Création d'une core collection des ressources génétiques forestières du pin de Salzmann en France

André de Souza Rodriguez

Mémoire de deuxième année de Master Biodiversité Ecologie Evolution (BEE)

2021 - 2022

Création d'une *core collection* des ressources génétiques forestières du pin de Salzmann en France

Un peuplement de Pin de Salzmann dans le Conflent, au pied des sommets des Pyrénées Orientales © Daniel Cambon, 2007

André DE SOUSA RODRIGUES

Structure d'accueil : URFM (INRAE) **Date de soutenance :** 22 juin 2022
Dates du stage : 15 mars – 31 août 2022
Co-encadrant.e.s : Bruno FADY et Caroline SCOTTI-SAINTAGNE

Remerciements :

Je tiens tout d'abord à remercier mes co-encadrants de stage Caroline Scotti-Saintagne et Bruno Fady pour leur accueil, leur bienveillance et leur connaissance sans faille sur le bel arbre qu'est le pin de Salzmann. Merci Caroline pour les longues journées d'analyses et d'explication du fonctionnement des logiciels, je maîtrise tout cela parfaitement (ou presque) à présent. Merci Bruno pour les check-up réguliers et l'expertise sur les analyses génétiques.

Je remercie également tous les collègues de l'URFM, découvrir un environnement de travail en unité de recherche en présentiel après les restrictions sanitaires fait chaud au cœur, et les échanges sur le travail de chacun sont toujours passionnants. Remerciements particuliers à mes camarades doctorant.e.s et stagiaires qui sont tous.tes très gentil.le.s et très cools, on s'amuse bien autour du patio bientôt envahi d'ipomées.
Merci également à la salle de pause pour les innombrables cafés et thés fournis, c'est un grand plaisir. Merci aussi à la cantine et surtout aux mardis-frites sans lesquelles la vie est moins belle.

Je voudrai aussi remercier mes ami.e.s en région parisienne avec qui on a pu maintenir le contact durant ma première longue absence de la région, big up aux membres de la Secte et du Carrousel qui m'ont accompagné quotidiennement tout au long de mon stage pour de longues heures de conversations passionnantes.

Enfin, je souhaiterai remercier ma playlist Spotify, pour avoir toujours été là durant mon travail. Merci.

Table des matières

Introduction

La conservation de la biodiversité est un enjeu majeur pour la société humaine. Les tendances observées de perte de biodiversité ainsi que l'accroissement des menaces qui pèsent sur celle-ci font que l'on considère vivre actuellement une 6ème crise de la biodiversité (Ceballos et al. 2017). C'est dans ce contexte que la conservation d'espèces vulnérables ou en danger est nécessaire et urgente. Les programmes de conservation déployés peuvent généralement être classés en deux catégories : la conservation in situ et ex situ. La conservation in situ consiste à préserver les espèces dans leurs milieux de vie, en assurant une vigilance sur l'ensemble de l'habitat que l'on souhaite protéger, avec des suivis réguliers et des interventions plus ou moins importantes selon les plans de gestion. On peut ainsi maintenir ou restaurer des populations dans le milieu même où elles ont acquis leurs caractéristiques. Cependant, pour conserver des espèces in situ il est impératif d'avoir des tailles de populations suffisantes, afin d'assurer une diversité génétique nécessaire à la survie de la population ainsi que des moyens de contrôler les espaces de protection. Dans le cas de petites populations, telles que les espèces en risques d'extinction, la translocation est devenue de plus en plus courante dans la gestion des espèces de manière globale, notamment pour les plantes. On met alors en place des mesures ex situ, en plaçant une partie de la population dans un environnement plus ou moins contrôlé. La conservation ex situ est en effet actuellement un des moyens les plus efficaces pour préserver la diversité végétale (Mounce et al. 2017). La conservation des écosystèmes forestiers est quant à elle assez particulière et se fait principalement in-situ dans des aires protégées et notamment pour la conservation de la diversité génétique au sein des espèces. En France, la *Commission Nationale pour la Conservation des Ressources Génétiques Forestières* (CRGF créée en 1991) compte en 2021 près de 103 unités conservatoires in situ, couvrant 10 espèces protégées d'arbres forestiers (Desgroux et al 2020). Mais pour des populations particulièrement en danger dû à des risques environnementaux forts ou des espèces rares, des collections ex situ peuvent être mises en place. Cela concerne pour la CRGF 6 espèces, dont le pin de Salzmann.

Le pin de Salzmann (*Pinus nigra subsp. salzmannii*) est une des cinq sous-espèces de pin noir (*Pinus nigra*) (Scotti-Saintagne et al. 2019), endémique à la région méditerranéenne et au sud de l'Europe. Il a été découvert en 1810 par Philipp Salzmann, botaniste allemand, dans l'Hérault et a été décrit par Michel Félix Dunal en 1851. Si le pin de Salzmann constitue de grands peuplements forestiers producteurs de bois en Espagne couvrant environ 350 000

hectares, sa situation en France est différente. L'aire de répartition du pin de Salzmann était vaste et étendue à l'origine, mais elle a été morcelée du fait de nombreuses activités humaines (récolte de bois, écobuage, culture, pâturage) au cours du XIXème siècle (Vernet 2010). Suite à cette forte déforestation, d'importants travaux de reboisement ont été mis en place, dans le cadre de la loi de Restaurations des Terrains de Montagne (RTM) de 1860. Lors de ces reboisements, plusieurs espèces furent introduites, notamment des sous-espèces exotiques de pin noir, telles que le pin noir d'Autriche *(Pinus nigra subsp. nigra)* et le pin noir laricio de Corse *(Pinus nigra subsp. laricio)* qui furent plantées proches ou au sein même des zones restantes de pin de Salzmann. Ainsi, le pin de Salzmann est aujourd'hui présent en France sous forme de lambeaux résiduels, on recence en France sept stations isolées, en habitats de moyenne montagne, qui recouvrent approximativement 3000 hectares, ce qui fait du pin de Salzmann une des espèces forestières les plus rares en France (Quezel et Barbero 1988). Ce pin de Salzmann est aujourd'hui soumis à trois menaces principales qui tendent à accentuer sa raréfaction :

- Divers accidents liés au climat et aux actions de l'Homme : principalement les incendies auxquelles les petites populations de pin de Salzmann peuvent être très sensibles, mais également des tempêtes.
- Risque d'introgression par les autres pins noirs utilisés en reboisements. En effet, les pins de Salzmann peuvent se reproduire avec les autres espèces de pin noir (Arbez 1980), comme le pin laricio et le pin noir d'Autriche qui ont été très employés dans les reboisements en France. Ces flux de gènes peuvent conduire à des hybridations et engendrer une pollution génétique de l'espèce qui avoir des effets sur la durabilité et la valeur adaptative des individus.
- Isolement des populations par fragmentation de l'aire d'habitat et faiblesse des effectifs de certains groupes d'individus isolés (noyaux). Cette petite taille de population peut conduire, via une augmentation de la consanguinité, à un « vortex d'extinction » (Gilpin, E. et M. Soulé 1986)

Les forêts de pin de Salzmann sont incluses dans les sites Natura 2000 (annexe I Directive Habitats de 1992, code 9530* : « Pinèdes (sub-)méditerranéennes de pins noirs endémiques »), et sont considérés comme un habitat d'intérêt communautaire prioritaire. Ces espèces endémiques présentent ainsi un intérêt de conservation particulier. De plus, de par ses faibles exigences écologiques et de sa rusticité, le pin de Salzmann peut être particulièrement

intéressant en foresterie dans le cadre de l'adaptation des peuplements forestiers soumis à l'impact du changement climatique.

Du fait des menaces importantes et de l'intérêt que porte cette espèce, des modes de gestion **in situ** (recommandés dans les cahiers d'habitats Natura 2000) ont été mis en place afin de préserver ces peuplements : ouverture des peuplements pour favoriser la régénération naturelle, gestion sylvicole classique et élimination des pins noirs introduits. Les menaces pesant sur le pin de Salzmann ont également mené la CRGF à le considérer dans les espèces prioritaires fin années 2000 et à conseiller une conservation des ressources génétiques de ce pin endémique. Afin de conserver au mieux cette espèce, des mesures de conservation **ex-situ** ont été mises en place, avec un objectif de sauvegarder les individus remarquables existant dans les peuplements naturels en les copiant par greffage et en les installant dans des plantations conservatoires. C'est ainsi qu'une collection nationale de conservation a pu être mise en place. Celle-ci compte 694 clones issus de greffes d'arbres originaires de peuplements naturels. Cette collection présente néanmoins des contraintes, comme la place occupée par les arbres, le maintien coûteux de la collection entière ainsi que le risque d'incendie important de la zone (Cadarache, Bouches-Du-Rhône). C'est pour faire face à ces contraintes que la CRGF propose la réalisation d'une core collection (collection de base), qui permettra de diminuer le nombre d'arbres sauvegardés en maximisant la représentativité de la diversité génétique de la collection globale.

De par les efforts globaux pour conserver des ressources génétiques de plantes rares ou utilisées en agriculture, le nombre de collections ex-situ a grandement augmenté ces dernières années (Odong et al. 2013). Le besoin de mettre en place des core collections suscite ainsi un réel intérêt pour de nombreuses espèces différentes. La démarche d'établir des core collections a été pensée dans le but de faciliter la caractérisation et l'utilisation des collections tout en préservant le plus possible la diversité génétique de la collection entière. Frankel (1984) a défini le concept de core collection comme « un ensemble limité d'accessions représentant, avec le minimum de redondance, la diversité d'une espèce ». Depuis lors, le concept de core collection n'a cessé de se développer et plusieurs publications, aussi bien sur l'aspect théorique que pratique de la mise en place de core collection, ont vu le jour. Il existe actuellement de nombreuses approches et méthodes différentes proposées et utilisées (M-Strat par Gouesnard et al 2001, PowerCore Kim et al 2007, CoreHunter Thachuk et al 2009, GenoCore Jeong 2017). En se basant sur une étude comparative de Jeong et al (2017), nous

nous sommes concentrés sur la méthode Core Hunter qui permet de choisir parmi plusieurs critères de d'évaluation de qualité de la core collection. D'après Odong et al. (2013), la plupart des core collections dans la littérature sont construites en se basant sur des critères similaires alors que les objectifs peuvent être différents. Or, comme l'explique Odong et al. (2013), Il n'existe pas, une seule bonne core collection, mais plusieurs puisque selon les objectifs désirés le critère de sélection des individus peut varier. Il est donc important de définir clairement l'objectif d'une core collection, ce que nous verrons en détail dans la partie Méthode sur notre cas d'étude.

Dans cette étude effectuée dans le cadre de mon stage de M2 à l'URFM d'Avignon, le but est de mettre au point une méthode analytique efficace et généralisable pour réaliser une core-collection. Cette méthode est appliquée à la collection nationale de pin de Salzmann. L'évaluation finale de la core collection est réalisée par comparaison des paramètres de diversité génétique et de différenciation génétique avec la collection nationale globale. L'objectif final est de pouvoir appliquer cette méthode à d'autres collections et quel que soit leur objectif de conservation

Matériels et Méthodes

1) Présentation des données obtenues au préalable de l'étude

1.1) Constitution de la collection

Afin de constituer la collection de pin de Salzmann des greffons d'individus issus de peuplements naturels ont été récoltés de 2008 à 2012. Les arbres dont les greffons ont été prélevés proviennent de sept groupes de peuplements de pins de Salzmann : Saint Guilhem, Conflent, Ardèche, Col d'Uglas, Gachas (Gard), La Tour sur Orb et Gorges du Tarn (cf. figure 6 pour leur emplacement géographique). Ces peuplements représentent l'ensemble des peuplements de pin de Salzmann qui ont de grandes chances d'être autochtone. L'autochtonie est en effet un élément primordial dans le choix des essences à prélever afin de s'assurer d'obtenir des arbres représentatifs de la diversité naturelle des pins des Salzmann sans considérer les potentiels hybrides.

Afin de choisir les individus représentatifs pour les analyses génétiques et la constitution de la collection, une identification des arbres issus de peuplements autochtones a été réalisée. Le repérage concernait les individus âgés de plus de 140 ans en 2000, et cela car on peut ainsi

considérer qu'ils sont nés avant les premières campagnes de reboisement de la RTM (1870), et qu'ils ont une faible chance d'avoir été planté ou d'être hybride avec un pin noir non autochtone. L'âge des arbres susceptibles d'avoir l'âge requis a été estimé à l'aide d'un comptage des cernes sur des carottes de bois et ce sont 810 arbres qui ont été retenus pour les actions de greffage et génotypage sur plus de 4500 arbres carottés.

Figure 1 : Carotte prélevée sur un pin de Salzmann candidat. Les cernes annuels bien visibles permettent une estimation précise de l'âge de l'individu (celui-ci est un vieil arbre de plus de 140 ans, retenus dans l'échantillon). Photo © Norbert Turion.

Afin d'obtenir des greffons pour cloner les individus sélectionnés et les planter dans la collection ex-situ, des récoltes de rameaux ont été effectuées pour les 810 arbres retenus de 2008 à 2012, en privilégiant les rameaux les plus jeunes possibles (les plus apicaux).

Figure 2 : Récolte d'échantillons végétaux de pin de Salzmann pour les analyses ADN et le greffage, par grimpage dans le site des Gorges du Tarn. Photo © Daniel Cambon.

La collection de travail est installée à St Paul lez Durance (Bouches du Rhônes). Elle se compose en mars 2022 de 667 génotypes au total, pour un total de 1257 arbres, avec un nombre de copies variant entre 1 et 6 par génotype.

1.2) Données génétiques

Afin de réaliser des analyses génétiques sur les peuplements de pin de Salzmann, des marqueurs génétiques ont été mis au point Après une étude de différentes méthodes de marquage réalisée par G. Giovanelli en 2017, l'approche utilisant des marqueurs microsatellites s'est révélée la plus pertinente pour l'étude des sous espèces de *Pinus nigra*. Dans cette étude, des marqueurs microsatellites nucléaires ont été retenus pour l'analyse génétique d'espèces de pin noir, dont le pin de Salzmann : 9 marqueurs ont été mis au point *de novo* spécifiquement pour les pins noirs, 2 marqueurs ont été transférés de *Pinus tadea* et 2 autres de *Pinus halepensis*. C'est donc un total de 13 marqueurs microsatellites nucléaires qui

ont été utilisés pour caractériser les génotypes diploïdes des individus de la collection de conservation des pins de Salzmann. Les données à notre disposition regroupent 684 génotypes en tout. Parmi ces 684 génotypes, 64 ne sont pas présents dans la collection de Cadarache actuellement, dû à des morts d'individus dans la collection ou à des échecs de greffage.

