:489-522
- Avise, J C. 2009. « Phylogeography: Retrospect and Prospect ». *Journal of Biogeography* 36 (1): 3–15..
- Bandelt, H. J., P. Forster, et A. Rohl. 1999. « Median-Joining Networks for Inferring Intraspecific Phylogenies ». *Molecular Biology and Evolution* 16 (1): 37–48.
- Barrès, B, et al. 2012. « Understanding the Recent Colonization History of a Plant Pathogenic Fungus Using Population Genetic Tools and Approximate Bayesian Computation ». *Heredity* 109 (5): 269–79.
- Carraro, L, et al. 1998. « Transmission of the european stone fruit yellows phytoplasma by *Cacopsylla pruni* », *Journal of Plant Pathology*, Vol. 80, No. 3, pp. 233-239
- Chen, C., E. et al. 2007. « Bayesian Clustering Algorithms Ascertain Spatial Population Structure: A New Computer Program and a Comparison Study ». *Molecular Ecology Notes* 7 (5): 747–56.
- Cornuet, J M, et al. 2014. « DIYABC v2.0: A Software to Make Approximate Bayesian Computation Inferences about Population History Using Single Nucleotide Polymorphism, DNA Sequence and Microsatellite Data ». *Bioinformatics* 30 (8): 1187–89.
- Cornuet, J M, et al. 2008. « Inferring Population History with DIY ABC: A User-Friendly Approach to Approximate Bayesian Computation ». *Bioinformatics* 24 (23): 2713–19.
- Csilléry, K., et al. 2010. « Approximate Bayesian Computation (ABC) in Practice ». *Trends in Ecology & Evolution* 25 (7): 410–18.
- Danet, J L, et al. 2011. « Multilocus sequence analysis reveals the genetic diversity of European fruit tree phytoplasmas and the existence of inter species recombination ». *Microbiology*, vol 157, issue 2

- Dayrat, B. 2005. « Towards Integrative Taxonomy: INTEGRATIVE TAXONOMY ». *Biological Journal of the Linnean Society* 85 (3): 407–15.
- Durand, E., et al. 2009. « Spatial Inference of Admixture Proportions and Secondary Contact Zones ». *Molecular Biology and Evolution* 26 (9): 1963–73.
- Durand, E., C. Chen, et O. François. 2009. « Tess Version 2.3 - Reference Manual August 2009 ».
- Earl, D.A., et B.M. vonHoldt. 2012. « STRUCTURE HARVESTER: A Website and Program for Visualizing STRUCTURE Output and Implementing the Evanno Method ». *Conservation Genetics Resources* 4 (2): 359–61.
- Estoup, A., et al. 2004. « GENETIC ANALYSIS OF COMPLEX DEMOGRAPHIC SCENARIOS: SPATIALLY EXPANDING POPULATIONS OF THE CANE TOAD, *BUFO MARINUS* ». *Evolution* 58 (9): 2021–36.
- Estoup, A., et al. 2001. « Inferring Population History From Microsatellite and Enzyme Data in Serially Introduced Cane Toads, *Bufo Marinus* ». *Genetics* 159 (4): 1671–87.
- Evanno, G., S. Regnaut, et J. Goudet. 2005. « Detecting the Number of Clusters of Individuals Using the Software Structure: A Simulation Study ». *Molecular Ecology* 14 (8): 2611–20.
- Fišer, C., C.T. Robinson, et F. Malard. 2018. « Cryptic Species as a Window into the Paradigm Shift of the Species Concept ». *Molecular Ecology* 27 (3): 613–35.
- Fiser, C., et M. Zgajmajster. 2009. « Cryptic Species from Cryptic Space: The Case of *Niphargus Fongi* Sp. n. (Amphipoda, Niphargidae) ». *Crustaceana*, Vol. 82, No. 5, pp. 593–614.
- François, O. et E. Durand. 2010. « Spatially Explicit Bayesian Clustering Models in Population Genetics: SPATIAL CLUSTERING MODELS ». *Molecular Ecology Resources* 10 (5): 773–84.
- Fresia, P., A.M.L. Azeredo-Espin, et M.L. Lyra. 2013. « The Phylogeographic History of the New World Screwworm Fly, Inferred by Approximate Bayesian Computation Analysis ». Édité par Keith A Crandall. *PLoS ONE* 8 (10): e76168.
- Guindon, S., et al. « New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0 ». *Systematic Biology* 59 (3): 307–21.
- Guindon, S., et O. Gascuel. 2003. « A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood ». Édité par Bruce Rannala. *Systematic Biology* 52 (5): 696–704.
- Hardy, G. H. 1908. « Mendelian Proportions in a Mixed Population ». *Science, New Series* 28 (706): 49–50.
- Hausdorf, B., et C. Hennig. 2010. « Species Delimitation Using Dominant and Codominant Multilocus Markers ». *Systematic Biology* 59 (5): 491–503.
- Hewitt, G. 2000. « The Genetic Legacy of the Quaternary Ice Ages ». *Nature* 405 (6789): 907–13.
- . 2001. « Speciation, Hybrid Zones and Phylogeography - or Seeing Genes in Space and Time ». *Molecular Ecology*, 10, 537 – 549.
- Hurst, G.D.D., et F.M. Jiggins. 2005. « Problems with Mitochondrial DNA as a Marker in Population, Phylogeographic and Phylogenetic Studies: The Effects of Inherited Symbionts ». *Proceedings of the Royal Society B: Biological Sciences* 272 (1572): 1525–34.
- Jombart, T. 2008. « ADEGENET: A R Package for the Multivariate Analysis of Genetic Markers ». *Bioinformatics* 24 (11): 1403–5.
- Jombart, T., et C. Collins. s. d. « A Tutorial for Discriminant Analysis of Principal Components (DAPC) Using ADEGENET 2.0.0 ».
- Jörger, K.M., et M. Schrödl. 2013. « How to Describe a Cryptic Species? Practical Challenges of Molecular Taxonomy ». *Frontiers in Zoology* 10 (1): 59.
- Kalinowski, S.T. 2004. « Counting Alleles with Rarefaction: Private Alleles and Hierarchical Sampling Designs ». *Conservation Genetics* 5 (4): 539–43.
- . 2005. « Hp-Rare 1.0: A Computer Program for Performing Rarefaction on Measures of Allelic Richness ». *Molecular Ecology Notes* 5 (1): 187–89.
- Lee, I.-M., et al. 1998. « Revised Classification Scheme of Phytoplasmas Based on RFLP Analyses of 16S RNA and Ribosomal Protein Gene Sequences ». *International Journal of Systematic Bacteriology* 48 (4): 1153–69.
- Leigh, J.W., et D. Bryant. 2015. « POPART: Full-Feature Software for Haplotype Network Construction ». Édité par Shinichi Nakagawa. *Methods in Ecology and Evolution* 6 (9): 1110–16.
- Lemoine, F., et al. 2019. « NGPhylogeny.Fr: New Generation Phylogenetic Services for Non-Specialists ». *Nucleic Acids Research* 47 (W1): W260–65.
- Letunic, I., et P. Bork. 2021. « Interactive Tree Of Life (ITOL) v5: An Online Tool for Phylogenetic Tree Display and Annotation ». *Nucleic Acids Research* 49 (W1): W293–96.
- Li, W.W., et al. 2021. « Phylogeography of *Prunus Armeniaca* L. Revealed by Chloroplast DNA and Nuclear Ribosomal Sequences ». *Scientific Reports* 11 (1): 13623.
- Liebhold, A.M., et P.C. Tobin. 2008. « Population Ecology of Insect Invasions and Their Management ». *Annual Review of Entomology* 53 (1): 387–408.
- Liu, S., s. d. « Evolutionary History of Apricot (*Prunus Armeniaca* L.) and Impact of Different Processes of Evolution on Genetic Diversity », 225.
- Liu, S., et al. 2019. « The Complex Evolutionary History of Apricots: Species Divergence, Gene Flow and Multiple Domestication Events ». *Molecular Ecology* 28 (24): 5299–5314.
- Mallez, S., et al. 2021 « Inference of the Worldwide Invasion Routes of the Pinewood Nematode *Bursaphelenchus Xylophilus* Using Approximate Bayesian Computation Analysis », bioRxiv, 452326, version 6.

