



# Structure génétique des populations de l'Apollon et influence du paysage dans la région Auvergne Rhône-Alpes

**Caroline Kebaili**

Master 2 Biodiversité, Écologie et Évolution, parcours Dynamique et Modélisation de la  
biodiversité  
2018-2019

sous la direction de Laurence Després

Enseignant référent : François Munoz



*Je tiens en premier lieu à remercier Laurence Després pour sa disponibilité, sa gentillesse et toutes les opportunités qu'elle m'a offertes tout au long de mon master. J'ai appris énormément et rencontré beaucoup de personnes d'un grand intérêt. Et confirmé mon envie de poursuivre dans la recherche. Merci beaucoup !*

*Merci aux équipes des parcs et de FlaviaAPE pour la récolte des échantillons, à la poursuite d'un Apollon pas très facile à dénicher, et pour leurs éclairages plus que bienvenus sur le terrain et le comportement de ce beau papillon. Votre aide m'a été précieuse.*

*Merci également aux autres membres du LECA, Stéphanie, Christian, Florence et Olivier pour leur accueil, leur disponibilité et leurs conseils, ainsi qu'à mes collègues, Elizabeth, Laura, Marie, Théophile et les deux Camilles pour les échanges nombreux et intéressants.*

*Et enfin, merci à Vincent et à mes parents, qui me supportent dans mes moments de doute et de fatigue et m'encouragent tous les jours.*

# TABLE DES MATIÈRES

INTRODUCTION.....	1
MATÉRIEL ET MÉTHODES.....	2
<i>Échantillonnage.....</i>	<i>2</i>
<i>Extraction de l'ADN et séquençage.....</i>	<i>2</i>
<i>Tri des données génétiques et construction des catalogues de SNPs.....</i>	<i>3</i>
<i>Analyse descriptive des données génétiques.....</i>	<i>3</i>
<i>Étude de la structure génétique des populations.....</i>	<i>3</i>
<i>Étude du paysage et de l'environnement climatique.....</i>	<i>4</i>
<i>Modélisation de l'influence du paysage et de la distance géographique sur la variabilité génétique.....</i>	<i>5</i>
<i>Inférences démographiques et des temps de divergence entre populations.....</i>	<i>7</i>
RÉSULTATS.....	9
<i>Échantillonnage et tri des données.....</i>	<i>9</i>
<i>Analyse descriptive des données génétiques.....</i>	<i>10</i>
<i>Structure génétique des populations.....</i>	<i>11</i>
<i>Modélisation de l'influence du paysage et de la distance géographique sur la variabilité génétique.....</i>	<i>14</i>
<i>Inférences démographiques et des temps de divergence entre populations.....</i>	<i>17</i>
DISCUSSION.....	18
CONCLUSION.....	22
RÉFÉRENCES BIBLIOGRAPHIQUES.....	23
ANNEXES.....	26

# LISTE DES FIGURES ET CARTES

<b>Figure 1</b> : L'Apollon (F. Michalke).....	1
<b>Figure 2</b> : Raster binaire de forêts de conifères.....	5
<b>Figure 4</b> : Scenarii alternatifs de divergence entre les populations d'Auvergne et des Alpes..	8
<b>Figure 5</b> : Scenarii alternatifs de divergence entre les populations des Alpes.....	9
<b>Figure 6</b> : Hétérozygotie observée ( $H_o$ ), hétérozygotie attendue ( $H_e$ ) et coefficient de consanguinité ( $F_{is}$ ) au sein de chacun des massifs échantillonnés.....	11
<b>Carte 1</b> : Zone d'étude et répartition des sites d'échantillonnage dans la région Auvergne Rhône-Alpes.....	12
<b>Figure 7</b> : Représentation graphique de la DAPC pour les individus alpins, pour $K = 3$ .....	13
<b>Figure 8</b> : Part de la variance génétique expliquée par les catégories de modèles d'isolement par la résistance et les modèles basés sur les catégories d'occupation du sol, dans les Alpes...	15
<b>Figure 9</b> : Part de la variance génétique expliquée par les catégories de modèles d'isolement par la résistance et les modèles basés sur les catégories d'exposition, en Chartreuse.....	16
<b>Figure 10</b> : Résultats des inférences démographiques par ABC pour le meilleur scénario de divergence des populations auvergnate et alpine.....	18
<b>Figure 11</b> : Résultats des inférences démographiques par ABC pour le meilleur scénario de divergence des populations alpines.....	18