2) Analyses génétiques : étude de la diversité génétique intra-population et de la différenciation génétique inter-populations

2.1) Allèles nuls

La détection d'allèles nuls est une étape fondamentale en analyses de génétique des populations utilisant des marqueurs microsatellites. En effet, l'une des principales limites des analyses utilisant des marqueurs microsatellites est leur présence dans les données (Dabrowski et al. 2014). Les allèles nuls sont des allèles qui ne sont pas amplifiés lors des analyses PCR, le plus souvent à cause de changements (mutation insertion, délétion) dans les régions flanquantes des amorces d'ADN, empêchant ainsi la liaison à l'amorce (Callen et al. 1993). Ce sont donc des formes alléliques réelles qui ne sont pas dues à des erreurs de manipulation en laboratoire, alors que le résultat observé peut être d'aucun allèle détecté dans le cas où l'allèle nul est présent en forme homozygote pour un locus diploïde. Dans le cas d'un locus hétérozygote, un allèle nul va nous amener à penser que le locus est homozygote (un seul allèle apparent), alors qu'en réalité un allèle nul non détecté est présent, ce qui peut entraîner un excès d'homozygotes dans nos données et fausser les analyses (Oddou-Muratorio et al. 2009).

Pour détecter une potentielle présence significative d'allèles nuls dans nos données, nous avons utilisé d'une part MICRO-CHECKER (Oosterhout et al. 2004) afin d'estimer la présence et la fréquence d'allèles nuls au sein de nos 13 marqueurs microsatellites d'étude (5000 randomisations), ainsi que ML-NullFreq (Kalinowski et al. 2006) qui utilise un autre algorithme pour le calcul de fréquence d'allèles nuls. L'utilisation conjointe de ces deux méthodes, recommandée par Dabrowski et al. (2014), permet de minimiser les faux négatifs en termes de détection d'allèles nuls. C'est également la méthode utilisée par Giovanelli (2017) lors de son étude sur les mêmes marqueurs génétiques sur les peuplements de pin noirs, qui n'avait pas mis en évidence de présence significative d'allèles nuls. En plus de ces méthodes, nous avons également utilisé FreeNA (Chapuis et al. 2007) pour avoir une estimation supplémentaire de la présence d'allèles nuls, le but étant de combiner les résultats de ces 3 méthodes afin d'avoir une estimation la plus précise de la présence d'allèles nuls. Cette étude a été faite pour chaque groupe de structure/provenance de pin de Salzmann afin de

se rapprocher au maximum du cadre théorique d'une population à l'équilibre de Hardy Weinberg, les allèles nuls étant estimés en faisant l'hypothèse de l'équilibre de H-W. La structuration génétique entre les populations (effet Wallund) peut biaiser cette estimation. Après l'estimation des locus présentant des allèles nuls, nous avons réalisé un test de robustesse de la méthode à 13 locus en réalisant analyse de structure et une estimation des indices de diversité sans les locus avec une forte fréquence d'allèles nuls, afin de s'assurer du poids des allèles nuls dans nos analyses.

2.2) Inférence de la structure génétique sans à priori sur la localisation des arbres

Avant de constituer la core collection, l'étape de caractérisation de la structure génétique des peuplements est indispensable. Pour ce faire, nous avons utilisé STRUCTURE 2.3 (Pritchard et al. 2000), un logiciel très répandu en analyse de structure génétique. L'objectif de cette analyse est de mettre en évidence les groupes/clusters génétiques des individus génotypés afin de mieux construire la core collection et de s'assurer que des individus de chaque groupe soient présents dans la sélection. Ce logiciel utilise une approche de clustering Bayésien afin de regrouper des individus selon leur génotype renseigné en un nombre de groupes K. Nous avons considéré un modèle admix, pour lequel les individus peuvent avoir une ascendance mixte, et ainsi appartenir à certains groupes sous plusieurs taux différents. Les 13 marqueurs microsatellites ont été utilisés pour cette analyse. Nous avons effectué trois itérations indépendantes avec pour chacune des valeurs de K allant de 1 à 10, le nombre de provenances étant de 7 avec des regroupements possibles, il ne semble pas judicieux de considérer plus de groupes que 10. Ces analyses ont été réalisés avec un burn-in de 50 000 pas suivi de 500 000 répétitions en chaîne de Markov par méthode de Monte Carlo.

Nous avons ensuite utilisé l'application StructureHarvester (Earl et von Holdt et al. 2012) afin de calculer les valeurs et de réaliser les visualisations graphiques permettant d'estimer le nombre de K le plus probable pour notre analyse. Ces calculs permettent d'obtenir les moyennes vraisemblance L(K) ainsi que le « taux de variation » (ΔK) entre des valeurs de K successives décrit par Evanno et al (2005) (voir figure 4b). Des indicateurs supplémentaires proposés par Puechmaille (2016), « MedMeaK » et « MaxMeaK », ont été pris en compte. Ils correspondent respectivement un à un calcul de médiane et maximum de moyenne du nombre de groupes, en étudiant pour chaque provenance connu (sept provenances ici) si la moyenne de taux d'apparentement des individus à un groupe de structure donné est supérieure à un seuil de 50 %. Ces indicateurs du nombre de groupes ont l'avantage de ne pas être soumis à la différence d'effectif entre les provenances.

Les résultats des trois itérations ont ensuite été compilés et visualisés à l'aide de l'application web CLUMPAK (Kopelman et al. 2015) qui utilise le logiciel CLUMPP (Jakobsson et al. 2007) pour la compilation et le logiciel DISTRUCT (Rosenberg 2003) pour la visualisation graphique. Différentes analyses ont été effectués en suivant cette méthode afin de s'assurer de l'effet de taille de populations non homogènes dans les données, problème qui a déjà été relevé pour des analyses STRUCTURE dans d'autre études (Puechmaille 2016) :

- Une analyse structure avec tous les individus génotypés, 684 individus issus de 7 populations supposées. *(Ana1)*
- Deux analyses en se basant sur les résultats en K=2 de la première analyse. Les individus ont été séparés en 2 groupes selon leur score d'appartenance, afin de bien identifier les sous structures de chaque cluster. *(AnaQ1 et AnaQ2)*
- Une analyse avec effectifs homogènes pour les 7 populations. Sélection de 20 individus pour les 5 populations qui en avaient plus, choix systématique en évitant de prendre des individus proches géographiquement, proches dans l'ordre de la base de données et en prenant ceux avec le moins de données manquantes. (*Ana2)*
- Une analyse avec effectifs homogènes pour les 7 populations. Sélection de 41 individus pour les populations qui en avaient plus, on se base ici sur l'effectif de la population Gorges du Tarn car celles avec moins d'effectifs semblent se rattacher à d'autres populations d'après des précédentes analyses. *(Ana3)*
- Analyses par groupes de structuration (5 groupes considérés donc 5 analyses en tout) afin de s'assurer de l'uniformité ou non de ces groupes.

2.3) Analyse en composantes principales (ACP)

Afin d'estimer la diversité génétique entre les populations et au sein des populations, nous avons réalisé une analyse en composantes principales (ACP) sur R. Cela permet d'avoir une information visuelle sur la répartition des individus selon leur génotype (d'après les 13 marqueurs génétiques). Afin de réaliser cela sur R, nous avons utilisé package *adegenet* (Jombart et al. 2008) qui permet de réaliser ce genre d'étude sur des marqueurs génotypiques en formatant les données en objet *genind* contenant les génotypes individuels. Afin de pouvoir réaliser les ACP malgré des données manquantes, nous avons utilisé la fonction *scalegen* qui permet de mettre à l'échelle les présences/absences de génotypes par individu, en remplaçant les données manquantes par des moyennes.

2.4) Indices de diversité génétique et de différenciation génétique

Afin d'estimer la diversité génétique des individus de la collection, des indices de diversité ont été calculés en utilisant le logiciel Spagedi (Hardy et al. 2002) : Fis, moyenne et variance des tailles d'allèles, hétérozygotie observée et attendue et nombre d'allèles avec valeur raréfiées pour l'effectif le plus petit (nombre d'allèles estimés à toutes les populations si elles présenteraient le même effectif). Ces indices sont calculés pour chaque marqueur et chaque groupe de structure (résultats partie 2.2), avec des valeurs moyennes en considérant tous les marqueurs/tous les groupes.

Nous avons également calculé des indices de différenciation génétique entre les populations à l'aide de GenAlEx (Peakall and Smouse 2012), en utilisant les indices G'st de Hedrick (2005) et l'indice de D de Jost (2008), qui sont des alternatives aux indices Fst et Gst connus qui permettent de s'affranchir à la dépendance à la diversité intra population de ces derniers. Leur significativité est testée avec un test de 999 permutations.

3) Constitution de la core collection

3.1) Objectif de la collection

Avant de construire une core collection il est nécessaire d'établir quel est l'objectif recherché. Nous nous sommes basés sur l'étude de Odong et al. (2013) qui identifie trois grands objectifs que peut viser une core collection (CC). Pour la définition de ces trois types de CC, le terme « accession » fait référence aux éléments constituant la collection complète et « entrée » les éléments de la core collection. Toutes les entrées sont des accessions mais seul quelques accessions sont des entrées :

- CC-I « Individuel » : Représentation uniforme des individus, on ne considère pas le poids des allèles. Chaque accession de la collection est représentée par une entrée dans la CC qui est la plus similaire. Cela correspond à une représentation uniforme de l'espace génétique de la collection, avec un poids constant dans cet espace, c'est la vision la plus intuitive d'une CC. La CC de type CC-1 constitue la meilleure option pour obtenir une CC « passe-partout » ou généraliste par rapport aux autres.

- CC-X « Extrêmes » : Représentation des individus en conservant les extrêmes alléliques. Une bonne CC de ce type comprend des entrées les plus différentes possibles les unes des autres. L'objectif de ce type de CC est de maximiser la diversité des traits des entrées de la CC.

• CC-D « Distribution » : Représentation de la distribution d'origine. Les proportions des accessions représentées dans une CC de ce type reflètent les contributions numériques des différentes régions ou groupe pour la collection entière. L'objectif ici est que les distributions des traits d'intérêt parmi les entrées de la CC soient les plus similaires (en termes de moyenne, variance et fréquence) à celles de la collection entière. Ce type de CC n'a d'intérêt que si l'on souhaite donner un aperçu de la composition de la collection entière en utilisant qu'une partie de celle-ci.

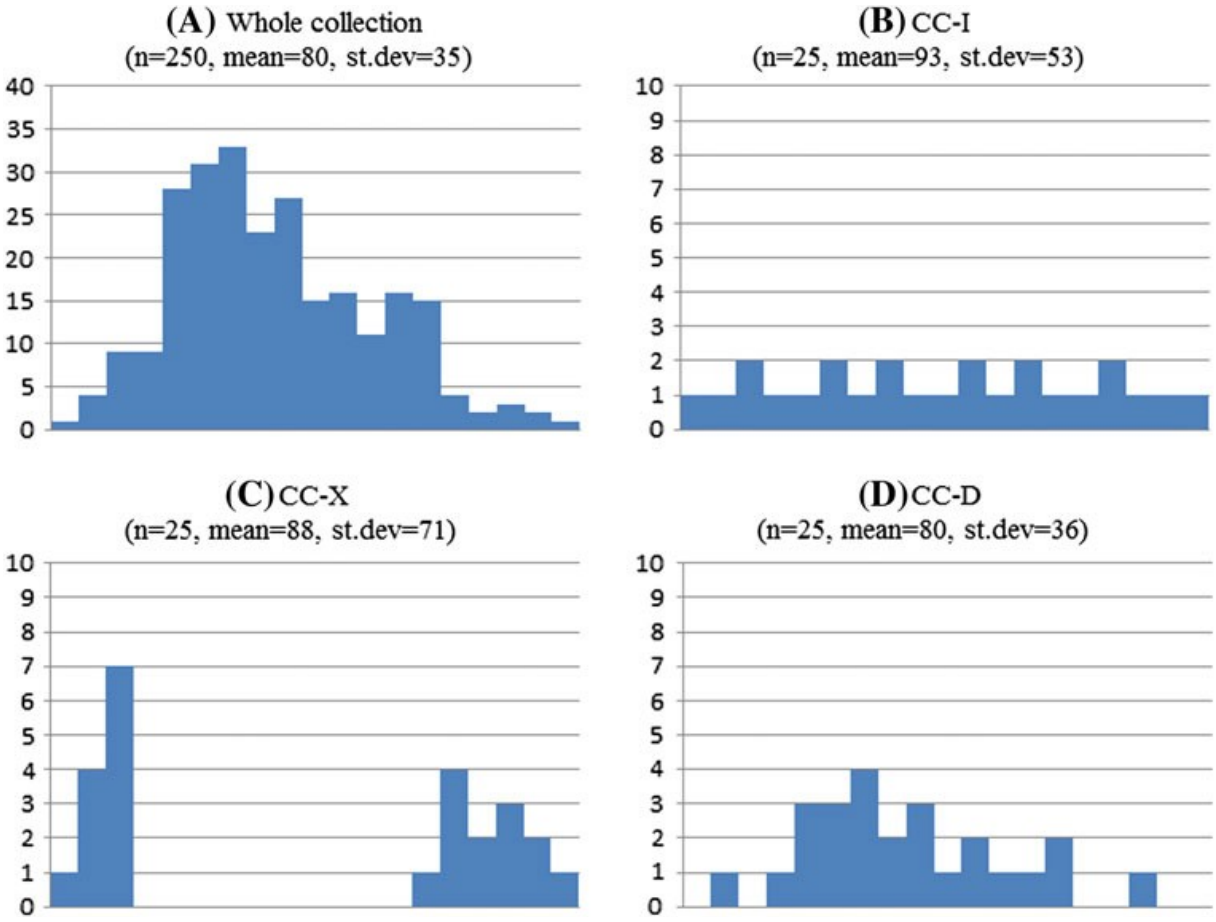


Figure 3 (issue de Odong et al. 2013) : a) Distribution multimodale pour un caractère d'une collection complète ; b) distribution du même trait pour une core collection de type CC-1 ; c) distribution du même trait pour une core collection de type CC-X ; c) distribution du même trait pour une core collection de type CC-D

Pour notre étude une core collection de type CC-I semble le plus adapté. En effet, l'objectif est d'avoir une représentation uniforme de la diversité de nos arbres avec le moins d'individus possibles qui se répètent. Ne garder que les extrêmes alléliques (CC-X) ne semble pas

pertinent dans notre cas, il n'y a pas vraiment de critères extrêmes rares à conserver, on cherche à obtenir une core collection avec le moins d'individus possibles qui sauvegarde la diversité de la population. L'objectif CC-D peut être intéressant pour représenter au mieux la population d'origine dans notre core collection mais elle peut entraîner des redondances fortes, notamment pour les provenances avec de nombreux individus ce qui pourrait affecter la diversité génétique de la CC.