- Marie-Jeanne, et al. 2020. « Multi-Scale Spatial Genetic Structure of the Vector-Borne Pathogen ‘Candidatus Phytoplasma Prunorum’ in Orchards and in Wild Habitats ». *Scientific Reports* 10 (1): 5002.
- Nei, M., et R. K. Chesser. 1983. « Estimation of Fixation Indices and Gene Diversities ». *Annals of Human Genetics* 47 (3): 253–59.
- Nielsen, R., et M A. Beaumont. 2009. « Statistical Inferences in Phylogeography ». *Molecular Ecology* 18 (6): 1034–47.
- Pascual, M., et al. 2007. « Introduction History of *Drosophila Subobscura* in the New World: A Microsatellite-Based Survey Using ABC Methods: INTRODUCTION HISTORY OF *D.SUBOBSCURA* ». *Molecular Ecology* 16 (15): 3069–83.
- Peccoud, J., G. Labonne, et N. Sauvion. 2013. « Molecular Test to Assign Individuals within the *Cacopsylla Pruni* Complex ». Édité par Valerio Ketmaier. *PLoS ONE* 8 (8): e72454.
- Peccoud, J., D R J. Pleydell, et N. Sauvion. 2018. « A Framework for Estimating the Effects of Sequential Reproductive Barriers: Implementation Using Bayesian Models with Field Data from Cryptic Species: BRIEF COMMUNICATION ». *Evolution* 72 (11): 2503–12.
- Percy, D M. 2003. « Radiation, diversity, and host-plant interactions among island and continental legume-feeding psyllids ». *Evolution* 57 (11): 2540–56.
- Pritchard, J K, M. Stephens, et P. Donnelly. 2000. « Inference of Population Structure Using Multilocus Genotype Data ». *Genetics* 155 (2): 945–59.
- Richards, V P, M J Stanhope, et M S Shivji. 2012. « Island Endemism, Morphological Stasis, and Possible Cryptic Speciation in Two Coral Reef, Commensal Leucothoid Amphipod Species throughout Florida and the Caribbean ». *Biodivers Conserv*, 21:343–361.
- Roderick, G K. 1996. « Geographic Structure of Insect Populations: Gene Flow, Phylogeography, and Their Uses », *Annu. Rev. Entomol.* 41:325-52
- Rousset, F. 2020. « Genepop 4.7 ».
- Rousset, F. 2008. « Genepop’007: A Complete Re-Implementation of the Genepop Software for Windows and Linux ». *Molecular Ecology Resources* 8 (1): 103–6.
- Rozas, J., et al. 2017. « DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets ». *Molecular Biology and Evolution* 34 (12): 3299–3302.
- Sauvion, N., et al. 2009. « Nine Polymorphic Microsatellite Loci from the Psyllid *Cacopsylla Pruni* (Scopoli), the Vector of European Stone Fruit Yellows ». *Molecular Ecology Resources* 9 (4): 1196–99.
- Sauvion, N., et al. 2021. « Occurrence Data for the Two Cryptic Species of *Cacopsylla Pruni* (Hemiptera: Psylloidea) ». *Biodiversity Data Journal* 9 (juillet): pp.e68860.
- Shi, S, et al. 2013. « Phylogeny and Classification of *Prunus Sensu Lato* (Rosaceae): Phylogeny and Classification of *Prunus* ». *Journal of Integrative Plant Biology* 55 (11): 1069–79
- Sim, Z., et al. 2016. « Genome-Wide Set of SNPs Reveals Evidence for Two Glacial Refugia and Admixture from Postglacial Recolonization in an Alpine Ungulate ». *Molecular Ecology* 25 (15): 3696–3705.
- Sjöstrand, A E, P. Sjödin, et M. Jakobsson. 2014. « Private Haplotypes Can Reveal Local Adaptation ». *BMC Genetics* 15 (1): 61.
- Song, H., et al. 2008. « Many Species in One: DNA Barcoding Overestimates the Number of Species When Nuclear Mitochondrial Pseudogenes Are Coamplified ». *Proceedings of the National Academy of Sciences* 105 (36):
- Song, W., et al. 2018. « Multiple Refugia from Penultimate Glaciations in East Asia Demonstrated by Phylogeography and Ecological Modelling of an Insect Pest ». *BMC Evolutionary Biology* 18 (1): 152.
- Stone, et al. 2007. « The Phylogeographical Clade Trade: Tracing the Impact of Human-Mediated Dispersal on the Colonization of Northern Europe by the Oak Gallwasp *Andricus Kollari* ». *Molecular Ecology* 16 (13): 2768–81.
- Struck, et al. 2018. « Finding Evolutionary Processes Hidden in Cryptic Species ». *Trends in Ecology & Evolution* 33 (3): 153–63.
- Sunnåker, et al. 2013. « Approximate Bayesian Computation ». Édité par Shoshana Wodak. *PLoS Computational Biology* 9 (1): e1002803.
- Thornton, K., et P. Andolfatto. 2006. « Approximate Bayesian Inference Reveals Evidence for a Recent, Severe Bottleneck in a Netherlands Population of *Drosophila Melanogaster* ». *Genetics* 172 (3): 1607–19.
- Wahlund, S. 1928. « Zusammensetzung von populationen und korrelationserscheinungen vom standpunkt der vererbungslehre aus betrachtet ».
- Wegmann, D., et L. Excoffier. 2010. « Bayesian Inference of the Demographic History of Chimpanzees ». *Molecular Biology and Evolution* 27 (6): 1425–35.
- Weinberg, W. 1908. « Ober Vererbungsgesetze beim Menschen. ».
- Woolley, S M., D. Posada, et K A. Crandall. 2008. « A Comparison of Phylogenetic Network Methods Using Computer Simulation ». Édité par Jason E. Stajich. *PLoS ONE* 3 (4): e1913.
- Wright, S. 1930. « Evolution in mendelian population ». *Genetics* 16.
- . 1938. « The Distribution of Gene Frequencies Under Irreversible Mutation ». *Proceedings of the National Academy of Sciences* 24 (7): 253–59.
- . 1942. « Isolation by distance ». *Genetics* 28.
- . 1940. « Breeding Structure of Populations in Relation to Speciation ». *The American Naturalist* 74 (752): 232–48.
- Zepeda-Paulo, et al. 2010. « The Invasion Route for an Insect Pest Species: The Tobacco Aphid in the New World ». *Molecular Ecology* 19 (21): 4738–52.

Annexes :

• Annexe 1 : Tableau récapitulatif des données

pop	localite	groupe	latitude	longitude	nombre de séquences microsattellites*	nombre de séquences ITS2	nombre de séquences COI-t
1	Larzac	A	43,9279	3,3117	62	37	0
2	Coursegoules	A	43,79	2,91	12	17	11
2	Coursegoules	B	43,79	2,91	5	7	0
3	Ht_Languedoc	A	3,3679	2,3016	5	8	0
3	Ht_Languedoc	B	3,3679	2,3016	9	14	0
4	La_Tieule	A	44,387	3,1243	28	23	0
4	La_Tieule	B	44,387	3,1243	2	4	0
5	Montgame	A	46,7338	0,5102	0	1	0
5	Montgame	B	46,7338	0,5102	7	25	0
6	Prades	A	42,6202	2,4356	11	13	0
6	Prades	B	42,6202	2,4356	0	2	0
7	PradesLez	A	3,7049	3,8502	8	30	0
7	PradesLez	B	3,7049	3,8502	2	4	0
8	Torreilles	A	42,7414	2,985	9	13	7
8	Torreilles	B	42,7414	2,985	3	14	1
9	Fesche	A	44,6528	4,4443	7	9	5
9	Fesche	B	44,6528	4,4443	0	3	0
10	Bellac	B	46,1068	1,0794	2	3	0
11	Romette	A	44,581994	6,106769	2	0	0
11	Romette	B	44,581994	6,106769	1	1	0
12	Udine	B	45,9235	9,5054	7	20	10
13	Cacak	B	43,8915	20,3502	12	9	3
14	Neustadt	B	49,3502	8,1487	13	0	5
15	Hattonville	B	48,9924	5,7145	10	18	4
16	Vouillon	B	46,8257	1,9203	21	14	5
17	Mirepoix	A	43,158472	1,988042	5	0	0
17	Mirepoix	B	43,158472	1,988042	12	3	0
18	Angers	A	47,3556	-0,5444	12	5	3
18	Angers	B	47,3556	-0,5444	0	0	1
19	sefra	A	44,8351	4,8911	18	16	0
19	sefra	B	44,8351	4,8911	2	4	0
20	Seysse	B	43,491	1,3193	7	8	6
21	Moia	A	41,8249	2,194	21	10	3
22	San_Pau	A	42,2577	2,3609	23	18	0
23	Col_d_Ares	A	42,3763	2,472	19	16	0
23	Col_d_Ares	B	42,3763	2,472	1	0	0
24	Col_de_Jau	A	42,6881	2,2509	2	1	0
24	Col_de_Jau	B	42,6881	2,2509	12	32	0
25	Lusignan	B	46,4303	0,1101	23	23	6
26	Aigoual	B	44,0292	3,4853	21	0	5
27	Seranne	A	43,7884	3,5433	22	12	0
27	Seranne	B	43,7884	3,5433	21	0	5
28	Grabels	A	43,6598	3,8202	69	27	0
28	Grabels	B	43,6598	3,8202	14	14	4
29	Levernois	B	47,0024	4,8472	18	5	5
30	Farcheville	A	48,4104	2,2889	13	11	4
30	Farcheville	B	48,4104	2,2889	6	8	3
31	Bouleternere	A	42,6586	2,5897	17	20	0
31	Bouleternere	B	42,6586	2,5897	2	3	0
32	Versailles	A	48,8453	2,0027	3	0	3