## LISTE DES TABLEAUX

<b>Tableau 1 :</b> Effets hypothétiques des catégories d'occupation du sol sur la dispersion de l'Apollon.....	6
<b>Tableau 2 :</b> Catégorisation des expositions des sites au soleil.....	6
<b>Tableau 3 :</b> Nombre de stations et d'individus échantillonnés par massif.....	9
<b>Tableau 4 :</b> Indices de diversité génétique, mesurés par station.....	10
<b>Tableau 5 :</b> Effet sur la dispersion et valeur de résistance pour les variables paysagères univariées.....	14

## LISTE DES ANNEXES

<b>Annexe I :</b> Coordonnées géographiques des stations d'échantillonnage.....	26
<b>Annexe II :</b> Tableau des indices de différenciation ( $F_{ST}$ ) inter-sites.....	27
<b>Annexe III :</b> Cartes des sites et des sous-populations dans les massifs.....	29

L'anthropocène est la période des temps géologiques caractérisée par l'impact majeur de l'Homme sur son environnement. La communauté scientifique parle de la sixième extinction de masse dont l'Homme est la cause principale (Barnosky *et al.*, 2011). Les changements climatique et d'utilisation des terres en sont des facteurs majeurs comme sources de réduction des aires de répartition des espèces et de fragmentation des habitats (Ashton, Gutiérrez et Wilson, 2009). Face à l'érosion de la biodiversité, des mesures de protection et de conservation des espèces sont mises en place, à l'image de la Convention sur le Commerce International des Espèces de faune et de flore Sauvages menacées d'extinction (CITES), signée à Washington en 1973.



*Figure 1 : L'Apollon (F. Michalke)*

Parmi les espèces concernées par la CITES se trouve l'Apollon (*Parnassius apollo*), un rhopalocère de la famille de Papilionidés, emblématique des prairies de montagnes. À l'échelle mondiale, il vit dans la majeure partie de l'Eurasie et est classé comme vulnérable par l'Union Internationale pour la Conservation de la Nature (IUCN).

En France, des populations subsistent dans les Pyrénées, dans le Haut Jura, dans les Alpes et dans le massif Central, à des altitudes le plus souvent comprises entre 800 et 2300 m (Deschamps-Cottin, 1998). Dans cette dernière zone, il est en danger d'extinction. Très héliophile, il affectionne les milieux ouverts, rocailleux, avec une végétation de faible taille (Deschamps-Cottin, 1998 ; Lafranchis, 2015). Une nouvelle génération est produite chaque année et ses chenilles, qui se nourrissent essentiellement sur des espèces des genres *Sedum* et *Sempervivum*, passent l'hiver dans l'œuf, protégées du gel par la neige (Ashton, Gutiérrez et Wilson, 2009 ; Deschamps-Cottin, 1998 ; Nakonieczny, Kędziorski et Michalczyk, 2007). En outre, sa capacité de dispersion est de l'ordre de 300 m en moyenne, mais peut dépasser le kilomètre (Brommer et Fred, 1999).

Le premier objectif de cette étude est d'étudier la structure génétique des populations de l'Apollon dans la région Auvergne Rhône-Alpes et d'identifier les éléments environnementaux et paysagers qui influencent les flux de gènes, à partir de marqueurs génétiques obtenus par double-digestion RAD Sequencing. Du fait de la distance géographique, une très forte différenciation génétique est attendue entre les populations des Alpes et de l'Auvergne. En outre, *P. apollo* étant un papillon de montagne exclusivement et les massifs échantillonnés étant

séparés par des vallées profondes (moins de 800 m d'altitude), nous avons posé l'hypothèse d'une différenciation inter-massifs forte, et d'une différenciation intra-massifs modérée, associée à la faible capacité de dispersion de l'Apollon et à son caractère sédentaire. De plus, une faible diversité génétique intra-population est supposée en lien avec son statut de protection, en particulier en Auvergne.

Le second objectif est d'évaluer les effectifs actuels et passés et les temps de divergence des populations, notamment entre le Massif Central et les Alpes, à l'aide d'une méthode Bayésienne de calcul approché. *P. apollo* est classé en danger d'extinction en Auvergne. Nous supposons donc que les effectifs auvergnats actuels sont plus faibles que les effectifs alpins. De plus, nous avons posé l'hypothèse que le réchauffement climatique qui a suivi le dernier maximum glaciaire (~ -20 000 ans BP) a causé la divergence entre les populations alpine et auvergnate. Enfin, du fait du déclin global de l'espèce, nous nous attendons à retrouver des signaux de goulots d'étranglement dans chacune des populations, dont les causes peuvent être climatique et/ou anthropique.