3.2) Critères de la core collection

Pour la construction de la core collection, nous avons utilisé le package Core Hunter 3 (Thachuk et al 2009 ; De Beukelaer 2018) utilisable sur R, permettant d'utiliser des critères de mesure appropriés à notre objectif de core collection. Avec cette méthode, le nombre d'entrées de la CC désiré doit obligatoirement être renseigné à priori, ainsi que les critères considérés. Core Hunter présente 5 critères différents selon les objectifs de la core collection, avec deux de critères de distance génétique et trois critères de diversité génétique :

a. EN : « Average Entry-to-Nearest-Entry distance » (Odong et al. 2013). Cela représente la distance moyenne entre toutes les entrées et leur plus proche voisin. Maximiser ce critère permet une grande diversité dans la core collection qui se traduit par une dissemblance maximum entre les individus de la core collection (De Beukelaer 2018). Cela s'apparente plutôt à une core collection de type CC-X.
b. AN : « Average Accesion-to-Nearest-Entry distance » (Odong et al. 2013). Cela représente la distance moyenne entre chaque accession de la collection entière et l'entrée de la core collection la plus similaire, incluant l'accession elle-même si elle a été sélectionnée. Minimiser ce critère permet une core collection qui représente au maximum tous les accessions individuelles de la collection entière (De Beukelaer 2018). Cela s'apparente plutôt à une core collection de type CC-I.

Pour ces deux critères, deux mesures de distance sont possibles, distance modifiée de Roger MR (Wright 1978) et distance de Cavalli-Sforza CE (Cavalli-Sforza and Edwards, 1967).

c. SH : Indice de diversité de Shannon (Shannon 1948). L'indice de diversité de Shannon est une mesure appropriée à la sélection de core collection lorsque l'on souhaite conserver le plus d'allèles rares possible (Thachuk et al 2009). Il atteint sa valeur maximum quand chaque allèle n'existe qu'une seule fois dans la core collection mesurée.

d. HE : Hétérozygotie attendue (Berg and Hamrick, 1997). Cela représente la proportion attendue de locus hétérozygotes. Contrairement à l'indice de Shannon, cette mesure prend en compte la diversité au sein de chaque locus. Comme chaque locus contribue également à la valeur globale de cette mesure, les CC sélectionnées avec cette mesure auront moins de chance d'être homozygotes pour un nombre donné de loci comparé aux sélections avec SH. Cette mesure est ainsi adaptée pour sélectionner des CC en favorisant la diversité allélique au sein et entre les loci. (Thachuk et al 2009).
e. CV : Couverture allélique. Cela correspond au pourcentage d'allèles observés dans la collection complète que l'on retrouve dans la core collection. C'est une mesure simple qui est particulièrement utile pour sélectionner des CC préservant les allèles dans des banques génétiques par exemple. (Thachuk et al 2009).

3.3) Analyse de stabilité et choix des critères

En entrant un critère et une taille de CC visée, différentes combinaisons d'accessions peuvent aboutir à des valeurs identiques des critères d'évaluation de la CC sur Core Hunter. Dans une phase exploratoire, nous avons réalisé une étude de stabilité des différents critères d'évaluation de la CC en fixant une taille de core collection de 50% des individus génotypés (342). Dix itérations ont été réalisées pour chaque critère afin d'évaluer la stabilité de la liste des individus qui constituent la core collection. Pour la métrique de distance, nous avons utilisé par défaut la distance modifiée de Roger (MR). D'après Soleimani (2020), le choix de cette métrique n'influe que très peu sur la sélection des entrées de la CC.

Dans une deuxième phase, nous nous sommes concentrés sur un seul d'évaluation, AN, qui correspondait le mieux à notre objectif de core collection CC-I (Odong et al. 2013). Le critère EN étant plus adapté pour des core collections de type CC-X, il n'a pas été envisagé dans notre étude. Les critères HE et SH, bien qu'étant les plus stables, ne sont pas considéré de bonnes qualités pour des core collections où l'objectif visé est CC-I ou CC-X, pour lesquels les critères AN et EN sont respectivement les plus adaptés (Soleimani 2020). Deux types d'analyses ont été réalisées avec le package Core Hunter sur R, chacune est répétée 10 fois pour évaluer la stabilité du choix des arbres pour différentes tailles de core collection :

- Analyse globale avec critère AN : On a considéré ici l'ensemble des individus, peu importe leur provenance, en ayant un objectif de CC-I en utilisant la métrique AN. On a donc fait cela pour différentes tailles de core collections (5%, 10%, 20%, ..., 80% de la collection complète), en relevant le nombre d'arbres stables et la valeur de AN pour

chaque cas avec la liste des arbres sélectionnés. Nous avons également relevé le pourcentage de couverture allélique (CV) de la core collection pour chaque sélection, afin de pouvoir se rendre compte à partir de quelle taille de CC tous les allèles peuvent être conservés.

- Analyse globale avec critère AN et CV : Le package Core Hunter a la possibilité de se baser sur plusieurs critères combinés afin de réaliser la sélection de core collection. Pour être sûr de pouvoir garder tous les allèles présents de la collection entière dans la core collection, nous avons réalisé la même analyse que précédemment mais en utilisant le critère de CV en plus du critère AN, pour ainsi obtenir des core collections qui maximisent la couverture allélique en suivant un objectif CC-I avec le critère AN. Le but étant de comparer les valeurs de AN et de CV avec la précédente analyse, afin de voir si le compromis d'une distance AN moins optimal est important pour un gain en couverture allélique (idéalement à 100%).

Afin de prendre en compte l'hétérogénéité des résultats obtenus sur 10 itérations (différentes combinaisons d'arbres aboutissent aux mêmes valeurs de critère d'évaluation) nous avons établi une méthode qui permet sélectionne les individus par ordre décroissant de stabilité : tous les arbres présents 10 fois/10, 9fois/10…jusqu'à obtenir la taille de CC souhaitée. S'il y a plusieurs choix d'arbres dans la dernière classe de stabilité avant d'arriver à la taille souhaitée, alors nous effectuons un tirage aléatoire.

3.4) Construction de la core collection cas d'étude

La méthode mise au point a ensuite été appliquée à la constitution de notre « cas d'étude », une core collection du pin de Salzmann. Les critères ont été AN et CV et nous avons choisi une taille de 30% de la collection. Ce choix s'est basé sur les résultats des valeurs CV notamment, c'est la taille minimale pour laquelle on atteint une couverture allélique de 100% (cf. partie résultat 2.2). L'évaluation de la core collection a finalement été réalisée en comparant les indices de diversité et de structure génétique avec celles obtenus dans la collection nationale globale.

Pour construire cette collection cas d'étude nous avons également pris en compte d'autres informations : la présence du génotype dans la collection et l'adéquation entre le groupe génétique (résultat de l'analyse structure) et la provenance réelle.

- *Présence dans la collection :* Pour la mise au point de la méthode, nous avons considéré l'ensemble des 684 génotypes à disposition. Comme ces génotypes ne sont pas tous présents sous forme de greffons, nous avons inclus une option (dans la routine sur R) qui permet de garder/ou éliminer les individus absents de la collection. Dans notre cas d'étude nous avons éliminé les arbres non greffés et donc absents de la collection (684-64 = 620 génotypes).
- *Inadéquation entre groupe de structure et provenance :* Certains individus appartiennent à un groupe génétique (coefficient d'admixture, Q, supérieur à 80%, résultats de Structure), qui ne correspond pas au groupe génétique majoritairement observé dans leur provenance géographique d'origine (voir Résultats structure). Pour prendre cela en compte, nous avons incorporé dans notre routine R une option afin de vérifier l'adéquation entre les groupes génétiques inférées par Structure et les provenances réelles, permettant d'identifier ces arbres pour les garder ou non dans la sélection de la core collection. Nous n'avons pas considéré ces individus pour la core collection cas d'étude (620-8 = 612 génotypes).

Une fois la core collection d'étude obtenue en suivant la routine R mise en place, nous avons recalculé les indices de diversité, la différenciation ainsi que la structure génétique de notre CC obtenue afin de la comparer à la collection complète.

Résultats

1) Analyses génétiques

1.1) Allèles nuls

Selon les méthodes d'estimation des allèles nuls, les résultats ont été sensiblement différents. Il semblerait néanmoins que, quel que soit la méthode, des fréquences d'allèles nuls supérieures à 8% (seuil d'influence des allèles nuls sur les analyses selon Oddou-Muratorio et al. 2009) sont retrouvées dans au moins deux populations pour les marqueurs pn6175, PtTx4001 et pn6266. L'analyse Structure et le calcul des indices de diversité sans ces 3 marqueurs ne présentent pas de différence majeure avec les analyses à 13 marqueurs. Les résultats complets de ces analyses sont présentés en annexe. Les 13 marqueurs ont été gardés pour les analyses de cette étude.

1.2) Analyses de Structure

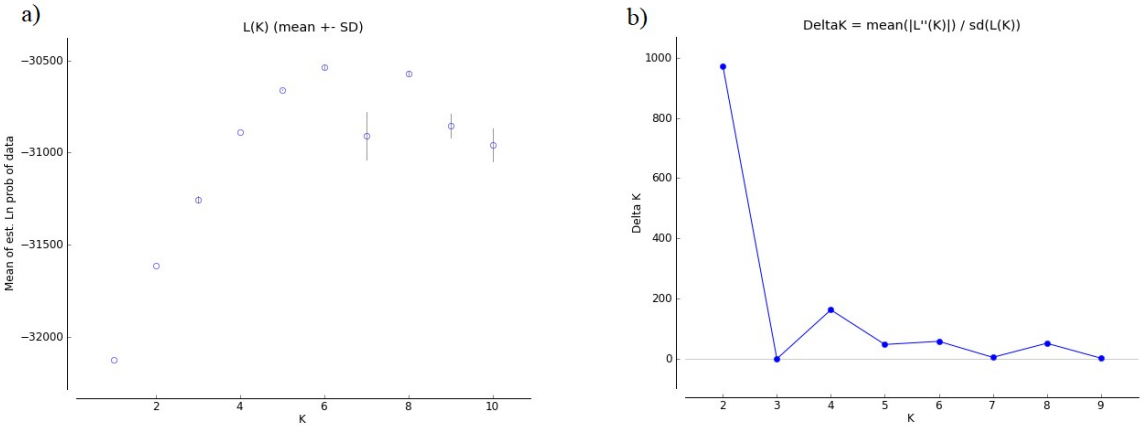


Figure 4 : Représentation de la moyenne de vraisemblance (a) et du « saut de vraisemblance » deltaK (b) en fonction du nombre K de groupes de structure pour l'analyse avec les 684 individus génotypés Ana1.

Pour Ana1, la valeur de la moyenne de vraisemblance augmente pour des valeurs allant de K=1 à K=6 où elle atteint son maximum sans que son intervalle de confiance n'augmente (figure 4a). Concernant le deltaK, ici on observe un saut de vraisemblance le plus grand pour K=2, indiquant une structuration en deux groupes (figure 4b). On observe également des sauts de vraisemblance pour K=4 et K=6 qui indiqueraient des sous-structurations.

Sur les analyses AnaQ1 et AnaQ2, les valeurs de vraisemblance et deltaK indiquent 3 groupes pour le premier ensemble et 2 groupes pour le deuxième (table 1). Avec le résultat graphique de répartition du taux d'appartenance des individus aux différents groupes selon leur provenance (figures 5), on peut noter un rapprochement entre les provenances St Guilhem et Tour sur Orb, ainsi que Gard et Ardèche.

Analyse	Echantillonnage	Nombre individus au total	Max delta K	K choisi par delta K	Max Vraisemblance	K choisi par vraisemblance
Ana1	Tous les individus	684	971.72427	2	-30537	6
Ana2	20 individus par pop, choix systématique	140	65.916411	4	-6148.733	4
Ana3	41 individus max par pop, aléatoire	240	83.19904	3	-10188.03	7
AnaQ1	Groupe majoritairement cluster 1 (Ana1 à K=2)	385	1160.2	3	-16722.5	3

AnaQ2	Groupe majoritairement cluster 2 (Ana1 à K=2)	299	131.19388	2	-13833.6	2

Table 1 : Valeurs des indicateurs du nombre de groupes de structure K le plus probable des différentes analyses réalisées.

Concernant les analyses réalisées sur des effectifs réduits mais homogènes entre les provenances, chacune donne des résultats différents de Ana1 sur les valeurs de vraisemblance et de deltaK pour le nombre de groupes à considérer (table 1). Les indicateurs, « MedMeaK » et « MaxMeaK » qui ne sont pas soumis à l'effet des effectifs hétérogènes entre provenance, nous indique un résultat de K=5 pour l'analyse Ana1. Ces critères et la séparation en 5 groupes au total suite à AnaQ1 (3 groupes) et AnaQ2 (2 groupes) indiquent ainsi une structure probable en 5 groupes (figure 6).