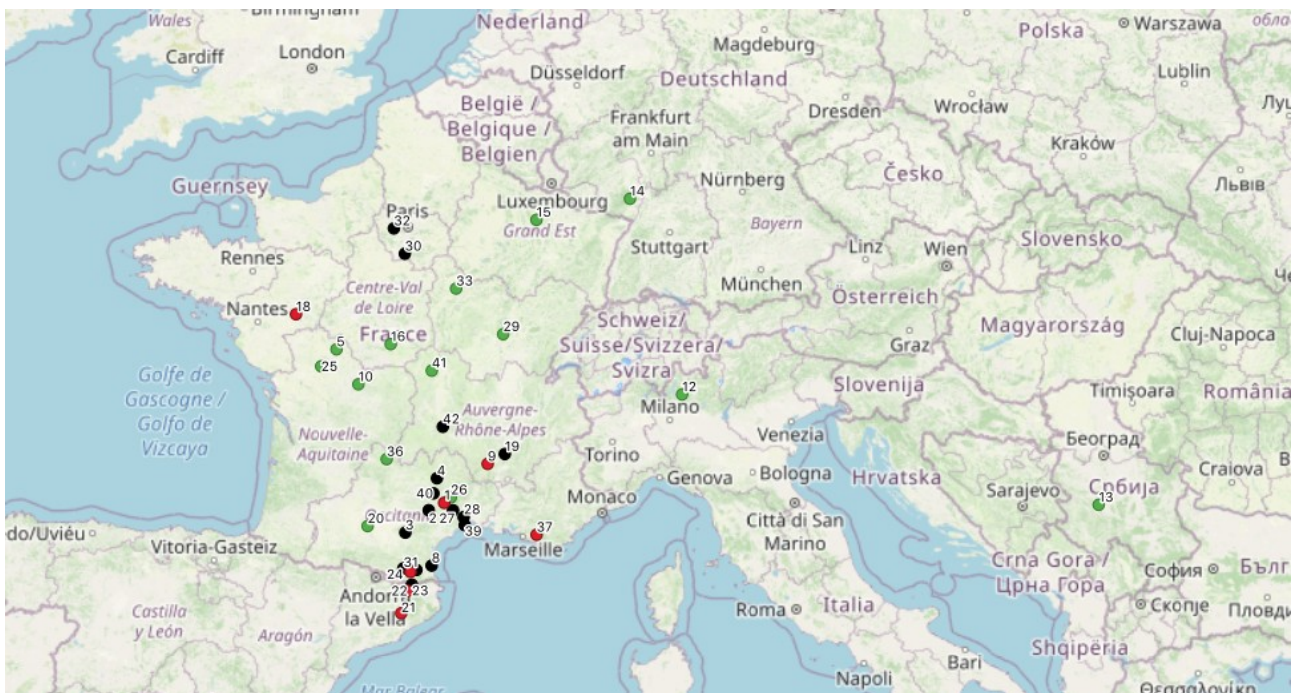
32	Versailles	B	48,8453	2,0027	14	14	6
33	Auxerre	B	47,8056	3,6232	21	28	3
34	Monterniers	A	46,161694	4,623819	2	0	1
34	Monterniers	B	46,161694	4,623819	6	3	0
35	Col_Crie	A	46,206667	4,540336	1	1	1
35	Col_Crie	B	46,206667	4,540336	20	5	0
36	Gramat	B	44,7408	1,8151	10	4	0
37	Ste_Baume	A	43,3274	5,7176	9	6	4
38	Communay	A	45,615842	4,850328	1	0	0
38	Communay	B	45,615842	4,850328	8	6	1
39	Maguelonne	A	43,5098	3,8468	5	7	0
39	Maguelonne	B	43,5098	3,8468	6	13	1
40	Millau	A	44,1053	3,032	18	11	0
40	Millau	B	44,1053	3,032	1	1	0
41	Montmarault	B	46,3524	2,9896	16	21	5
42	Lorlanges	A	45,3361	3,2774	1	1	2
42	Lorlanges	B	45,3361	3,2774	9	10	0
43	Pont_du_Chtx	B	45,803069	3,272247	9	8	0
44	Tordera	A	41,7725	2,7593	0	5	5
44	Tordera	B	41,7725	2,7593	0	18	0
45	Firmi	B	44,5468	2,3052	0	4	2
46	Montagnac	A	43,4811	3,4675	0	3	5
46	Montagnac	B	43,4811	3,4675	0	4	5
47	Thezan	A	43,1067	2,7447	0	3	6
47	Thezan	B	43,1067	2,7447	0	5	5
48	Palalda	A	42,4959	2,6683	0	5	5
49	Boule_d_Amont	A	42,5815	2,6174	0	0	5
50	Cuq_Toulza	B	43,5869	1,8487	0	9	5
51	Barthe	B	43,0786	0,4403	0	9	5
52	Orthez	B	43,4457	-0,7082	0	16	5
53	St_Sever	B	43,7809	-0,5868	0	9	5
54	St_J_Pied_Port	A	43,2419	-1,3019	0	8	10
54	St_J_Pied_Port	B	43,2419	-1,3019	0	0	3
55	Camargue	B	43,0884	1,2881	0	8	5
56	Limoux	A	43,0559	2,1316	0	0	5
57	Nyons	A	44,3294	5,1071	0	0	6
58	La_Souterraine	B	46,2279	1,456	0	7	5
59	Ile_de_Ré	B	46,187	-1,405	0	3	8
60	Moreilles	B	46,187	-1,405	0	1	1
61	Jaulnay	A	46,9291	0,4274	0	0	5
62	Le_Mans	A	48,0061	0,3579	0	0	5
63	Orleans	A	47,8771	1,5647	0	3	9
63	Orleans	B	47,8771	1,5647	0	1	4
64	Ludon	B	44,9942	-0,6072	0	0	10
65	Velesovice	B	49,1746	16,861	0	0	5
66	Mrsinci	B	43,8042	20,4919	0	0	6
67	Villenave_d_Ornon	A	44,7764	-0,5829	0	0	10
67	Villenave_d_Ornon	B	44,7764	-0,5829	0	0	1
68	La_Cavalerie	A	44,0124	3,1269	0	0	5
69	Guillemestre	A	44,6603	6,6053	0	0	5
70	Magara	B	36,7026	33,9083	0	0	25
71	Tautavel	A	42,7941	2,6985	0	0	5
71	Tautavel	B	42,7941	2,6985	0	0	12
72	Almese	B	45,1167	7,3953	0	0	5
73	Austria	B	48,3155	16,3839	0	0	2
74	Bosnia	B	44,7667	17,1833	0	0	5
75	Bursa	B	40,1918	28,9854	0	0	10
76	Orhaneli	B	40,0365	28,8964	0	0	6
77	Keles	B	39,8957	29,2218	0	0	2
78	Pozanti	B	36,6273	33,8741	0	0	22
79	Ayvali	B	40,6206	42,0117	0	0	13
80	Montrichard	A	47,301442	1,183114	0	0	7

81	Ponti	A	44,5293	8,3125	0	0	8
82	Loeche	A	46,380024	7,628834	0	0	8
82	Loeche	B	46,380024	7,628834	0	0	8
83	Rouen	A	49,3045	1,1285	0	0	3
84	Mantes la Jolie	A	48,9861	1,6891	0	0	3
84	Mantes la Jolie	B	48,9861	1,6891	0	0	8
85	Langon	A	44,5024	-0,2579	0	0	8
86	Biarritz	A	43,3627	-1,5155	0	0	8
87	Santander	A	43,344	-3,8201	0	0	8
88	Quiroga	A	42,4761	-7,3388	0	0	8
89	Venta de Banos	A	41,9333	-4,5016	0	0	8
90	Alfaro	A	42,1365	-1,78	0	0	7
91	Meckenheim	B	49,4022	8,2186	0	0	8
92	Ventiseri	B	41,954917	9,382633	0	0	8
93	Vescovato2	B	42,53065	9,4806	0	0	7
94	Brest	B	48,140198	-4,204939	0	0	8
95	Pomáz	B	47,651035	19,03957	0	0	8
96	Ljabojno	B	40,89666	21,14521	0	0	8
97	Beutenaken	B	50,77764	5,856584	0	0	8

Tableau récapitulatif des données par groupe et population.