## **Matériel et méthodes**

### *Échantillonnage*

L'échantillonnage a été effectué lors de la période de vol des imagos de l'été 2018, dans les massifs alpins de la Chartreuse, du Vercors, des Bauges, de Belledonne, et du Puy du Sancy en Auvergne, avec l'aide des équipes de l'association FlaviaAPE, de la Réserve Naturelle des Hauts Plateaux du Vercors, de la Réserve Naturelle des Hauts de Chartreuse, du Parc Naturel Régional du Massif des Bauges et de la Réserve Naturelle Nationale de Chastreix-Sancy. Pour chaque massif, 5 individus par site devaient être capturés sur 5 sites différents. Pour chaque individu, une patte a été prélevée pour les analyses génétiques. Marschalek, Jesu et Berres (2013) ont montré qu'un tel prélèvement n'avait pas d'effet sur la survie et le comportement de *Lycaena hermes*, un lépidoptère de la famille de Lycaenidés. Les coordonnées géographiques, les noms et le code associé des sites se trouvent dans l'annexe I. Le sexe des individus n'a pas été déterminé lors de la capture.

### *Extraction de l'ADN et séquençage*

Des marqueurs Single Nucleotide Polymorphism (SNP) ont été obtenus par ddRAD

Sequencing. Après extraction, l'ADN a subi une double-digestion avec les enzymes de restriction *SbfI* et *MspI* et la ligation d'adaptateurs afin de permettre l'identification des échantillons après séquençage. Les fragments ont été sélectionnés en fonction de leur taille sur Pipin® avant d'être amplifiés par PCR. Les bibliothèques produites ont ensuite été purifiées et séquencées par la technique Illumina.

### *Tri des données génétiques et construction des catalogues de SNPs*

Les reads d'ADN obtenus ont été traités à l'aide du programme STACKS (Catchen *et al.*, 2011) afin d'obtenir les SNPs sur lesquels ont été faits les analyses. Après une étape de filtration pour obtenir les reads ayant une longueur minimale de 110 paires de bases (bp), ces derniers ont été démultiplexés et filtrés en fonction de leur qualité à l'aide de la fonction *process radtags*. Les reads de qualité insuffisante, c'est-à-dire avec une probabilité supérieure à 0,1 d'être incorrects du fait d'un trop grand nombre d'erreurs de séquençage, ont été éliminés. Les séquences identiques ont ensuite été assemblées *de novo* sous forme de stacks (fonction *ustacks*), avec une tolérance d'au maximum 7 nucléotides différents. Cette étape permet d'identifier des loci putatifs et leurs allèles associés. Puis, par le biais de la fonction *cstacks*, un catalogue des loci a été construit, en identifiant les allèles consensus pour chacun des loci. Enfin, la fonction *populations* a été utilisée pour créer un jeu de données SNPs à partir du génotype de chaque individu à chacun des loci polymorphes. Seuls les SNPs présents chez au moins 80% des individus de la population et avec une fréquence allélique minimale de 5% ont été inclus. Un seul SNP par locus a été retenu.

### *Analyse descriptive des données génétiques*

Les hétérozygoties attendue ( $H_e$ ) et observée ( $H_o$ ), mesures de la diversité génétique, ainsi que le coefficient de consanguinité  $F_{is}$  (mesure de l'écart à la panmixie, supposée par l'équilibre de Hardy-Weinberg) et les fréquences alléliques à chaque locus, ont été calculés sous R (packages *hierfstat* (Goudet, 2004) et *adegenet* (Jombart, 2008)). 1000 bootstraps ont été effectués pour obtenir les intervalles de confiance à 95% des  $H_e$ ,  $H_o$  et  $F_{is}$ .

### *Étude de la structure génétique des populations*

La structure de la population a ensuite été étudiée à l'aide de plusieurs méthodes. D'une part,



le programme *sNMF* (Frichot *et al.*, 2014), qui estime les coefficients d'admixture individuels, et implémenté dans le package R *LEA*, permet d'inférer le nombre  $K$  de sous-populations génétiquement distinctes. Ce nombre est celui qui minimise la valeur du critère d'entropie croisée. Les valeurs de  $K = 1$  à  $K = 25$ , correspondant au nombre de sites échantillonnés, ont été testés, avec 20 répétitions pour chaque  $K$ .