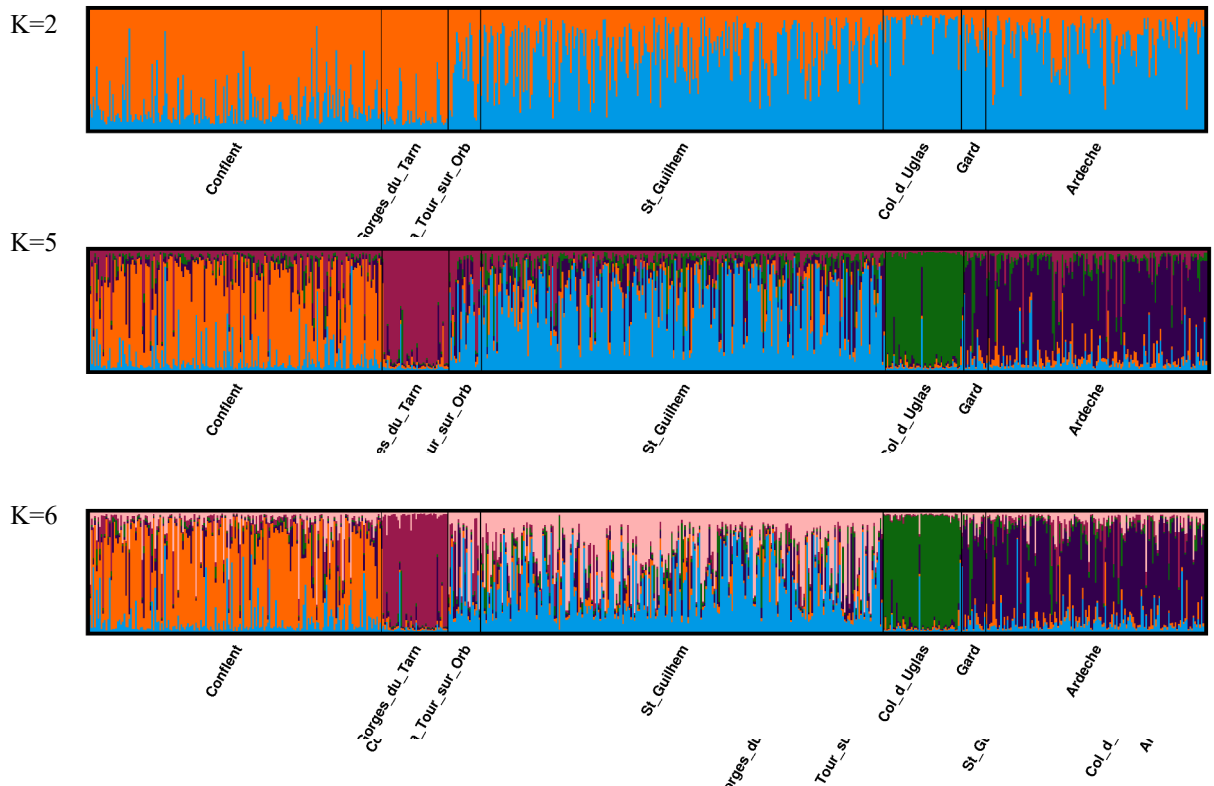


Figure 5a : Résultats graphiques des analyses Structure pour Ana1 à K=2, K=5 et K=6.Chaque bande représente un individu et chaque couleur représente son appartenance à un groupe de façon indépendante entre chaque analyse. K=2 et K=6 sont des structurations probables d'après les analyses de vraisemblance, mais c'est la structuration en K=5 qui est retenue en étudiant les sous-structures pour K=2 (Individus Q1 en majorité bleu et individus Q2 en majorité orange étudiés séparément dans AnaQ1 et AnaQ2).

K=3

Figure 5b : Résultats graphiques des analyses Structure pour AnaQ1 à K=3 et AnaQ2 à K=2.

K=2

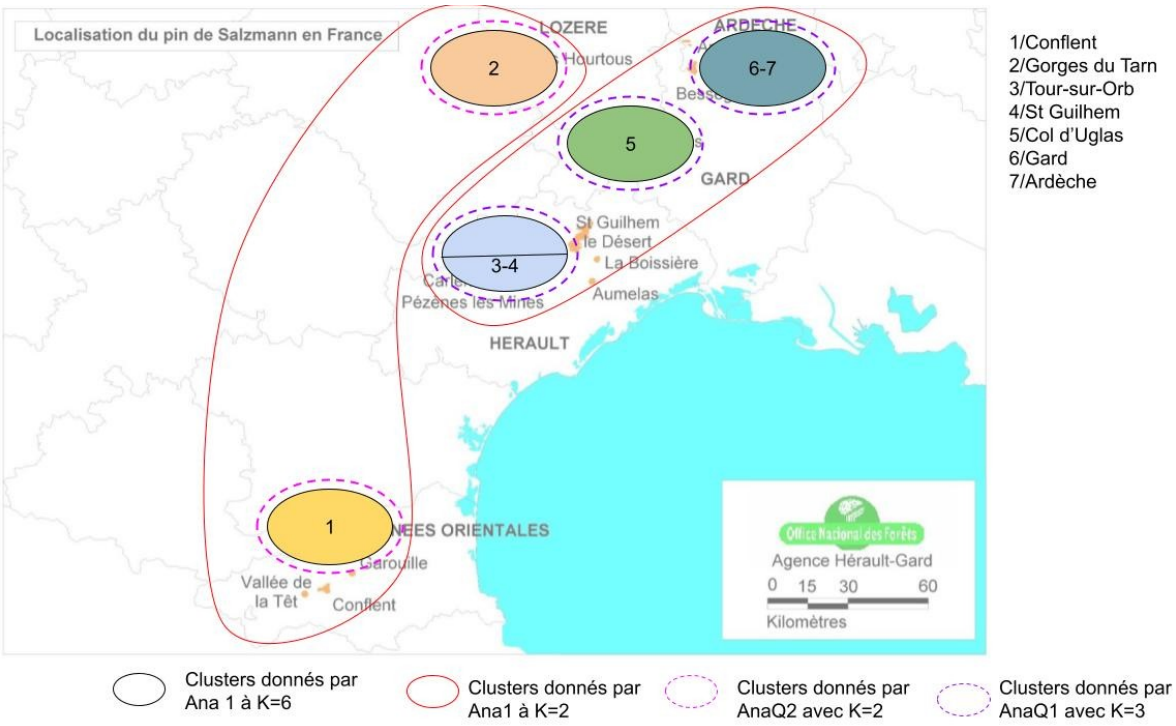


Figure 6 : Schéma résumé des groupes de structure génétiques prises en compte après les différentes analyses réalisées.

Les résultats des analyses effectuées pour chacun de ces 5 groupes de Structure confirment la présence de groupes uniformes (résultats complets des valeurs de vraisemblance et deltaK en annexe) pour quatre groupes. Seul le groupe des Gorges du Tarn fait exception, avec une sous-structuration claire en 2 groupes différents. (Figure 7).

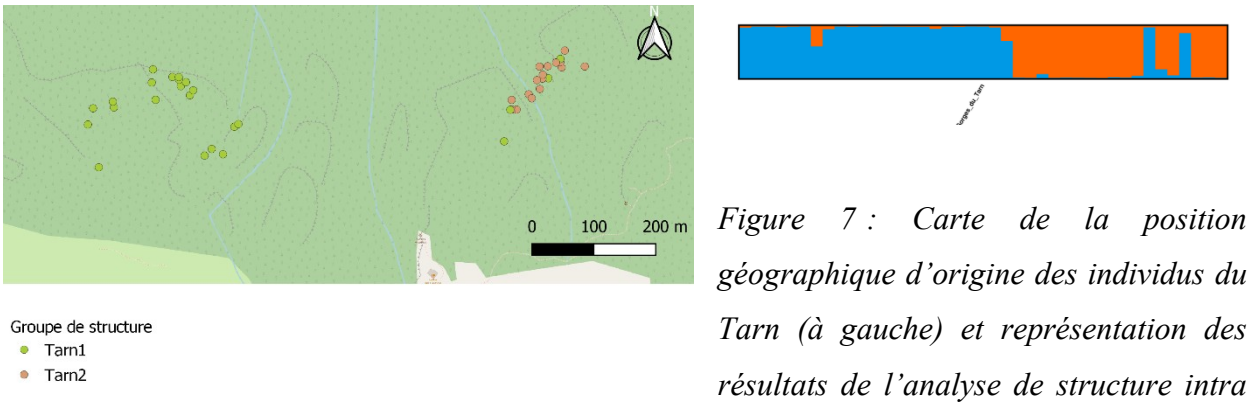


Figure 7 : Carte de la position géographique d'origine des individus du Tarn (à gauche) et représentation des résultats de l'analyse de structure intra population pour les individus du Tarn (à droite)

	Admix	Q1	Q2	Q3	Q4	Q5	Total
Guilhem-Orb	198	66	0	1	0	2	267
Conflent	103	2	72	0	0	2	179
Gard-Ardèche	88	0	0	60	1	0	149
Uglas	6	0	0	0	42	0	48
Tarn	8	0	0	0	0	33	41
							684

Table 2 : Appartenance des individus aux groupes structures avec Ana1 pour K=5 selon la provenance. Un individu est considéré appartenant à un groupe Qx si son score d'appartenance à ce groupe est supérieur à 80%, sinon il est considéré admix. En vert les individus appartenant au groupe structure majoritaire de la provenance, en rouge les non adéquats.

1.3) Indices de diversité génétique

Groupe	Effectif	NA	Nae	AR(k=12)	He	Ho	Fis	Pval(Fis<>0)	Moyenne taille allèles	Variance taille allèles
Tous groupes	684	16.38	6.34	5.55	0.7393	0.658	0.11	**0**	301.5	68.3
Conflent	179	13.46	6.07	5.38	0.738	0.681	0.077	**0**	301.7	65.5
Tarn	41	7.77	4.71	4.68	0.6706	0.575	0.144	**0**	302.3	82.3
Guilhem-Orb	267	14.08	5.56	5.31	0.7211	0.657	0.089	**0**	301.4	63.7
Uglas	48	6.23	3.54	4.05	0.6425	0.558	0.134	**0**	299.4	59.8
Gard-Ardèche	149	12.77	6.12	5.4	0.7309	0.68	0.069	**0**	301.9	72.7

Table 3 : Indices de diversité de la collection complète. NA : Nombre d'allèle moyen par marqueurs, Nae : « Effective allele number » (Nombre d'allèles à fréquences égales selon le He de la population, Nielsen et al. (2003)), AR : Nombre d'allèles raréfié pour 12 observations, He : Hétérozygotie attendue, Ho : Hétérozygotie observée, Fis : coefficient de consanguinité, Pval(Fis<>0) : P-value du test de 2000 permutations de gènes parmi les individus, Moyenne et variance de taille d'allèles indiqués en paire de bases.

L'hétérozygotie globale de la collection complète est de 0,739. On observe la diversité la plus faible pour le Col d'Uglas avec une hétérozygotie de 0,643 et la diversité la plus forte au Conflent avec une hétérozygotie de 0,738. Le coefficient de consanguinité est significativement supérieur à 0 pour toutes les populations indiquant un déficit en hétérozygotes.

1.4) Indices de différenciation génétique

G'st	Conflent	Tarn	Guilhem-Orb	Uglas	Gard-Ardèche
Conflent		**0.001**	**0.001**	**0.001**	**0.001**
Tarn	**0.184**		**0.001**	**0.001**	**0.001**
Guilhem-Orb	**0.082**	**0.154**		**0.001**	**0.001**
Uglas	**0.256**	**0.369**	**0.213**		**0.001**
Gard-Ardèche	**0.094**	**0.189**	**0.094**	**0.207**	

D	Conflent	Tarn	Guilhem-Orb	Uglas	Gard-Ardèche
Conflent		**0.001**	**0.001**	**0.001**	**0.001**
Tarn	**0.137**		**0.001**	**0.001**	**0.001**
Guilhem-Orb	**0.062**	**0.113**		**0.001**	**0.001**
Uglas	**0.192**	**0.278**	**0.156**		**0.001**
Gard-Ardèche	**0.071**	**0.141**	**0.070**	**0.152**	

Tables 4 : Indices de différenciation entre provenances de la collection complète. Valeurs de G'st (Hedrick) et D (Jost) sous la diagonale, p-value de significativité de ces indices au-dessus de la diagonale (valeurs significatives à 5% en gras)

Les indices de différenciation sont significatifs pour toutes les paires de provenances. Les deux indices indiquent le même ordre de grandeur de différenciation. Les valeurs les plus élevés de différenciation s'observent entre la provenance Col d'Uglas et toutes les autres, avec la différenciation génétique la plus grande entre Col d'Uglas et Gorges du Tarn.

2) Critères et qualité de la core collection

2.1) Etude de stabilité des critères (10 itérations, taille de 50% de la collection complète, 5 critères étudiés)

Le critère CV (couverture allélique) présente très peu d'arbres stables qui sont sélectionné sur les 10 itérations (24 sur 342 dans la CC), il existe de nombreuses combinaisons qui maximisent la couverture allélique à 1, c'est un critère peu discriminant à lui tout seul pour établir une core collection. Ensuite on a le critère AN avec 182 arbres stables, EN avec 274, et enfin SH et HE qui sont très stables (335 et 339). Parmi les 684 arbres génotypés, seul un arbre a été choisi avec 5 critères (PNS 793), 53 avec 4 critères, 159 avec 3 critères, 139 avec 2 critères, 182 avec 1 critère et 150 jamais sélectionnés.

2.2) Etude du critère AN (10 itérations, taille de 5% à 80 % de la collection complète, critère AN).