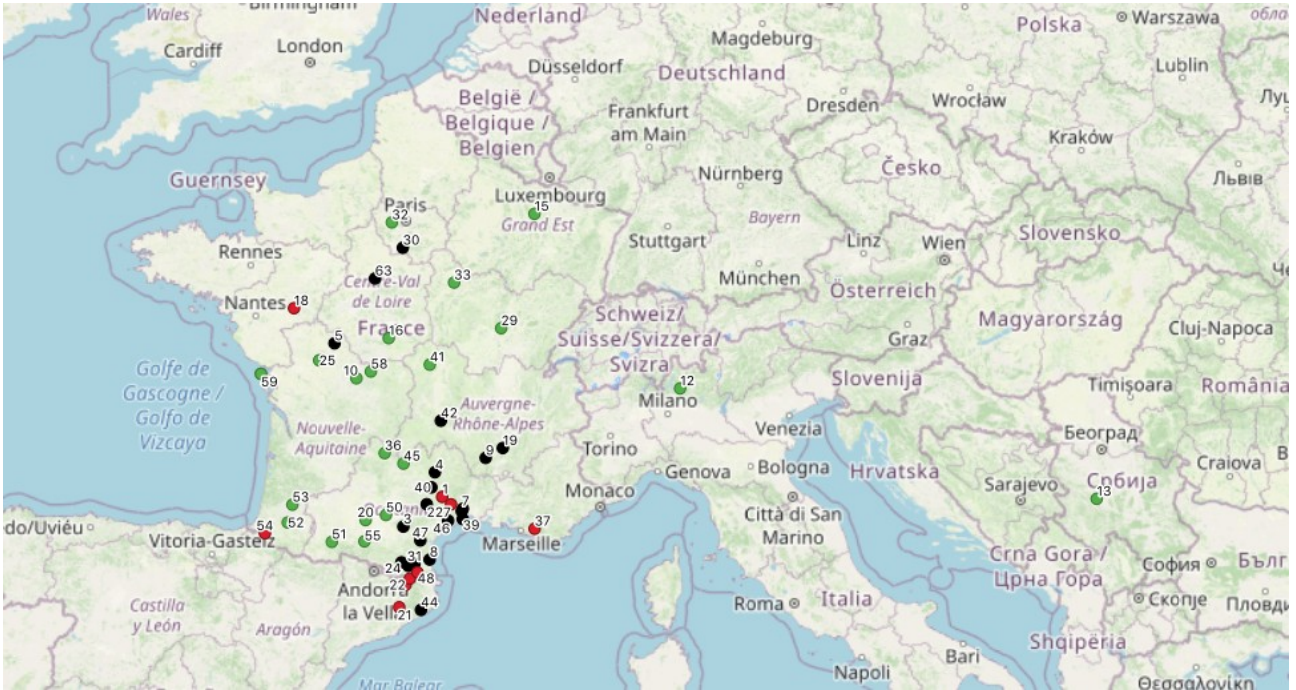
* nombre d'individus pour lequel on a les séquences des 8 marqueurs microsatellites

● **Annexe 2 : Cartes de répartition des différents sites d'échantillonnage pour les différents marqueurs.**



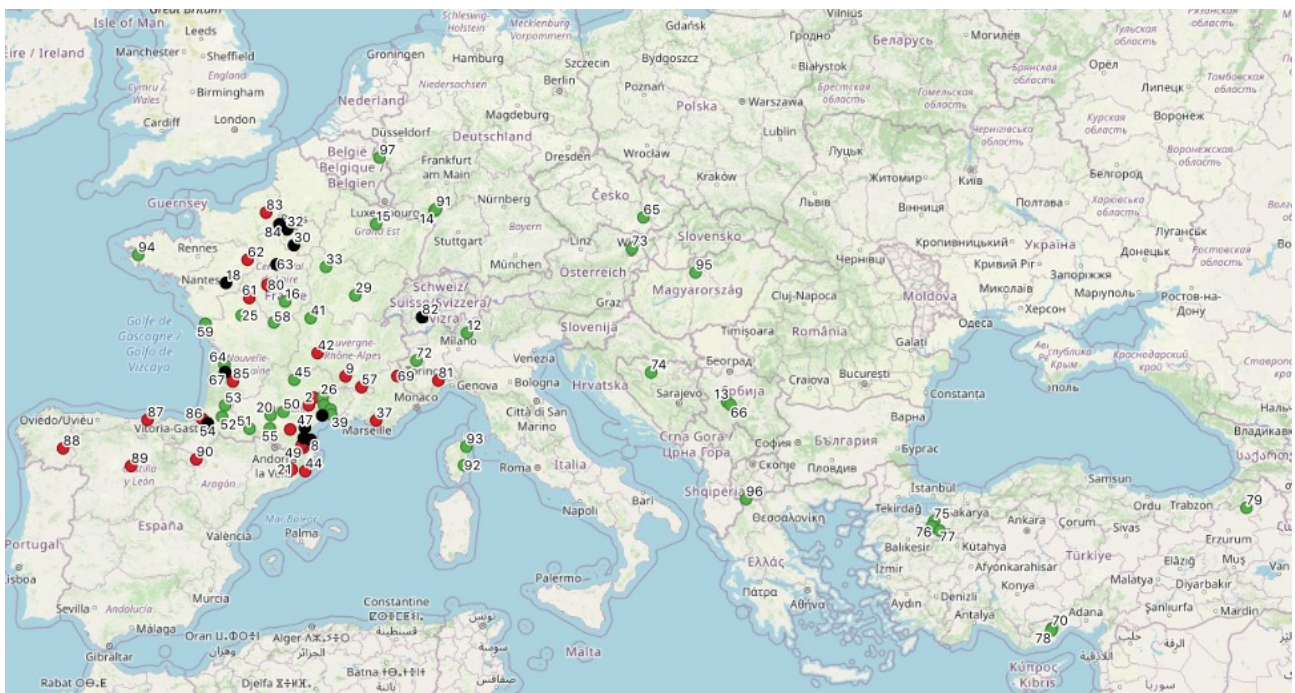
Carte de la répartition des différents sites d'échantillonnages des données microsatellites

(Les points rouges représentent les populations pour lesquelles seuls des individus A ont été échantillonnés, les points verts représentent celles pour lesquelles seuls des individus B ont été échantillonnés et les points noirs représentent celles pour lesquelles des individus A et B ont été échantillonnés.)



Carte de la répartition des différents sites d'échantillonnages des données ITS2

(Les points rouges représentent les populations pour lesquelles seuls des individus A ont été échantillonnés, les points verts représentent celles pour lesquelles seuls des individus B ont été échantillonnés et les points noirs représentent celles pour lesquelles des individus A et B ont été échantillonnés.)



Carte de la répartition des différents sites d'échantillonnages des données COI-t

(Les points rouges représentent les populations pour lesquelles seuls des individus A ont été échantillonnés, les points verts représentent celles pour lesquelles seuls des individus B ont été échantillonnés et les points noirs représentent celles pour lesquelles des individus A et B ont été échantillonnés.)

● **Annexe 3 : Nombre Burnin, itérations, kmax et runs pour Structure et TESS**

Logiciel	Groupe testé	Burnin (phase d'initiation)	Nombre d'itérations de la MCMC	Kmax	Nombre d'essais (run)
STRUCTURE	A+B	10000	20000	2	20
	A	10000	20000	5	50
	B	10000	20000	5	50
TESS	A+B	10000	20000	De 2 à 10	De 20 à 50
	A	10000	20000	De 2 à 10	De 20 à 50
	B	10000	20000	De 2 à 10	De 20 à 50

Tableau récapitulatifs des analyses de clustering bayésien réalisées sur STRUCTURE et TESS et des paramètres associés entrés dans le logiciels

Pour toutes ces analyses, le Brunin (phase d'initialisation) et le nombre d'itérations (Number of MCMC Reps after Burnin) sont fixés en suivant François et Durand (2010). On observe que les modèles convergent rapidement, ce qui semble indiquer que ces valeurs sont suffisantes.

● **Annexe 4 : Statistiques résumées utilisées pour l'ABC**

	Microsatellites	Séquences d'ADN (ITS2 ou COI)
Statistiques résumées pour un échantillon	<ul style="list-style-type: none"> • Nombre moyen d'allèles • Diversité génétique moyenne • Moyenne des variances de taille • Moyenne des M de Garza- Williamson 	<ul style="list-style-type: none"> • Nombres d'haplotypes • Nombre de sites de ségrégation • Moyenne des différences par paire • Variance des différences par paire • Sites ségrégeant privés • Moyenne des nombres du nucléotide le plus rare aux sites de ségrégation • Variance des nombres du nucléotide le plus rare au site de ségrégation • D de Tajima
Statistiques résumées pour deux échantillons	<ul style="list-style-type: none"> • Nombre moyen d'allèles • Diversité génétique moyenne • Moyenne des variances de taille • Fst • Indice de Classification • Distances alléliques partagées • Distance $(d\mu)^2$ 	<ul style="list-style-type: none"> • Nombres d'haplotypes • Nombre de sites de ségrégation • Moyenne des différences par paires (W) • Moyenne des différences par paire (B) • Fst (Hudson et al., 1992)

Tableau récapitulatifs des statistiques résumées utilisées dans les analyses ABC.