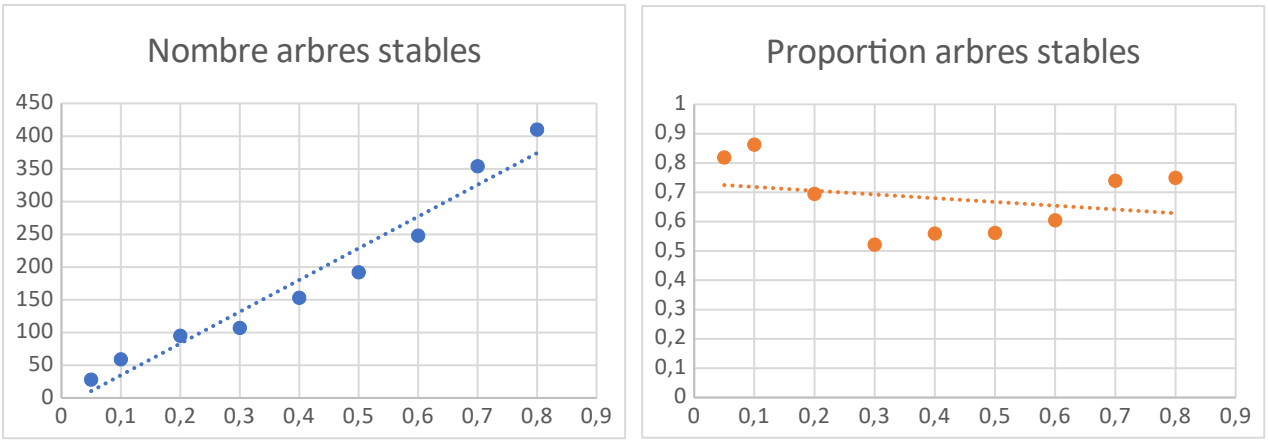
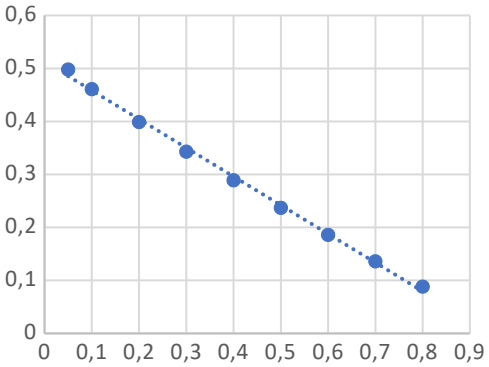


Figure 8 : Nombre (à gauche) et proportion (à droite) d'arbres stables aux 10 itérations de core collection en fonction de la taille de core collection (de 5 à 80% de la collection complète).

Le nombre d'arbres stables augmente avec la taille de core collection visé sur une tendance plutôt linéaire (Figure 8). On observe plus d'arbres stables en proportion pour des tailles de core collection extrêmes. Pas de plateau n'est observé pour la tendance du nombre d'arbres stables avec des grandes CC.

Figure 9 : Valeur de AN en fonction de la taille de Core collection



Les valeurs de AN moyennes pour les 10 itérations diminuent pour des tailles de core collection plus grande (Figure 9), ce qui fait sens, plus on a d'entrées dans la core collection plus on peut réduire la distance AN entre accessions et entrées. On peut noter une forte linéarité de cette décroissance qui indique une proportionnalité entre taille de core collection et valeur de AN. Il semblerait donc qu'il n'y a pas d'effet de saturation qui stabilise la valeur AN à partir d'une certaine taille de population.

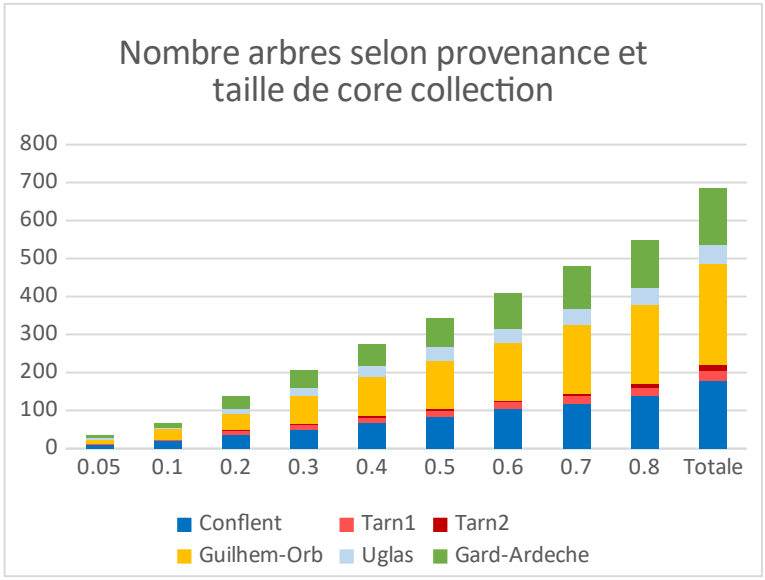
Les deux core collections obtenues en utilisant uniquement le critère AN pour l'une et AN et CV pour l'autre présentent des valeurs différentes (table 5). La valeur du critère de distance AN est inférieur lorsque AN est utilisé seul. Dans ce cas, la core collection optimise la valeur AN (valeur la plus basse possible), Pour les collections avec cet unique objectif, la couverture allélique (CV) n'atteint jamais 100%, même pour la core collection qui représente 80% de la

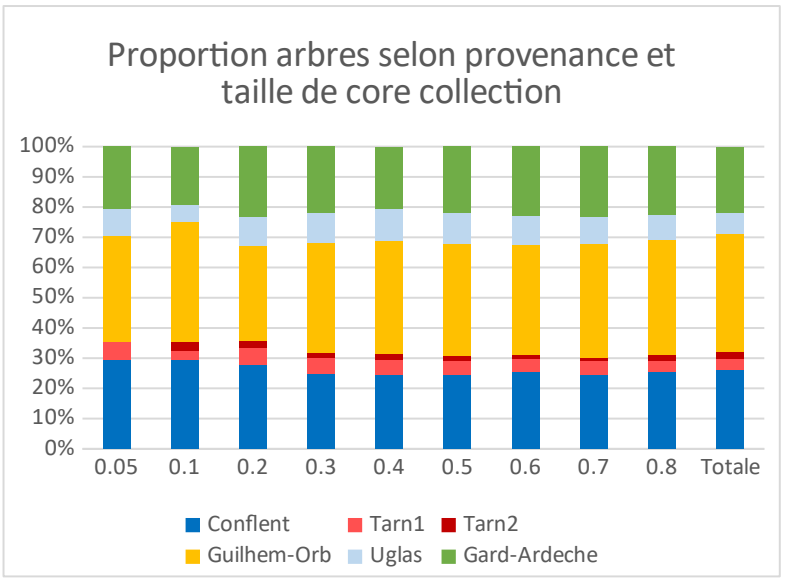
taille d'origine. En considérant ensemble les objectifs AN et CV, la valeur de la distance AN dans les collections est moins optimale de quelques décimales mais la couverture allélique atteint 100% dès une taille de core collection de 30% de la collection complète. Nous avons ainsi privilégié l'utilisation des objectifs AN et CV combinés afin de s'assurer de ne perdre aucun des 213 allèles de la collection complète même si l'on perd un peu d'optimisation du critère AN.

Objectif	Valeur Critère	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
AN	AN	0.4984	0.4616	0.3990	0.3429	0.2895	0.2367	0.1866	0.1364	0.0884
	CV	0.5962	0.7230	0.8122	0.8638	0.8873	0.9390	0.9484	0.9718	0.9765
AN et CV	AN	0.5092	0.4694	0.4020	0.3454	0.2894	0.2372	0.1866	0.1363	0.0900
	CV	0.8357	0.9671	0.9953	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table 5 : Valeurs des critère AN et CV pour les core collection de tailles allant de 5% à 80% de la collection complète en suivant l'objectif AN seul ou AN et CV combinés, avec la méthode de sélection des arbres les plus stables sur les 10 itérations.

Figure 10 : Histogramme compilé du nombre et de la proportion d'individus en fonction de la taille de la core collection et des provenances. Les provenances identifiées en groupe de structure ont ici été jointes. La provenance Tarn a été séparé en deux groupes selon les résultats de l'analyse de structure intra population.





En étudiant la représentation des provenances dans les différentes core collection réalisées (avec critères AN et CV), on peut voir que toutes les provenances sont représentées dès une taille de 10% de la collection complète (Figure 10). On remarque également que les proportions des différentes provenances au sein des core collections sont représentatives des proportions retrouvées dans la collection totale.

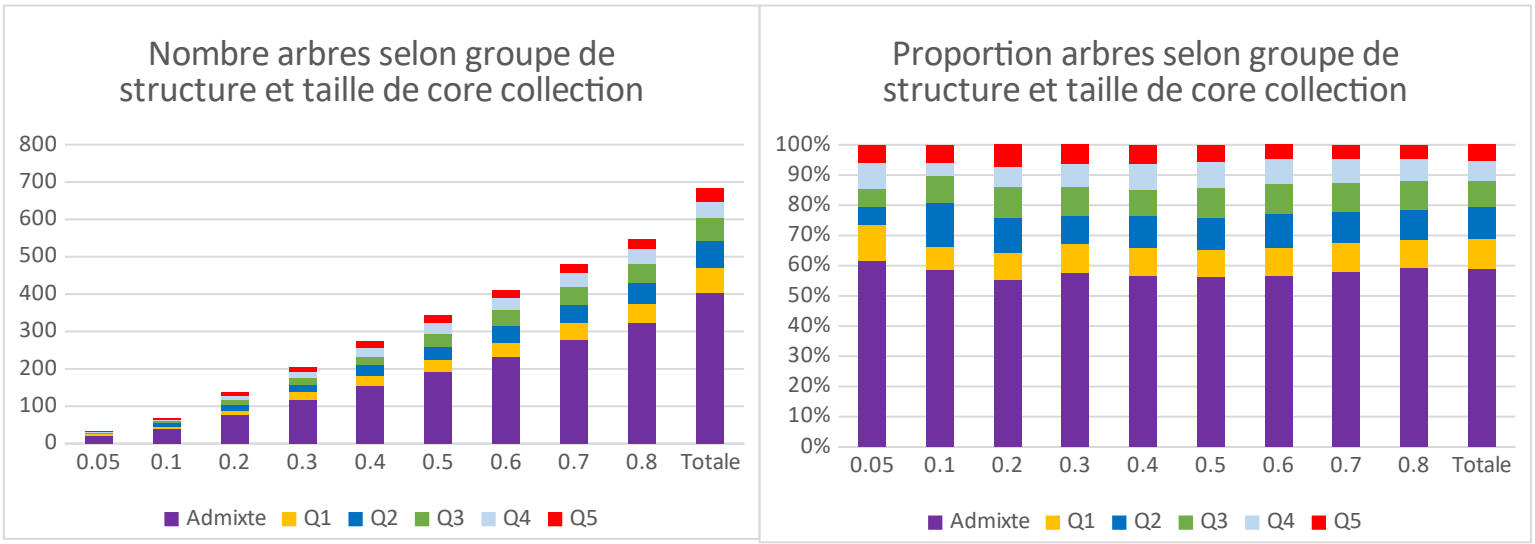


Figure 11 : Histogramme compilé du nombre et de la proportion d'individus en fonction de la taille de la core collection et des groupes de structure. Les admix correspondent à des individus dont le score d'appartenance à un groupe après l'analyse de structure n'atteint pas 80%. Les individus apparentés aux clusters de Q1 à Q5 correspondent à des individus de provenance Guilhem-Orb, Conflent, Ardèche, Uglas et Tarn dans l'ordre avec un score de plus de 80%.

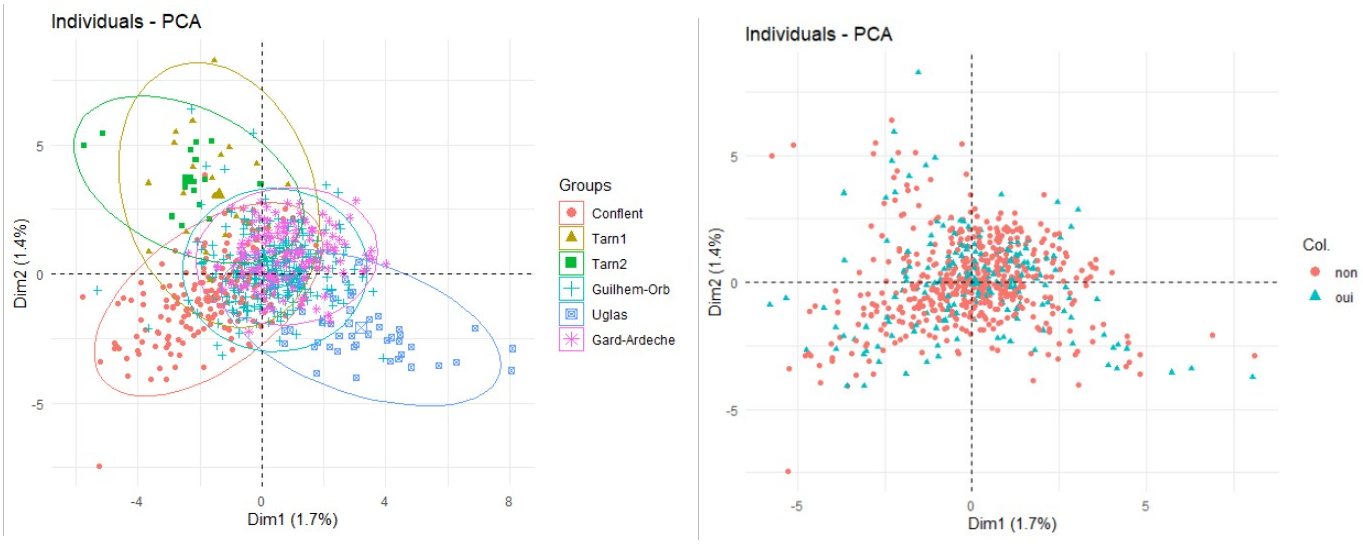
Concernant les proportions d'individus *admix* (appartenant à deux lignées à la fois, ascendance mixte) et fortement apparentés aux groupes de structure pour les core collections de différentes tailles, elles sont également représentatives de celles de la collection totale (Figure 11).

2.3) Collection cas d'étude

La collection cas d'étude a donc été réalisée pour une taille de 30% comme détaillé en méthode. Nous avons représenté sur l'analyse ACP (figure 12) l'ensemble des individus de la collection, en indiquant ceux sélectionnés pour la core collection. Les points représentants les individus sélectionnés semblent se répartir dans l'ensemble des directions de l'espace, ce qui

semble indiquer une bonne diversité sélectionnée parmi les individus de la collection grâce aux critères AN et CV.

Figure 12 : Analyse en composantes principales pour les dimensions 1 et 2 des individus de la collection selon leur groupe de provenance. Pour la figure de droite, le oui indique que l'individu est sélectionné dans la core collection, et le non qu'il ne l'est pas.



Concernant les indices de diversité, les résultats de la core collection semblent être très similaires aux résultats de la collection complète (Table 6). Tous les indices de diversité semblent être proches de la collection complète pour les moyennes de toutes provenance confondue. Afin de confirmer cela, nous avons réalisé une comparaison des valeurs de He (Hétérozygotie attendue), critère important dans la caractérisation de diversité génétique, entre la core collection et la collection complète à l'aide du package *adegenet* (Jombart et al. 2008) sur R. Un test de comparaison de He a ainsi été effectué et retourne une valeur de p-value de 0.632, ce qui ne rejette pas l'hypothèse d'égalité entre les He. On peut ainsi considérer que les He de la CC et de la collection complète ne sont pas significativement différents.