On note que pour les tests sur une seule population avec microsatellites+ITS2, 12 statistiques résumées sont utilisées, pour les tests sur 2 populations avec les mêmes marqueurs, 24 sont utilisées, et pour ceux à 4 populations pour COI-t seul, 62 sont utilisées

• **Annexe 5 : Méthodologie et détails des scénarios de l'ABC**

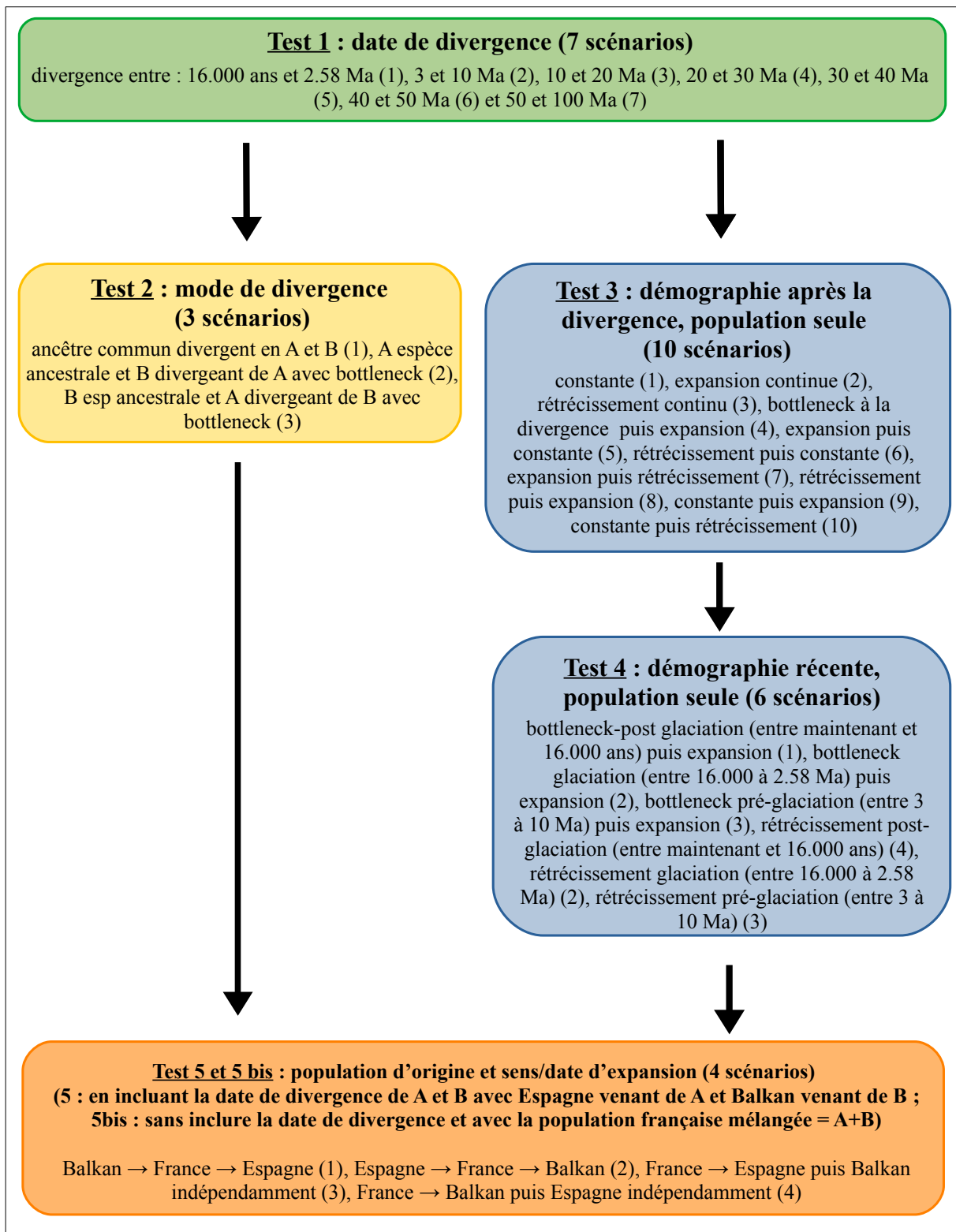


Schéma récapitulatif de la méthodologie et des différents tests et scénarios réalisés au sein des analyses ABC.

● Annexe 6 : Priors des différents tests de l'ABC

Étape	Paramètres	Nom du paramètre	distribution	Intervalle [min-max]	Conditions
étape 1	effectif efficace de la population ancestrale	N1+N2	Log-uniforme	[20-100.000]	-
	effectif efficace de la population A après divergence	N1	Log-uniforme	[10-50.000]	-
	effectif efficace de la population B après divergence	N2	Log-uniforme	[10-50.000]	-
	date de divergence testée de A & B	t1	Log-uniforme	[16.000-2.580.000]	-
	date de divergence testée de A & B	t2	Log-uniforme	[3.000.000-10.000.000]	-
	date de divergence testée de A & B	t3	Log-uniforme	[10.000.000-20.000.000]	-
	date de divergence testée de A & B	t4	Log-uniforme	[20.000.000-30.000.000]	-
	date de divergence testée de A & B	t5	Log-uniforme	[30.000.000-40.000.000]	-
	date de divergence testée de A & B	t6	Log-uniforme	[40.000.000-50.000.000]	-
	date de divergence testée de A & B	t7	Log-uniforme	[50.000.000-100.000.000]	-
étape 2	effectif efficace de la population ancestrale du sc1	N1+N2	Log-uniforme	[20-100.000]	-
	effectif efficace de la population A	N1, N3, N5	Log-uniforme	[10-50.000]	-
	effectif efficace de la population B	N2, N4, N6	Log-uniforme	[10-50.000]	-
	effectif efficace pendant le bottleneck après la divergence	N4b, N5b	Log-uniforme	[10-1.000]	N4b<N4&N3, N5b<N5&N6
	date de divergence de A & B	tdiv	Log-uniforme	[30.000.000-65.000.000]	-
	date de fin du bottleneck	t1	Log-uniforme	[1-65.000.000]	t1<tdiv
étape 3	effectifs efficaces de la population	N1, N2, N3, N4, N5, N7, N8, N9, N10, N11, N12, N13, N2b, N4b	Log-uniforme	[10-50.000]	N3<N2<N1, N1<N4<N5, N8<N1, N1<N9, N2b>N1&N10, N4b<N1&N11, N12<N1, N13>N1
	effectif efficace de la population pendant le bottleneck	N6	Log-uniforme	[10-100]	N6<N1 & N7
	date de divergence de A & B	tdiv	Log-uniforme	[30.000.000-65.000.000]	-
	dates d'évènements démographiques	tc, te, tr, tb	Log-uniforme	[1-65.000.000]	tc, te, tr & tb<tdiv
	effectifs efficaces de la population	Navdiv, N1, Nexp	Log-uniforme	[10-50.000]	Nexp>Navdiv&N1
étape 4	effectif efficace de la population pendant le bottleneck	N2	Log-uniforme	[10-100]	N2<Nexp&N1
	date de divergence de A & B	tdiv	Log-uniforme	[30.000.000-65.000.000]	-
	dates d'évènements démographiques	t1, tf1	Log-uniforme	[1-16.000]	t1>tf1
	dates d'évènements démographiques	t2, tf2	Log-uniforme	[3.000.000-10.000.000]	t2>tf2
	dates d'évènements démographiques	tg, tfg	Log-uniforme	[16.000-2.580.000]	tg>tfg
	effectifs efficaces des populations	N1, N2, N3, N4	Log-uniforme	[10-50.000]	-
étape 5	effectifs efficaces des populations	N2a, N3a, N4a, N1b, N2b, N3b, N1c, N4c, N1d, N4d	Log-uniforme	[10-1.000]	N4a&N3b&N4c&N4d<N2, N2a&N1b&N1c&N1d<N3, N3a<N1, N2b<N4
	date de divergence de A & B	tdiv	Log-uniforme	[30.000.000-65.000.000]	-
	dates d'évènements démographiques et de divergences	ta,b, t1a,b, t2a,b, t3a,b, t4a,b	Log-uniforme	[1-100.000.000]	ta,b<t1a,b<t2a,b<tdiv<t3a,b<t4a,b
	dates d'évènements démographiques et de divergences	tc,d, t1c,d t2c,d, t3c,d	Log-uniforme	[1-65.000.000]	tc,d<t1c,d<t2c,d<tdiv
	effectifs efficaces des populations	N1, N2, N3	Log-uniforme	[10-50.000]	-
étape 5bis	effectifs efficaces des populations	N1b, N3b, N2ba, N2bb	Log-uniforme	[10-1.000]	N3b&N1b<N2, N2ba<N1, N2bb<N3
	dates d'évènements démographiques et de divergences	ta,b,c,d, t1a,b,c,d, t2a,b,c,d, t3a,b,c,d	Log-uniforme	[1-100.000.000]	ta,b,c,d<t1a,b,c,d<t2a,b,c,d<t3a,b,c,d

Tableau récapitulatifs des priors utilisés dans l'analyse ABC pour les différentes étapes et de leur distribution. - = pas de données

● **Annexe 7 : Tests exacts de Hardy Weinberg, et Rho-statistiques, détails par locus**