Groupe	Effectif	NA	Nae	AR(k=12)	He	Ho	Fis	Pval(Fi<>0)	Moyenne taille allèles	Variance taille allèles
Tous groupes	184	15.92	6.23	5.53	0.7375	0.65	0.119	**0**	301.4	73.5
Conflent	45	11.15	6.45	5.49	0.7418	0.695	0.064	**0**	301.9	70.7
Tarn	17	6.23	4.61	4.64	0.665	0.552	0.17	**0**	301.7	79.6

							5			
Guilhem-Orb	70	12.15	5.26	5.25	0.7203	0.649	0.1	**0**	301.2	66.4
Uglas	19	5.08	3.49	3.99	0.635	0.579	0.093	**0.036**	299	58.8
Gard-Ardèche	33	9.69	6.07	5.35	0.7125	0.671	0.06	**0.014**	302.3	76.5

Table 6 : Indices de diversité pour la core collection d'étude. Voir description des indices table x de la collection complète.

G'st	Conflent	Tarn	Guilhem-Orb	Uglas	Gard-Ardèche
Conflent		**0.001**	0.431	0.050	0.186
Tarn	**0.082**		**0.011**	0.061	**0.004**
Guilhem-Orb	0.001	**0.043**		0.719	0.305
Uglas	0.032	0.034	-0.008		**0.020**
Gard-Ardèche	0.013	**0.059**	0.004	**0.041**	

D	Conflent	Tarn	Guilhem-Orb	Uglas	Gard-Ardèche
Conflent		**0.001**	0.431	0.050	0.186
Tarn	**0.060**		**0.012**	0.061	**0.004**
Guilhem-Orb	0.001	**0.032**		0.715	0.303
Uglas	0.023	0.025	-0.006		**0.020**
Gard-Ardèche	0.010	**0.044**	0.003	**0.031**	

Table 7 : Indices de différenciation pour la core collection d'étude. Voir description table x de la collection complète pour explication.

Les deux indices de différenciation présentent des résultats similaires (Table 7). La différenciation est significative pour quatre paires de provenances ici, contrairement à la collection complète où toutes les provenances sont différenciées entre elles. Les valeurs de G'st et D sont visiblement plus faibles pour la core collection, la différenciation entre provenances est moins marquée. L'analyse Structure pour la core collection présente les groupes Gorges du Tarn et Col d'Uglas assez bien identifiés, le Conflent est légèrement identifiable et les groupes Ardèche et St-Guilhem/Tour sur Orb ne sont plus distingués entre eux (Figure en annexe).

Discussion

Core collection cas d'étude

La méthode mise au point dans cette étude nous a permis d'obtenir efficacement une une core collection représentative de la collection complète de pin de Salzmann. Sur le plan de la diversité génétique, la core collection cas d'étude obtenue présente des caractéristiques similaires à la collection complètes en se basant sur tous les indices de diversités étudiés ici (table 6). L'optimisation du critère de distance AN permet également de s'assurer d'obtenir une core collection représentant de façon homogène la collection complète (Odong et al. 2013). Notre core collection a également l'avantage de représenter la distribution de la collection complète tant pour les provenances (figure 10) que pour les groupes de Structure (figure 11), notamment avec la proportion d'individus identifiés d'ascendance mixte qui reste la même dans la core collection. Sur le plan de la différenciation génétique, notre core collection ne permet pas de différencier les groupes de provenance entre eux comme on pouvait le faire avec la collection complète (table 7). Le cas de la provenance du Col d'Uglas est particulièrement surprenant car ce groupe présente la différenciation la plus grande avec les autres groupes dans la collection complète, alors qu'il n'est différencié qu'avec le groupe Gard-Ardèche dans la core collection avec des indices de différenciation génétique beaucoup plus faibles. Une hypothèse concernant ces différences importantes est la réduction importante des effectifs pour la core collection. En effet, la core collection présente des effectifs assez faibles par population, 19 individus pour le Col d'Uglas par exemple, alors que le nombre d'allèles reste lui proche de celui de la collection complète (on a cherché à conserver tous les allèles en réduisant les effectifs). Dans une étude sur les indices Gst et D, Gerlach et al. (2010) ont mis en évidence que pour des cas de populations à faibles effectifs avec un nombre d'allèles élevés, les indices peuvent présenter des biais et ne pas se révéler significatifs, alors qu'une structuration est présente. L'analyse Structure réalisée sur la core collection est moins évidente sur la structuration de certains groupes comme St-Guilhem et Ardèche mais les groupes du Col d'Uglas et du Tarn sont bien identifiés et séparés, contrairement à ce que laisse penser les indices de différenciation. Ainsi les indices utilisés de différenciation utilisés ici présente certainement des biais.

Notre core collection cas d'étude a été construite en se basant sur le critère AN pour un objectif de core collection CC-I avec une taille de 30% de la collection complète en ne considérant pas les individus absents de la collection. Cela peut ouvrir plusieurs points de discussion :

-La taille : Pour choisir la taille de la CC cas d'étude, une méthode envisageable aurait été de se baser sur la valeur du critère de sélection, AN ici, afin de choisir la taille idéale de CC. Cependant, nous avons pu mettre en évidence que la valeur AN diminuait de façon linéaire avec la taille de la CC (donc pas de saut de AN particulièrement intéressant entre une taille et une autre), et la littérature ne nous a pas permis de déterminer une valeur de AN seuil acceptable pour une CC. Nous n'avons ainsi pas considéré le critère AN pour le choix de la taille de la CC cas d'étude. Un autre avantage de cette taille de CC est qu'elle est assez faible pour pouvoir mettre en évidence des effets potentiels du choix restreint d'individus sur la structure génétique de la CC (taux d'hétérozygotie peut être différent par exemple). Cela représente également une surface assez facilement gérable d'arbres en condition pratique. Une perspective concernant ce paramètre est d'augmenter le seuil de taille progressivement en calculant la différenciation génétique à chaque fois, afin de trouver une taille minimale avec laquelle les groupes de structure sont bien défini. Cela pourrait permettre d'éviter les biais des indices de différenciation génétique à faibles effectifs pour la core collection retenue.

-Le critère : Comme on a pu le voir, les critères utilisés présentent une variabilité importante (partie 2.1 résultats) et déterminer quel critère on souhaite utiliser impacte le résultat de la sélection. Il pourrait être envisageable de vouloir garder des individus les plus extrêmes possibles pour conserver des caractéristiques rares, cela est assez commun en conservation. Un tel objectif, de type CC-X, nous ferait ainsi utiliser le critère EN en priorité.

-Les individus absents de la collection : Nous ne les avons pas considérés mais selon les cas de figure cela est envisageable. En effet, si pour la réalisation de la core collection les moyens sont limités, que l'on ne peut pas aller chercher les individus sélectionnés non présents dans la collection de base afin de les copier (coût et ressources nécessaires), on peut ne pas les considérer lors de la réalisation de la core collection. Par contre, si l'on souhaite la core collection la plus optimale en considérant tous les individus dont nous avons le génotype, alors on peut garder ces individus non présents dans la collection et décider d'aller récolter les greffons directement sur les habitats naturels.

-L'outil : Dans une étude réalisée par Jeong et al (2017) qui présente le logiciel GenoCore, une comparaison est effectuée pour différents logiciels de construction de core collection (GenoCore, Core Hunter, MSTRAT et PowerCore). Pour deux jeux de données testés, les résultats indiquent de meilleurs scores de diversité au sein de la core collection sélectionnée pour GenoCore et Core Hunter. L'un des principaux avantages de GenoCore sur Core Hunter

est qu'il est mieux adapté pour des jeux de données très grands en utilisant moins de mémoire et de temps d'exécution sur ordinateur. Cependant, Core Hunter est un outil qui permet de bien définir les critères de mesure de qualité de la core collection lors de la sélection des accessions, afin d'obtenir une core collection mieux adapté aux objectifs vu précédemment.) Idéalement, tester les différents outils existants avec nos données pour ensuite pouvoir comparer les résultats obtenus aurait été la méthode la plus complète à adopter. Cependant, ce type d'analyse, bien que complet, aurait nécessité des ressources qui sortent du cadre de notre étude, notamment en termes de temps. Pour notre étude, nous avons ainsi décidé de choisir l'outil Core Hunter en nous basant sur la littérature et nous consacrer en particulier sur celui-ci en l'étudiant de la façon la plus précise possible.

Perspectives sur la méthode de constitution de la core collection

La démarche analytique de sélection de la core collection construite sous R a été testée par un utilisateur lambda pour un jeu de données d'une autre espèce génotypée avec des marqueurs microsatellites (sapin de Bornmuller) et cela a abouti à une sélection concluante. La méthode est ainsi réutilisable pour d'autres espèces qui utilisent les mêmes données d'entrée, avec des critères d'optimisation de la core collection qui peuvent être personnalisés selon les objectifs de la core collection. Cette méthode présente plusieurs perspectives. La prise en compte de phénotypes en tant que critères de sélection en est une première. En effet, cela est possible avec l'outil Core Hunter mais n'a pas été réalisé ici par manque de temps. Des caractéristiques phénotypiques comme l'âge des arbres, le nombre de cônes produits et le nombre de clones dans la collection (proxy de la facilité de greffage) sont des traits dont nous possédons déjà les données et qui pourraient apporter des critères supplémentaires dans la sélection de core collection. Le critère de l'âge notamment été étudié brièvement dans notre étude, nous avons pu mettre en évidence un effet de la provenance sur l'âge des arbres (test Anova en annexe). L'étude de ces traits phénotypiques constituent une réelle perspective future de notre étude. Une autre perspective est l'adaptation de la méthode à différents marqueurs génétiques. En effet, notre méthode n'a été testée que pour des données basées sur des marqueurs microsatellites. Il existe néanmoins une dizaine de marqueurs génétiques différents qui pourraient être utilisés en caractérisation de génotype d'une collection qu'on voudrait réduire. On peut notamment citer le polymorphisme nucléotidique (SNPs) ou le séquençage nouvelle génération (NGS) qui constituent des méthodes de marquage permettant d'obtenir de très grands jeux de données. Si le jeu de donnée n'est pas de taille trop importante notre méthode peut prendre en compte différents marqueurs. Nous aurons

prochainement à disposition des données génétiques de pin de Salzmann avec des marqueurs SNPs (80 marqueurs bialéliques). Une perspective d'étude serait de comparer les résultats donnés par ces deux types de marqueurs. Notre méthode peut prendre en compte différents marqueurs mais elle n'est pas adaptée aux grands jeux de données. Une alternative assez récente est le programme GenoCore (Jeong 2017) qui peut supporter ce genre de données, mais il a le désavantage de ne pas prendre en compte différents critères de core collection, ce qui rend l'obtention d'un objectif de core collection de différents types (CC-I, CC-X ou CC-D) plus complexe.

Structure et conservation de la diversité génétique des peuplements de pin de Salzmann

Avec cette étude nous avons pu étudier la Structure des populations de pin de Salzmann. Une première analyse (Ana1 à K=2) a pu mettre en évidence une structuration en deux groupes qui viendraient probablement de deux lignées anciennes, avec rapprochement entre les populations du Conflent et des Gorges du Tarn d'une part et les autres provenances d'autre part. Ce résultat n'avait pas été mis en évidence lors d'études précédentes à notre connaissance. La structuration en cinq groupes de structures était attendue avec une appartenance à un même groupe génétique pour les provenances Tour sur Orb et St-Guilhem ainsi que Gard et Ardèche dû à la proximité géographique. Ces groupes avaient pu être mis en évidence dans des études précédentes (Compte rendu Fady 2017). Les analyses Ana2 et Ana3 qui reposent sur des effectifs réduits homogènes ont montré des résultats assez différents en termes de nombre de groupes et structuration de ces groupes qui diffèrent des structurations déjà décrites, nous ne les avons pas considérés mais elles confirment un effet non négligeable de la taille des effectifs sur les analyses Structure (déjà discuté par Puechmaille 2016). La sous-structuration présente dans les Gorges du Tarn est également intéressante à observer, en étudiant la position spatiale d'origine des individus, nous avons pu nous rendre compte que la délimitation entre les deux groupes est très marquée. La forte variance de la taille des allèles (table 3), plus élevée que pour les autres provenances, en est une preuve supplémentaire. Les individus génotypés du Tarn ont été récoltés sur deux zones assez distantes qui pourraient expliquer la présence de deux groupes de structure pour cette provenance. Une analyse de l'apparentement entre ces arbres pourrait être envisagée afin de tester si la structuration résulte d'une structuration par famille. Une autre hypothèse serait une origine plus ancienne, par exemple un isolement géographique qui aurait engendré une divergence génétique entre ces deux groupes. C'est une perspective d'étude afin de mieux comprendre la structuration génétique de ces populations. La population du Col d'Uglas a pu être identifié comme étant la moins riche en diversité génétique (plus faible He) et également la plus différenciée aux autres (forts indices de différenciation). On peut supposer un phénomène de type goulot d'étranglement qui aurait eu lieu, réduisant la diversité de cette population en la différenciant davantage des autres.

La relation entre provenance géographique et groupe de structure est globalement cohérente sauf pour 8 individus génotypés qui présentent des groupes de structure atypiques pour leur provenance (table 5). Nous nous sommes donc interrogés sur la provenance de ces individus, notamment 2 arbres qui appartiennent à la lignée St Guilhem alors qu'ils ont été

échantillonnés au Conflent (deux provenances éloignées de plus de 160 km). Nous ne pouvons pas exclure l'introduction par plantation d'une provenance « non locale ». Néanmoins, si l'on considère les âges des arbres concernés (donner les âges), ces plantations auraient dû être réalisées avant 1860. Or, à connaissance d'experts, il n'y a pas eu de reboisements ou de déplacements entre les peuplements de pin de Salzmann avant le RTM en 1860 (Calas 1900.). L'hypothèse de déplacement anthropique antérieur au RTM semble à écarter. Pour ce qui est du déplacement naturel, on pourrait considérer du pollen ou des semences emportés par le vent, mais les distances sont forcément faibles dans ce cas, ce qui ne marche pas avec nos observations. On pourrait penser à un transport de graines par les oiseaux sur de longues distances mais cela devrait être rarissime, alors que nous avons 8 cas ici ce qui traduirait un phénomène relativement fréquent qui, sur la durée, aurait provoqué une certaine homogénéisation génétique des peuplements. Le déplacement naturel semble également à écarter. Ces résultats troublants pourraient ainsi être dû à des erreurs lors des différentes manipulations : un individu d'une provenance aurait été catégorisé dans une autre par exemple, ou alors ces résultats seraient dû à des biais de l'analyse Structure qui n'aurait pas pu identifier clairement l'appartenance à un certain groupe pour des individus avec certains allèles.

La core collection que nous proposons va permettre de conserver la diversité génétique de la collection nationale de pin de Salzmann en la réduisant efficacement. Cette conservation est ici à l'échelle globale, en considérant l'ensemble des provenances de pin de Salzmann en France. A l'échelle des territoires, il existe une demande pour s'engager à la conservation des ressources génétiques locales (e.g. cas de la vallée du Gardon de Mialet, site Natura 2000). La routine R que nous avons mise en place permettra aussi de réaliser des core collection à l'échelle des provenances pour le maintien de la diversité génétique locale. On peut facilement prévoir des programmes de restauration de la diversité génétique locale qui verront le jour et pour lesquelles cette méthode pourra être utilisée. La routine mise au point dans ce projet a été pensée pour être facilement employable et réutilisable, elle pourra ainsi être utilisé dans un cadre plus général avec d'autre espèces au sein de la CRGF.

Bibliographie

Arbez M. “Intraspecific hybridizations in European black pines (Pinus nigra Arn.) ”, Rapport interne INRA, 1980.

Berg, E. E., and J. L. Hamrick. “Quantification of Genetic Diversity at Allozyme Loci.” *Canadian Journal of Forest Research*, vol. 27, no. 3, Mar. 1997, pp. 415–24, https://doi.org/10.1139/x96-195.

Calas J. “Le pin laricio de Salzmann. Imprimerie nationale”, 1900, 50p.

Callen, D. F., et al. “Incidence and Origin of ‘Null’ Alleles in the (AC)n Microsatellite Markers.” *American Journal of Human Genetics*, vol. 52, no. 5, May 1993, pp. 922–27.

Cavalli-Sforza, L. L., and A. W. F. Edwards. “Phylogenetic Analysis. Models and Estimation Procedures.” *American Journal of Human Genetics*, vol. 19, no. 3 Pt 1, May 1967, pp. 233–57, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1706274/.

Ceballos, Gerardo, et al. “Biological Annihilation via the Ongoing Sixth Mass Extinction Signaled by Vertebrate Population Losses and Declines.” *Proceedings of the National Academy of Sciences*, vol. 114, no. 30, July 2017, https://doi.org/10.1073/pnas.1704949114.

Chapuis, Marie-Pierre, and Arnaud Estoup. “Microsatellite Null Alleles and Estimation of Population Differentiation.” *Molecular Biology and Evolution*, vol. 24, no. 3, Mar. 2007, pp. 621–31, https://doi.org/10.1093/molbev/msl191.

Dąbrowski, M. J., et al. “Reliability Assessment of Null Allele Detection: Inconsistencies between and within Different Methods.” *Molecular Ecology Resources*, vol. 14, no. 2, 2014, pp. 361–73, https://doi.org/10.1111/1755-0998.12177.

De Beukelaer, Herman, et al. “Core Hunter 3 : Flexible Core Subset Selection.” *BMC BIOINFORMATICS*, vol. 19, 2018, https://doi.org/10.1186/s12859-018-2209-z.

Desgroux, A. Joyeuau, C. Bastianelli, C. Lefevre, F. “Programme national pour la Conservation des Ressources Génétiques Forestières en France”, poster de la CRGF, 2020, https://crgf.inrae.fr/.

Earl, Dent A., and Bridgett M. vonHoldt. “STRUCTURE HARVESTER: A Website and Program for Visualizing STRUCTURE Output and Implementing the Evanno Method.” *Conservation Genetics Resources*, vol. 4, no. 2, 2012, pp. 359–61, https://doi.org/10.1007/s12686-011-9548-7.

Evanno, G., et al. “Detecting the Number of Clusters of Individuals Using the Software Structure: A Simulation Study.” *Molecular Ecology*, vol. 14, no. 8, 2005, pp. 2611–20, https://doi.org/10.1111/j.1365-294X.2005.02553.x.

Fady B. “ Programme global de conservation des populations françaises de pin de Salzmann : Compte rendu scientifique final du partenaire INRA”, 2017.

Gerlach, Gabriele, et al. “Calculations of Population Differentiation Based on GST and D: Forget GST but Not All of Statistics!: NEWS AND VIEWS: COMMENT.” *Molecular Ecology*, vol. 19, no. 18, 2010, pp. 3845–52, https://doi.org/10.1111/j.1365-294X.2010.04784.x.

Gilpin, E. et M. Soulé, M. E. “Minimum Viable Populations : Processus d'extinction des espèces”. In M. E. Soulé, ed. Conservation Biology: The Science of Scarcity and Diversity.Sinauer, Sunderland, Mass, 1986, pp. 19-34.

Giovannelli G. “Histoire évolutive et diversité adaptative du pin noir, Pinus nigra Arn., à l'échelle de son aire de répartition”. Thèse de doctorat en sciences, Aix-Marseille Université, 2017.

Hardy, Olivier J., and Xavier Vekemans. “Spagedi: A Versatile Computer Program to Analyse Spatial Genetic Structure at the Individual or Population Levels.” *Molecular Ecology Notes*, vol. 2, no. 4, 2002, pp. 618–20, https://doi.org/10.1046/j.1471-8286.2002.00305.x.

Hedrick, Philip W. “A STANDARDIZED GENETIC DIFFERENTIATION MEASURE.” *Evolution*, vol. 59, no. 8, 2005, pp. 1633–38, https://doi.org/10.1111/j.0014-3820.2005.tb01814.x.

Jakobsson, M., and N. A. Rosenberg. “CLUMPP: A Cluster Matching and Permutation Program for Dealing with Label Switching and Multimodality in Analysis of Population Structure.” *Bioinformatics*, vol. 23, no. 14, July 2007, pp. 1801–06, https://doi.org/10.1093/bioinformatics/btm233.

Jeong, Seongmun, et al. “GenoCore: A Simple and Fast Algorithm for Core Subset Selection from Large Genotype Datasets.” *PLOS ONE*, edited by Alexandre G. de Brevern, vol. 12, no. 7, July 2017, p. e0181420, https://doi.org/10.1371/journal.pone.0181420.

Jombart, Thibaut. “Adegenet: A R Package for the Multivariate Analysis of Genetic Markers.” *Bioinformatics*, vol. 24, no. 11, June 2008, pp. 1403–05, https://doi.org/10.1093/bioinformatics/btn129.

Jost, Lou. “G_{ST} and Its Relatives Do Not Measure Differentiation.” *Molecular Ecology*, vol. 17, no. 18, 2008, pp. 4015–26, https://doi.org/10.1111/j.1365-294X.2008.03887.x.

Kalinowski, Steven T., et al. “Ml-Relate: A Computer Program for Maximum Likelihood Estimation of Relatedness and Relationship.” *Molecular Ecology Notes*, vol. 6, no. 2, 2006, pp. 576–79, https://doi.org/10.1111/j.1471-8286.2006.01256.x.

Kopelman, Naama M., et al. "CLUMPAK : A Program for Identifying Clustering Modes and Packaging Population Structure Inferences across *K*." *Molecular Ecology Resources*, vol. 15, no. 5, 2015, pp. 1179–91, https://doi.org/10.1111/1755-0998.12387.
Mounce, Ross, et al. "Ex Situ Conservation of Plant Diversity in the World's Botanic Gardens." *Nature Plants*, vol. 3, no. 10, 2017, pp. 795–802, https://doi.org/10.1038/s41477-017-0019-3.
Nielsen, Rasmus, et al. "Estimating Effective Paternity Number in Social Insects and the Effective Number of Alleles in a Population." *Molecular Ecology*, vol. 12, no. 11, 2003, pp. 3157–64, https://doi.org/10.1046/j.1365-294X.2003.01994.x.
Oddou-Muratorio, Sylvie, et al. "Population Estimators or Progeny Tests: What Is the Best Method to Assess Null Allele Frequencies at SSR Loci?" *Conservation Genetics*, vol. 10, no. 5, 2009, pp. 1343–47, https://doi.org/10.1007/s10592-008-9648-4.
Odong, T. L., et al. "Quality of Core Collections for Effective Utilisation of Genetic Resources Review, Discussion and Interpretation." *Theoretical and Applied Genetics*, vol. 126, no. 2, 2013, pp. 289–305, https://doi.org/10.1007/s00122-012-1971-y.
Peakall, R., and P. E. Smouse. "GenAlEx 6.5: Genetic Analysis in Excel. Population Genetic Software for Teaching and Research--an Update." *Bioinformatics*, vol. 28, no. 19, Oct. 2012, pp. 2537–39, https://doi.org/10.1093/bioinformatics/bts460.
Pritchard, Jonathan K., et al. "Inference of Population Structure Using Multilocus Genotype Data." *Genetics*, vol. 155, no. 2, June 2000, pp. 945–59, https://doi.org/10.1093/genetics/155.2.945.
Puechmaille, Sebastien J. "The Program STRUCTURE Does Not Reliably Recover the Correct Population Structure When Sampling Is Uneven: Subsampling and New Estimators Alleviate the Problem." *Molecular Ecology Resources*, vol. 16, no. 3, 2016, pp. 608–27, https://doi.org/10.1111/1755-0998.12512.
Quezel P. and Barbero M. "Signification phytoécologique et phytosociologique des peuplements de Pin de Salzmann en France. " Ecologia mediterranea – XIV, 1988, pp 41-63.
Rosenberg, Noah A. "Distruct: A Program for the Graphical Display of Population Structure: PROGRAM NOTE." *Molecular Ecology Notes*, vol. 4, no. 1, Dec. 2003, pp. 137–38, https://doi.org/10.1046/j.1471-8286.2003.00566.x.
Scotti-Saintagne, Caroline, et al. "Recent, Late Pleistocene Fragmentation Shaped the Phylogeographic Structure of the European Black Pine (Pinus Nigra Arnold)." *Tree Genetics & Genomes*, vol. 15, no. 5, 2019, p. 76, https://doi.org/10.1007/s11295-019-1381-2.
Shannon, C. E. "A Mathematical Theory of Communication." *Bell System Technical Journal*, vol. 27, no. 3, 1948, pp. 379–423, https://doi.org/10.1002/j.1538-7305.1948.tb01338.x.
Soleimani, Behnaz, et al. "Comparison Between Core Set Selection Methods Using Different Illumina Marker Platforms: A Case Study of Assessment of Diversity in Wheat." *Frontiers in Plant Science*, vol. 11, July 2020, p. 1040, https://doi.org/10.3389/fpls.2020.01040.
Thachuk, Chris, et al. "Core Hunter: An Algorithm for Sampling Genetic Resources Based on Multiple Genetic Measures." *BMC Bioinformatics*, vol. 10, no. 1, 2009, p. 243, https://doi.org/10.1186/1471-2105-10-243.
Van Oosterhout, Cock, et al. "Micro-Checker: Software for Identifying and Correcting Genotyping Errors in Microsatellite Data." *Molecular Ecology Notes*, vol. 4, no. 3, 2004, pp. 535–38, https://doi.org/10.1111/j.1471-8286.2004.00684.x.
Vernet, J.L and al. "Eco-histoire de la Forêt de Pinus nigra Arnold ssp. Salzmanni (Dunal) Franco de Saint-Guilhem–le-Désert (Hérault, France) ". Forêt, archéologie et environnement, Nancy, Office national des forêts, Institut national de la recherche agronomique et direction régionale des Affaires culturelles de Lorraine, 2004, p. 87-96.
Wright, Sewall. *Evolution and the Genetics of Populations, Volume 4: Variability Within and Among Natural Populations*. University of Chicago Press, 1978, https://press.uchicago.edu/ucp/books/book/chicago/E/bo3642015.html.