Locus	Groupe	Chi2	Df	Proba	Rhois	Rhost	Rhoit
Cp6_115	AB	> 331.4809	82	< 1.01e-31	0.5467	0.1519	-0.6155
	A	> 226.5947	50	< 2.55e-24	0.5467	0.0232	0.4552
	B	55.5346	54	0.416697	0.5467	-0.0155	0.5941
Cp5_45	AB	> 314.2285	74	< 2.35e-31	0.1031	0.0473	0.1455
	A	64.6108	48	0.055017	0.0307	0.0127	0.0431
	B	49.4848	28	0.007378	0.6557	0.0506	0.6731
Cp6_144	AB	> 917.6987	86	< 0.00e+00	0.8400	0.0756	0.8521
	A	> 564.6476	46	< 0.00e+00	0.8853	0.0224	0.8879
	B	> 356.6808	62	< 5.51e-43	0.4680	0.0610	0.5004
Cp4_127	AB	> 792.3536	86	< 0.00e+00	0.8426	0.5815	0.9342
	A	24.3395	50	0.999169	-0.1015	0.0007	-0.1023
	B	> 703.9771	66	< 0.00e+00	0.8193	0.0024	0.8198
Cp5_43	AB	> 649.6930	86	< 0.00e+00	0.1603	0.0558	0.2072
	A	> 308.5435	48	< 9.73e-40	0.2722	0.0439	0.3042
	B	147.9446	62	5.52e-09	0.1090	-0.0081	0.1018
Cp6_15	AB	> 380.2536	86	< 1.28e-38	0.1970	0.2333	0.3843
	A	> 66.7188	50	< 0.057034	-0.0025	-0.0050	-0.0075
	B	> 218.1699	66	< 3.65e-18	0.3138	0.0168	0.3253
Cp6_129	AB	233.9806	84	4.42e-16	0.3620	0.0229	0.3766
	A	15.6427	38	0.999501	0.0509	0.0032	0.0539
	B	228.3937	56	1.10e-22	0.6748	-0.0237	0.6671
Cp4_108	AB	> 787.0020	86	< 0.00e+00	0.5058	0.1174	0.5639
	A	> 460.1875	50	< 0.00e+00	0.6308	0.0203	0.6382
	B	> 276.9084	66	< 1.22e-27	0.4165	0.0271	0.4324
All locus, All pop	AB	> 3582.8118	670	High. sign	0.5625	0.3947	0.7352
	A	> 1403.5052	380	High. sign	0.3511	0.0132	0.3597
	B	> 1771.3262	460	High. sign	0.4590	0.0079	0.4633

Tableau des résultats par Locus, pour toutes les populations, du test exact de Hardy-Weinberg (test probabilité) et des Rho-statistiques pour la population totale AB (noir), A (rouge) et B (vert).

Df = degré de liberté ; Proba = probabilité

● **Annexe 8 : Tests de Mantel, isolation par la distance (IBD)**

Population	Intercept (a)	Pente (b)	Pr(corrélation > corrélation observée)	Pr(corrélation < corrélation observée)
AB	0,071162	0,0123314	0 sous l'hypothèse nulle	1 sous l'hypothèse nulle
A	0,00612135	-0,004784	0,76	0,24
B (totaux)	-0,000401078	0,000997877	0 sous l'hypothèse nulle	1 sous l'hypothèse nulle
B (France)	0,00546009	-0,00419607	0,772	0,228

Tableau des résultats des tests de Mantel (IBD) pour les différents groupes étudiés. L'intercept a et la pente b donnent une courbe de la forme $F_{st}/(1-F_{st}) = a + b \ln(\text{distance})$. Les deux Probabilités du tableau sont unilatérales et résultent du test de Mantel pour 1000 permutations. Pr = probabilité

● **Annexe 9 : Détails des tests de choix de scénarios dans DIYABC**

Les tableaux suivants donnent la probabilité à postériori de chaque scénario (colonnes) pour chaque population testée (lignes), ainsi que son minimum et maximum dans les différents jeux de données inférés. Le code couleur suit la légende suivante :

	: scénario gagnant
	: scénario sous 5%
	: scénario sous 10%
x	inutilisable

Tests de l'étape 1 : date de divergence de *C. pruni*

Scénarios / pop	1 – 16.000 à 2.58Ma	2 – 3Ma à 10Ma	3 – 10Ma à 20Ma	4 – 20Ma à 30Ma	5 – 30Ma à 40Ma	6 – 40Ma à 50Ma	7 – 50Ma à 100Ma
28	0.0000 [0.0000,0.0195]	0.0002 [0.0000,0.0195]	0.0103 [0.0000,0.0294]	0.1080 [0.0000,0.2391]	0.0613 [0.0000,0.1399]	0.7921 [0.6103,0.9739]	0.0281 [0.0000,0.0667]
3	0.0000 [0.0000,0.5470]	0.0088 [0.0000,0.5494]	0.1198 [0.0000,0.8186]	0.5315 [0.1822,0.8808]	0.0028 [0.0000,0.5469]	0.2567 [0.0000,0.6501]	0.0804 [0.0000,0.5488]
2	0.0000 [0.0000,0.0075]	0.0004 [0.0000,0.0074]	0.0005 [0.0000,0.0073]	0.0371 [0.0000,0.1684]	0.0006 [0.0000,0.0073]	0.9303 [0.7259,1.0000]	0.0311 [0.0000,0.1477]
39	0.0000 [0.0000,0.0968]	0.0002 [0.0000,0.0969]	0.0091 [0.0000,0.1031]	0.0469 [0.0000,0.1391]	0.0186 [0.0000,0.1094]	0.6411 [0.4009,0.8813]	0.2840 [0.0440,0.5241]
1 & 25	0.0000 [0.0000,0.1014]	0.0001 [0.0000,0.1014]	0.0012 [0.0000,0.1016]	0.1066 [0.0000,0.2383]	0.2623 [0.0115,0.5132]	0.6007 [0.3187,0.8828]	0.0290 [0.0000,0.1236]
19 & 15	0.0000 [0.0000,0.4835]	0.0013 [0.0000,0.4836]	0.0169 [0.0000,0.4870]	0.2922 [0.0000,0.7058]	0.0100 [0.0000,0.4856]	0.1753 [0.0000,0.7433]	0.5044 [0.1378,0.8709]
21 & 24	0.0000 [0.0000,0.3029]	0.0670 [0.0000,0.3347]	0.3103 [0.0000,0.6695]	0.1359 [0.0000,0.3903]	0.0279 [0.0000,0.3112]	0.1303 [0.0000,0.5099]	0.3286 [0.0000,0.6629]
23 & 33	0.0000 [0.0000,1.0000]	0.0174 [0.0000,1.0000]	0.0001 [0.0000,1.0000]	0.0001 [0.0000,1.0000]	0.9805 [0.8818,1.0000]	0.0017 [0.0000,1.0000]	0.0001 [0.0000,1.0000]
30	x	x	x	x	x	x	x
1 & 15	0.0000 [0.0000,0.2163]	0.0090 [0.0000,0.2165]	0.0069 [0.0000,0.2159]	0.3337 [0.0000,0.7341]	0.0999 [0.0000,0.3063]	0.5123 [0.1102,0.9143]	0.0381 [0.0000,0.2386]
1 & 24	0.0000 [0.0000,0.0988]	0.0044 [0.0000,0.0999]	0.0161 [0.0000,0.1052]	0.2624 [0.0327,0.4921]	0.1149 [0.0000,0.2400]	0.5241 [0.2647,0.7836]	0.0780 [0.0000,0.18650]
1 & 33	0.0000 [0.0000,0.2278]	0.0018 [0.0000,0.2278]	0.0187 [0.0000,0.2302]	0.3198 [0.0017,0.6378]	0.0423 [0.0000,0.2419]	0.3712 [0.0579,0.6844]	0.2463 [0.0000,0.5497]
19 & 25	0.0000 [0.0000,0.1375]	0.0001 [0.0000,0.1375]	0.0219 [0.0000,0.1512]	0.0633 [0.0000,0.1981]	0.0228 [0.0000,0.1496]	0.6066 [0.2674,0.9458]	0.2853 [0.0000,0.6177]
19 & 24	0.0000 [0.0000,0.8129]	0.0168 [0.0000,0.8212]	0.0200 [0.0000,0.8224]	0.0326 [0.0000,0.8288]	0.0222 [0.0000,0.8220]	0.1548 [0.0000,1.0000]	0.7543 [0.4913,1.0000]
19 & 33	0.0000 [0.0000,1.0000]	0.0000 [0.0000,1.0000]	0.0151 [0.0000,1.0000]	0.0894 [0.0000,1.0000]	0.0221 [0.0000,1.0000]	0.1273 [0.0000,1.0000]	0.7460 [0.3442,1.0000]
21 & 25	0.0000 [0.0000,0.3098]	0.0000 [0.0000,0.3098]	0.1205 [0.0000,0.3667]	0.0794 [0.0000,0.3084]	0.2206 [0.0000,0.5726]	0.2908 [0.0000,0.6969]	0.2887 [0.0000,0.6949]
21 & 15	0.0000 [0.0000,0.5339]	0.0003 [0.0000,0.5339]	0.0479 [0.0000,0.5503]	0.0939 [0.0000,0.5706]	0.0464 [0.0000,0.5463]	0.2400 [0.0000,0.8582]	0.5715 [0.1888,0.9542]
21 & 33	0.0000 [0.0000,0.0159]	0.0000 [0.0000,0.0159]	0.0126 [0.0000,0.0341]	0.0893 [0.0000,0.2438]	0.0127 [0.0000,0.0358]	0.0160 [0.0000,0.0439]	0.8694 [0.6747,1.0000]
23 & 25	0.0000 [0.0000,0.0111]	0.0001 [0.0000,0.0111]	0.0102 [0.0000,0.0331]	0.0096 [0.0000,0.0297]	0.0222 [0.0000,0.0642]	0.8862 [0.7183,1.0000]	0.0717 [0.0000,0.2076]
23 & 15	0.0000 [0.0000,0.0048]	0.0002 [0.0000,0.0049]	0.0098 [0.0000,0.0292]	0.0086 [0.0000,0.0249]	0.0065 [0.0000,0.0188]	0.9276 [0.8158,1.0000]	0.0472 [0.0000,0.1408]
23 & 24	0.0000 [0.0000,0.3810]	0.0116 [0.0000,0.3878]	0.0676 [0.0000,0.4311]	0.0289 [0.0000,0.3967]	0.0271 [0.0000,0.3957]	0.3974 [0.0000,0.8168]	0.4674 [0.0402,0.8945]