Annexes

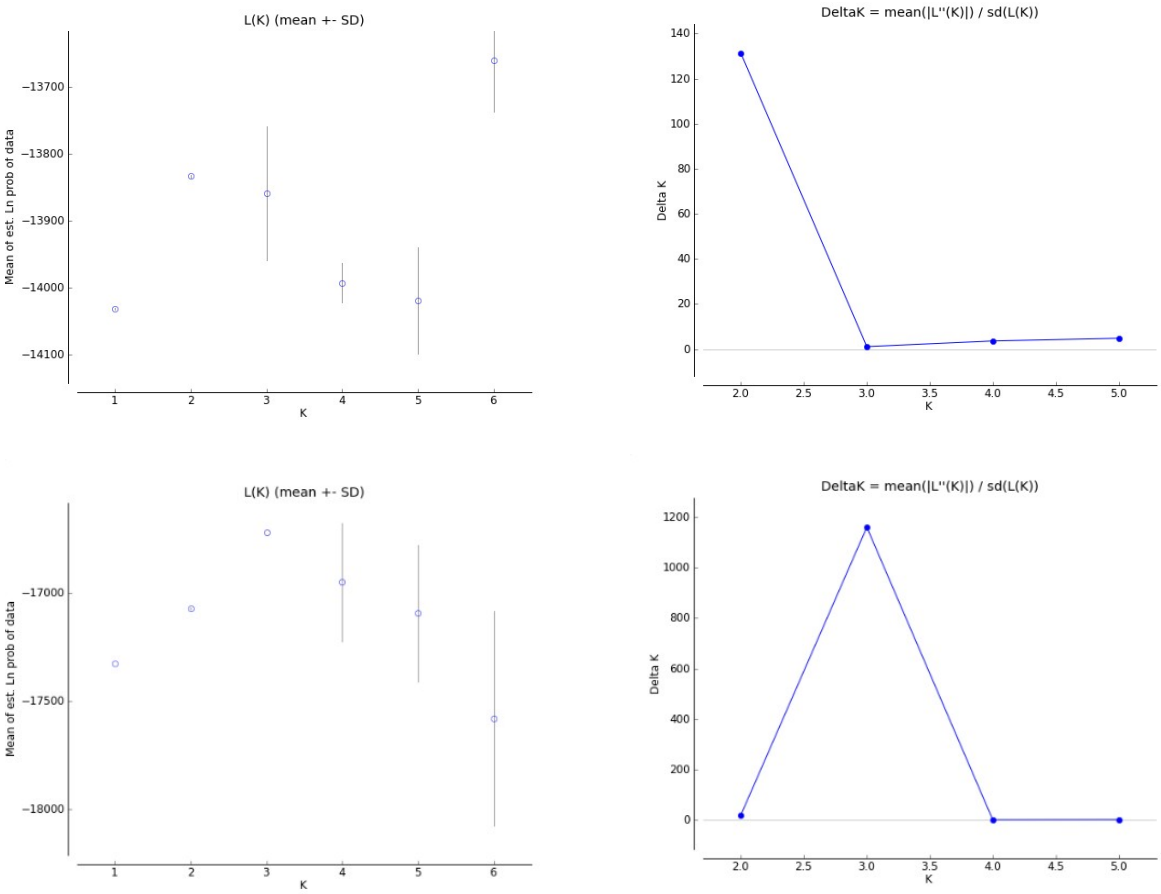


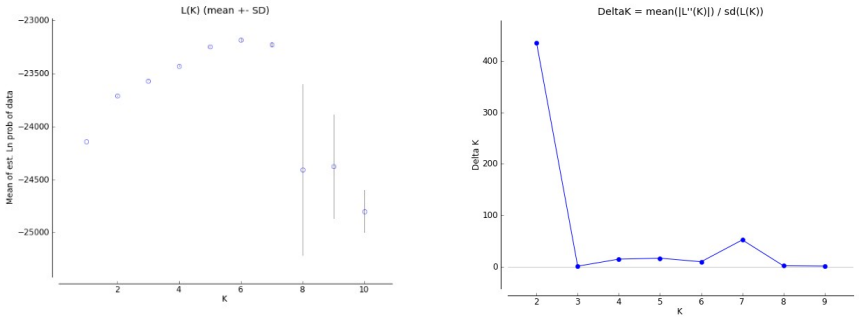
Figure A1 : Représentation de la moyenne de vraisemblance et du « saut de vraisemblance » deltaK en fonction du nombre K de groupes de structure pour les analyses AnaQ1 en haut et AnaQ2 en bas

Analyse	Nombre individus au total	Max delta K	K choisi par deltaK	Max Vraisemblance	K choisi par vraisemblance
AnaConflent	179	35.301202	3	-8398.3667	1
AnaTarn	41	363.97129	2	-1392.7	3
AnaOrb-Guilhem	267	4.135893	2	-12038.7	1
AnaUglas	48	4.710138	3	-1478.3	1
AnaGard-Ardeche	134	24.134905	2	-6789.3	1

Table A1 : Valeurs des indicateurs du nombre de groupes de structure K le plus probable des différentes analyses intra-provenances. Les valeurs de delta K inférieur à 50 sont très faibles et le K choisi ne représente pas une réalité à considérer, pour les cas probables de K=1 il faut se fier à la moyenne de vraisemblance.

	MICROCHECKER										ML-Null					FreeNA				
	Ardèche		Conflent		Guilhem		Tarn		Uglas		Ardèche	Conflent	Guilhem	Tarn	Uglas	Ardèche	Conflent	Guilhem	Tarn	Uglas
Locus	Nul présent	Freq	Nul présent	Freq	Nul présent	Freq	Nul présent	Freq	Nul présent	Freq	Nul présent	Nul présent	Nul présent	Nul présent	Nul présent	Freq	Freq	Freq	Freq	Freq
PHA_4783	Non	0.006	Non	0.010	Non	0.030	Non	-0.111	Non	-0.038	0.171	0.325	0.124	0.753	0.533	0.004	0.006	0.021	0.000	0.000
PHA_6062	Non	0.011	Non	-0.018	Non	-0.035	Non	-0.027	Non	-0.071	0.429	0.858	0.988	0.624	0.889	0.000	0.000	0.000	0.000	0.000
pn1403	Oui	0.033	Non	0.021	Non	0.013	Oui	0.190	Oui	0.114	0.077	0.001	0.087	0.000	0.023	0.021	0.022	0.011	0.180	0.083
pn2153	Non	-0.021	Non	0.024	Non	0.022	Non	0.082	Non	-0.011	0.569	0.362	0.103	0.275	0.000	0.000	0.000	0.017	0.000	0.000
pn2246	Oui	0.041	Oui	0.031	Oui	0.087	Non	-0.008	Oui	0.122	0.000	0.001	0.000	0.477	0.000	0.037	0.026	0.077	0.000	0.110
pn4379	Oui	0.055	Oui	0.031	Oui	0.033	Non	0.030	Non	0.043	0.004	0.023	0.001	0.159	0.038	0.041	0.024	0.031	0.012	0.036
pn6175	Oui	0.051	Oui	0.111	Oui	0.073	Oui	0.098	Oui	0.209	0.002	0.000	0.000	0.008	0.000	0.044	0.107	0.069	0.080	0.184
pn6266	Oui	0.073	Oui	0.066	Oui	0.103	Oui	0.170	Non	0.053	0.000	0.003	0.005	0.001	0.000	0.069	0.056	0.076	0.136	0.069
pn6360	Non	0.013	Non	-0.007	Non	0.027	Non	0.042	Non	0.055	0.602	0.108	0.004	0.240	0.012	0.000	0.008	0.019	0.000	0.051
pn7754	Non	0.002	Non	0.014	Oui	0.021	Non	0.038	Oui	0.113	0.037	0.104	0.071	0.012	0.000	0.013	0.008	0.012	0.038	0.100
pn8747	Oui	0.059	Non	0.014	Oui	0.053	Non	-0.031	Non	0.012	0.159	0.259	0.051	0.353	0.217	0.010	0.005	0.030	0.002	0.014
PtTX3107	Non	0.010	Non	0.008	Non	0.013	Oui	0.176	Non	0.031	0.037	0.407	0.257	0.002	0.172	0.027	0.000	0.009	0.148	0.027
PtTX4001	Oui	0.061	Oui	0.144	Oui	0.102	Oui	0.147	Non	-0.007	0.036	0.000	0.000	0.001	0.296	0.031	0.128	0.107	0.149	0.008

Table A2 : Résultats complets des analyses d'allèles nuls avec les 3 logiciels Microchecker, ML-Null et FreeNA. Les colonnes Nul présent indiquent la présence d'allèle nuls dans le locus considéré, en rouge les p-value<0.05 avec le logiciel considéré. Les colonnes Freq indiquent la fréquence estimée d'allèles nuls, en jaune les fréquences > 0.08.



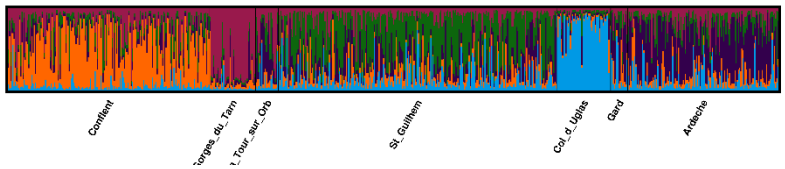


Figure A4 : Résultats de l'étude Structure avec 10 marqueurs, résultats similaires à l'analyse à 13 marqueurs

Groupe	Effectif	NA	Nae	AR(k=50)	He	Ho	Fis	Pval(Fis<>0)
Tout groupe	684	16.1	6.18	9.58	0.719	0.664	0.076	**0**
Conflent	179	13.1	5.74	8.88	0.7162	0.696	0.028	**0.0135**
Tarn	41	7.5	4.91	7.04	0.6531	0.593	0.093	**0.0025**
Guilhem-Orb	267	13.7	5.54	9.18	0.7056	0.666	0.057	**0**
Uglas	48	5.9	3.16	5.5	0.5985	0.534	0.109	**0**
Gard-Ardeche	149	12.2	6.12	9.01	0.7177	0.68	0.052	**0**

Table A3 : Indices de diversité pour l'analyse à 10 marqueurs

Figure A3 : Représentation de la moyenne de vraisemblance et du « saut de vraisemblance » deltaK en fonction du nombre K de groupes de structure pour l'analyse à 10 marqueurs

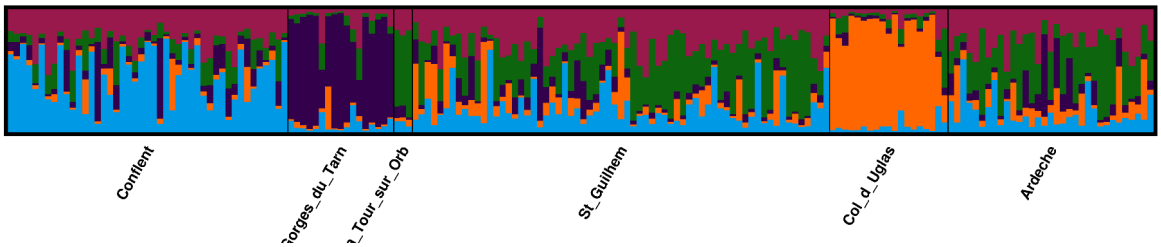


Figure A5 : Résultat graphique de l'analyse Structure de la core collection cas d'étude pour K=5.

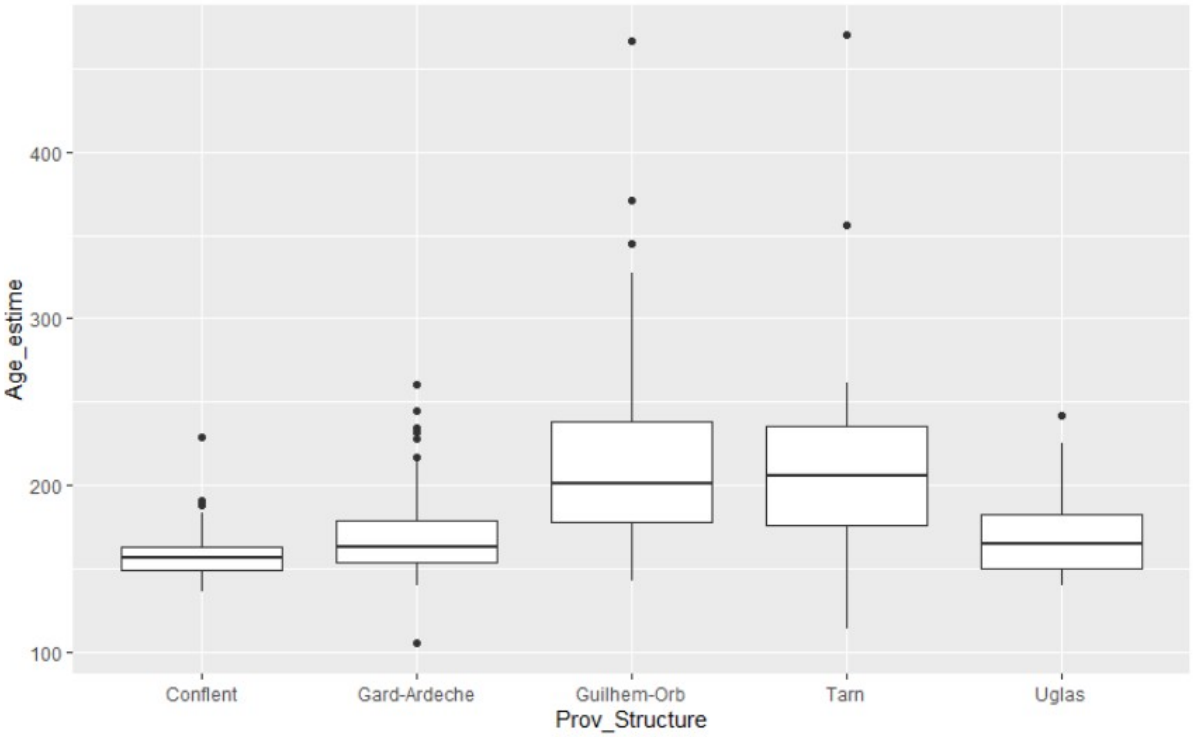


Figure A6 : Box-plot de l'âge estimé des arbres présents dans les différentes provenances. L'effet de la provenance sur l'âge est significatif d'après un test Anova et des tests de comparaisons par paire indiquent 2 groupes pour lesquelles les âges ne sont pas significativement différents selon la provenance, Conflent, Gard-Ardèche et Uglas d'un côté et St-Guillhem-Orb et Tarn de l'autre.

Résumé

En contexte de grande crise de la biodiversité, la conservation d'espèces vulnérables est nécessaire et urgente. Les écosystèmes forestiers ayant un rôle écologique majeur, notamment dans la lutte contre le changement climatique, leur conservation est une priorité. Une approche permettant de conserver ces écosystèmes est la conservation des ressources génétiques, dont la CRGF (Commission des Ressources Génétiques Forestières) en est responsable en France. Cette conservation passe par des unités conservatoires in situ mais également des approches ex situ pour des espèces très vulnérables comme le pin de Salzmann (Pinus nigra subsp. salzmannii), sous espèce du pin noir endémique en zone méditerranéenne. Dans cette étude, nous avons caractériser la structuration génétique de la collection nationale de pin de Salzmann afin de mettre au point une méthode permettant de créer une collection de base (Core Collection) réduite de cette collection en conservant la diversité génétique. Cette méthode permet efficacement de conserver la diversité génétique de la collection et a été pensée pour être facilement réutilisable avec d'autre espèces. Elle pourra ainsi être appliqué pour la constitution de core collection d'autres espèces forestières dans le cadre de la CRGF.

Abstract

In the context of a major biodiversity crisis, the conservation of vulnerable species is necessary and urgent. As forest ecosystems play a major ecological role, particularly in the fight against climate change, their conservation is a priority. One approach to conserve these ecosystems is the conservation of genetic resources, for which the CRGF (Commission des Ressources Génétiques Forestières) is responsible in France. This conservation involves in situ conservation units but also ex situ approaches for highly vulnerable species such as Salzmann pine (Pinus nigra subsp. salzmannii), a subspecies of black pine endemic to the Mediterranean area. In this study we studied the genetic structuring of the national Salzmann pine collection in order to develop a method to create a reduced core collection of this collection by preserving genetic diversity. This method effectively preserves the genetic diversity of the collection and was designed to be easily reusable with other species. It can thus be applied for core collection's constitution for other forest species within the framework of the CRGF.