Tests de l'étape 2 : mode de divergence de *C. pruni*

Scénarios / pop	1 – AC	2 – B vient de A	3 – A vient de B
28	0.8140 [0.6637,0.9643]	0.0605 [0.0000,0.6984]	0.1255 [0.0000,0.8222]
3	0.1861 [0.0062,0.3660]	0.8060 [0.6192,0.9928]	0.0079 [0.0000,0.7807]
2	0.1276 [0.0000,0.5843]	0.6059 [0.2943,0.9175]	0.2665 [0.0322,0.5008]
39	0.9224 [0.8362,1.0000]	0.0671 [0.0000,0.1482]	0.0105 [0.0000,0.0262]
1 & 25	0.6811 [0.4716,0.8907]	0.1035 [0.0056,0.2014]	0.2154 [0.0245,0.4062]
19 & 15	0.9307 [0.8208,1.0000]	0.0627 [0.0000,0.1688]	0.0067 [0.0000,0.0208]

Tests de l'étape 4 : précisions des scénarios démographiques sur les populations françaises

Scénarios / pop	Esp	1 – bottleneck anthropique (16.000 à maintenant) puis expansion	2 –bottleneck glaciation (16.000 à 2.58Ma) puis expansion	3 –bottleneck pré-glaciation (3Ma à 10Ma) puis expansion	4 – rétrécissement anthropique	5 – rétrécissement glaciation	6 – rétrécissement pré-glaciation
4	A	0.0000 [0.0000,0.0688]	0.2004 [0.1408,0.2599]	0.2144 [0.1537,0.2752]	0.1218 [0.0404,0.2032]	0.1975 [0.1326,0.2625]	0.2658 [0.1936,0.3381]
28	A	x	x	x	x	x	x
3	A	0.4212 [0.4120,0.4303]	0.0218 [0.0152,0.0285]	0.0243 [0.0176,0.0310]	0.4931 [0.4846,0.5017]	0.0205 [0.0138,0.0272]	0.0191 [0.0124,0.0258]
2	A	0.0002 [0.0000,0.0417]	0.2072 [0.1602,0.2542]	0.3329 [0.2785,0.3872]	0.1442 [0.0890,0.1994]	0.1676 [0.1234,0.2118]	0.1479 [0.1052,0.1906]
39	A	0.1010 [0.0963,0.1058]	0.0849 [0.0802,0.0895]	0.0937 [0.0890,0.0983]	0.5568 [0.5485,0.5650]	0.0781 [0.0738,0.0823]	0.0856 [0.0809,0.0903]
1	A	x	x	x	x	x	x
19	A	0.0001 [0.0000,0.0718]	0.3177 [0.2232,0.4122]	0.2920 [0.2080,0.3760]	0.1140 [0.0308,0.1972]	0.0824 [0.0141,0.1507]	0.1939 [0.1139,0.2738]
21	A	0.0000 [0.0000,0.0686]	0.2700 [0.1963,0.3436]	0.3443 [0.2661,0.4225]	0.1180 [0.0307,0.2053]	0.1250 [0.0616,0.1885]	0.1427 [0.0781,0.2074]
23	A	0.0002 [0.0000,0.0464]	0.1605 [0.1113,0.2097]	0.3666 [0.3096,0.4237]	0.0956 [0.0348,0.1563]	0.1456 [0.0989,0.1922]	0.2315 [0.1770,0.2861]
30	A	0.0003 [0.0000,0.0263]	0.2340 [0.1945,0.2735]	0.2047 [0.1706,0.2388]	0.1840 [0.1552,0.2129]	0.1554 [0.1232,0.1876]	0.2215 [0.1825,0.2606]
18	A	0.0060 [0.0000,0.0154]	0.1874 [0.1720,0.2029]	0.2319 [0.2156,0.2482]	0.2604 [0.2475,0.2733]	0.1510 [0.1377,0.1644]	0.1632 [0.1490,0.1774]
28	B	0.3830 [0.2972,0.4688]	0.0006 [0.0000,0.0539]	0.0109 [0.0000,0.0647]	0.6032 [0.5180,0.6883]	0.0021 [0.0000,0.0554]	0.0003 [0.0000,0.0535]
25	B	0.0009 [0.0000,0.3152]	0.1787 [0.0000,0.6244]	0.3593 [0.0000,0.8655]	0.3747 [0.0000,0.7668]	0.0220 [0.0000,0.3303]	0.0644 [0.0000,0.3835]
2	B	0.4806 [0.4451,0.5160]	0.0045 [0.0000,0.0373]	0.0050 [0.0000,0.0377]	0.4950 [0.4607,0.5294]	0.0079 [0.0000,0.0406]	0.0070 [0.0000,0.0397]
3	B	0.2329 [0.1725,0.2933]	0.0022 [0.0000,0.0208]	0.0156 [0.0000,0.0400]	0.7228 [0.6595,0.7862]	0.0172 [0.0000,0.0438]	0.0093 [0.0000,0.0307]
39	B	0.3081 [0.2581,0.3581]	0.0043 [0.0000,0.0266]	0.0088 [0.0000,0.0315]	0.6579 [0.6088,0.7069]	0.0094 [0.0000,0.0322]	0.0115 [0.0000,0.0346]
15	B	0.2058 [0.1614,0.2502]	0.0181 [0.0000,0.0363]	0.0299 [0.0088,0.0511]	0.6609 [0.6061,0.7158]	0.0455 [0.0143,0.0767]	0.0397 [0.0105,0.0690]
24	B	0.1125 [0.0712,0.1538]	0.0465 [0.0309,0.0620]	0.0750 [0.0546,0.0954]	0.4811 [0.4223,0.5399]	0.1527 [0.1115,0.1938]	0.1323 [0.0964,0.1681]
33	B	0.0017 [0.0000,0.0352]	0.1115 [0.0748,0.1482]	0.1296 [0.0919,0.1674]	0.3002 [0.2501,0.3503]	0.2067 [0.1527,0.2607]	0.2503 [0.1900,0.3107]
30	B	0.0021 [0.0000,0.2101]	0.2512 [0.0000,0.5155]	0.4110 [0.1693,0.6527]	0.0972 [0.0000,0.3631]	0.0908 [0.0000,0.2941]	0.1476 [0.0000,0.3675]
32	B	0.0287 [0.0185,0.0390]	0.0893 [0.0708,0.1078]	0.1656 [0.1380,0.1932]	0.4591 [0.4255,0.4927]	0.1347 [0.1099,0.1594]	0.1227 [0.0988,0.1465]
35	B	0.0211 [0.0000,0.1114]	0.0410 [0.0000,0.1320]	0.0796 [0.0000,0.1805]	0.4796 [0.3034,0.6559]	0.0874 [0.0000,0.1932]	0.2913 [0.0814,0.5012]
29	B	0.0073 [0.0000,0.0628]	0.1482 [0.0000,0.3507]	0.0866 [0.0000,0.1979]	0.5338 [0.3068,0.7609]	0.0814 [0.0000,0.1846]	0.1426 [0.0000,0.3267]
16	B	0.0021 [0.0000,0.2101]	0.2512 [0.0000,0.5155]	0.4110 [0.1693,0.6527]	0.0972 [0.0000,0.3631]	0.0908 [0.0000,0.2941]	0.1476 [0.0000,0.3675]
41	B	0.0072 [0.0000,0.0251]	0.1330 [0.1080,0.1579]	0.2065 [0.1783,0.2347]	0.2304 [0.2095,0.2513]	0.2234 [0.1928,0.2539]	0.1995 [0.1694,0.2296]

Tests de l'étape 5 : scénarios d'expansion de *C. pruni* en Europe

Scénarios / pop	1 : B → F → E	2 : E → F → B	3 : F → E puis B indépendant	4 → F → B puis E indépendant
75-84-87	0.0000 [0.0000,0.9110]	0.1251 [0.0000,0.9016]	0.8742 [0.7623,0.9860]	0.0007 [0.0000,1.0000]
78-85-94-88	0.0000 [0.0000,1.0000]	0.2062 [0.0000,1.0000]	0.7938 [0.3230,1.0000]	0.0000 [0.0000,1.0000]
79-64-67-89	0.9999 [0.9998,1.0000]	0.0000 [0.0000,1.0000]	0.0000 [0.0000,1.0000]	0.0000 [0.0000,1.0000]
70-46-90	0.0000 [0.0000,1.0000]	0.1968 [0.0000,1.0000]	0.8032 [0.2448,1.0000]	0.0000 [0.0000,1.0000]
76-71-87	0.0000 [0.0000,0.0001]	0.9887 [0.9791,0.9983]	0.0113 [0.0017,0.0209]	0.0000 [0.0000,0.0001]
96-84-88	0.0000 [0.0000,1.0000]	0.0382 [0.0000,1.0000]	0.9618 [0.7949,1.0000]	0.0000 [0.0000,1.0000]

95-85-94-89	0.0000 [0.0000,1.0000]	0.0346 [0.0000,1.0000]	0.9654 [0.8567,1.0000]	0.0000 [0.0000,1.0000]
91-64-67-90	0.0000 [0.0000,1.0000]	0.0028 [0.0000,1.0000]	0.9972 [0.9873,1.0000]	0.0000 [0.0000,1.0000]
75-46-88	0.0000 [0.0000,1.0000]	0.0935 [0.0000,1.0000]	0.9065 [0.6006,1.0000]	0.0000 [0.0000,1.0000]
78-71-89	0.0000 [0.0000,1.0000]	0.1060 [0.0000,1.0000]	0.8940 [0.6299,1.0000]	0.0000 [0.0000,1.0000]
79-84-90	0.0000 [0.0000,1.0000]	0.0851 [0.0000,1.0000]	0.9149 [0.6172,1.0000]	0.0000 [0.0000,1.0000]
70-85-94-87	0.0000 [0.0000,0.0000]	0.9985 [0.9974,0.9996]	0.0015 [0.0004,0.0026]	0.0000 [0.0000,0.0000]
76-64-67-88	0.0000 [0.0000,1.0000]	0.0836 [0.0000,1.0000]	0.9164 [0.5188,1.0000]	0.0000 [0.0000,1.0000]
96-46-89	0.0000 [0.0000,1.0000]	0.0640 [0.0000,1.0000]	0.9360 [0.7248,1.0000]	0.0000 [0.0000,1.0000]
95-71-90	0.0000 [0.0000,1.0000]	0.0343 [0.0000,1.0000]	0.9657 [0.8454,1.0000]	0.0000 [0.0000,1.0000]
91-84-87	0.0000 [0.0000,0.0003]	0.9798 [0.9639,0.9957]	0.0202 [0.0043,0.0361]	0.0000 [0.0000,0.0003]
75-85-94-89	0.0000 [0.0000,1.0000]	0.0635 [0.0000,1.0000]	0.9365 [0.6764,1.0000]	0.0000 [0.0000,1.0000]
78-64-67-90	0.0000 [0.0000,1.0000]	0.1303 [0.0000,1.0000]	0.8697 [0.5212,1.0000]	0.0000 [0.0000,1.0000]
79-46-87	0.0000 [0.0000,0.0000]	0.9984 [0.9973,0.9996]	0.0016 [0.0004,0.0027]	0.0000 [0.0000,0.0000]
70-71-88	0.0000 [0.0000,1.0000]	0.0955 [0.0000,1.0000]	0.9045 [0.5819,1.0000]	0.0000 [0.0000,1.0000]
76-84-89	0.0000 [0.0000,1.0000]	0.0285 [0.0000,1.0000]	0.9715 [0.8681,1.0000]	0.0000 [0.0000,1.0000]
96-64-67-90	0.0000 [0.0000,1.0000]	0.0220 [0.0000,1.0000]	0.9780 [0.8931,1.0000]	0.0000 [0.0000,1.0000]
95-71-87	0.0000 [0.0000,0.0000]	0.9983 [0.9935,1.0000]	0.0017 [0.0000,0.0065]	0.0000 [0.0000,0.0000]
91-46-88	0.0000 [0.0000,1.0000]	0.0613 [0.0000,1.0000]	0.9387 [0.6950,1.0000]	0.0000 [0.0000,1.0000]

Tests de l'étape 5bis : précisions des scénarios d'expansion de *C. pruni* en Europe

Scénarios / pop	1 : B → F → E	2 : E → F → B	3 : F → E puis B indépendant	4 → F → B puis E indépendant
75-84-87	0.0000 [0.0000,1.0000]	0.0000 [0.0000,1.0000]	1.0000 [0.9996,1.0000]	0.0000 [0.0000,1.0000]
78-85-94-88	0.0007 [0.0000,0.8331]	0.0402 [0.0000,0.8473]	0.9465 [0.9005,0.9925]	0.0126 [0.0000,0.8699]
79-64-67-89	0.0466 [0.0000,0.1309]	0.0107 [0.0000,0.0943]	0.4157 [0.2946,0.5368]	0.5270 [0.4062,0.6477]
70-46-90	0.0002 [0.0000,0.0028]	0.0003 [0.0000,0.0028]	0.9280 [0.8949,0.9611]	0.0714 [0.0384,0.1045]
76-71-87	0.0153 [0.0000,1.0000]	0.0000 [0.0000,1.0000]	0.0407 [0.0000,1.0000]	0.9440 [0.7804,1.0000]
96-84-88	0.0000 [0.0000,0.0009]	0.0028 [0.0008,0.0049]	0.9539 [0.9337,0.9740]	0.0432 [0.0231,0.0634]
95-85-94-89	0.2093 [0.0852,0.3334]	0.1722 [0.0687,0.2757]	0.4138 [0.2940,0.5336]	0.2047 [0.0224,0.3871]
91-64-67-90	0.2099 [0.0611,0.3588]	0.4105 [0.2455,0.5756]	0.3012 [0.1772,0.4252]	0.0784 [0.0000,0.2242]
75-46-88	0.0040 [0.0000,0.0151]	0.0059 [0.0000,0.0154]	0.1374 [0.0722,0.2025]	0.8527 [0.7862,0.9192]
78-71-89	0.0000 [0.0000,0.0007]	0.0158 [0.0056,0.0260]	0.9573 [0.9317,0.9828]	0.0269 [0.0045,0.0494]
79-84-90	0.0026 [0.0000,0.3333]	0.0045 [0.0000,0.3325]	0.6858 [0.5342,0.8374]	0.3071 [0.0000,0.6405]
70-85-94-87	0.0006 [0.0000,1.0000]	0.0000 [0.0000,1.0000]	0.6297 [0.3194,0.9400]	0.3697 [0.0000,0.8967]
76-64-67-88	0.1167 [0.0304,0.2031]	0.2236 [0.0939,0.3532]	0.5531 [0.4236,0.6825]	0.1066 [0.0367,0.1766]
96-46-89	0.0022 [0.0000,0.0091]	0.0001 [0.0000,0.0068]	0.1481 [0.1096,0.1866]	0.8497 [0.8112,0.8881]
95-71-90	0.0105 [0.0000,0.8528]	0.0024 [0.0000,0.8586]	0.2085 [0.0000,1.0000]	0.7785 [0.5351,1.0000]

91-84-87	0.0784 [0.0000,0.2242]	0.0000 [0.0000,1.0000]	0.0429 [0.0000,1.0000]	0.9481 [0.8050,1.0000]
75-85-94-89	0.0045 [0.0000,0.0100]	0.0035 [0.0000,0.0074]	0.0684 [0.0276,0.1091]	0.9236 [0.8816,0.9656]
78-64-67-90	0.0048 [0.0000,0.3002]	0.1122 [0.0000,0.3669]	0.7303 [0.6280,0.8326]	0.1527 [0.0000,0.4805]
79-46-87	0.0011 [0.0000,1.0000]	0.0000 [0.0000,1.0000]	0.0743 [0.0000,1.0000]	0.9246 [0.7334,1.0000]
70-71-88	0.0013 [0.0000,0.0302]	0.0022 [0.0000,0.0321]	0.8237 [0.6802,0.9672]	0.1729 [0.0297,0.3160]
76-84-89	0.0002 [0.0000,0.7361]	0.0159 [0.0000,0.7395]	0.9688 [0.9453,0.9923]	0.0152 [0.0000,0.7630]
96-64-67-90	0.0001 [0.0000,0.0148]	0.1847 [0.1197,0.2497]	0.8150 [0.7500,0.8800]	0.0002 [0.0000,0.0150]
95-71-87	0.3649 [0.0000,1.0000]	0.0000 [0.0000,1.0000]	0.0000 [0.0000,1.0000]	0.6351 [0.1482,1.0000]
91-46-88	0.0787 [0.0000,0.5958]	0.0029 [0.0000,0.5774]	0.0148 [0.0000,0.6461]	0.9036 [0.8473,0.9598]