De plus, une analyse discriminante en composantes principales (DAPC) a été réalisée. Cette méthode décompose la variance de l'échantillon en une variance intra-groupe et une variance inter-groupe et maximise cette dernière, ce qui lui permet d'estimer un nombre  $K$  de clusters génétiquement distincts. Des valeurs croissantes de  $K$  sont comparées en utilisant le Critère d'Information Bayésien (BIC), le  $K$  optimal étant celui qui minimise ce critère.

L'indice de différenciation génétique  $F_{ST}$  a été calculé pour chaque paire de localités à l'aide du package R *adegenet*, avec 1000 bootstraps. La significativité des résultats a été vérifiée par le biais de leur intervalle de confiance à 95% : les valeurs de  $F_{ST}$  dont l'intervalle de confiance inclut 0 ne sont pas significatives.

### *Étude du paysage et de l'environnement climatique*

Afin d'appréhender le paysage dans lequel l'Apollon évolue, différentes variables environnementales ont été considérées. L'occupation du sol a été déterminée à l'aide des données Corine Land Cover 2018 (Copernicus, 2018 ; résolution de 100 m x 100 m) et Cesbio 2018 (Inglada *et al.*, 2018 ; résolution de 10 m x 10 m). Ces données sont complémentaires : Corine Land Cover comporte des catégories plus précises et Cesbio une meilleure résolution. Un modèle numérique de terrain (résolution de 25 m x 25 m) a été construit grâce à la base de données ALTI ® de l'Institut national de l'information géographique et forestière (IGN). Ce modèle a permis d'obtenir des valeurs d'altitude, de pente et d'exposition pour les sites échantillonnés.

L'environnement physique et climatique a été étudié par le biais des 19 variables bioclimatiques, de la vitesse du vent et de la quantité de radiations solaires reçue au niveau des sites d'échantillonnage, issues de *WorldClim* (<http://worldclim.org>).

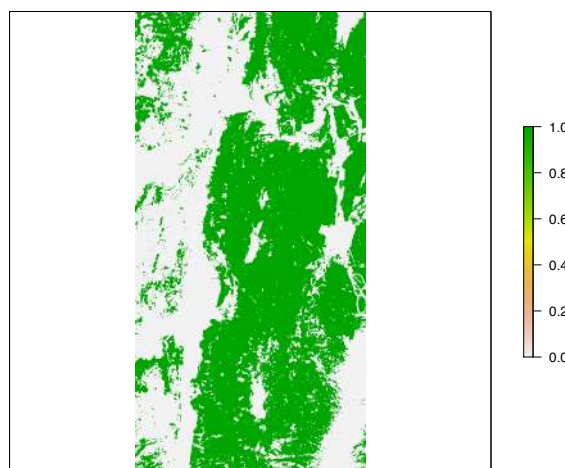
Afin d'étudier l'influence de la répartition des différentes espèces utilisées par la chenille, des données d'occurrences de ces plantes, provenant du Conservatoire Botanique National Alpin (CBNA) et du site *Global Biodiversity Information Facility* (GBIF) (<https://www.gbif.org>), ont

été rastérisées sous forme de données de présence. Les espèces suivantes ont été prises en compte (Deschamps-Cottin, Roux et Descimon, 1997 ; Deschamps-Cottin, 1998) : *Sedum album*, *Sedum sediforme* (= *altissimum* & *nicaeense*), *Sedum ochroleucum*, *Sedum sexangulare*, *Sedum telephium*, *Sempervivum montanum* et *Sempervivum arachnoideum*.

L'ensemble de ces données ont été manipulées à l'aide des logiciels ArcGIS 10.6 (Esri, 2018) et R (R Development Core Team, 2005), et découpées à l'échelle des Alpes (départements de la Drôme, de l'Isère, de la Savoie et de la Haute-Savoie) et des différents massifs échantillonnés (les Bauges, la Chaîne de Belledonne, la Chartreuse et le Vercors pour les Alpes, et du Puy de Sancy en Auvergne).

### *Modélisation de l'influence du paysage et de la distance géographique sur la variabilité génétique*

Connaissant l'habitat de l'Apollon et à partir des hypothèses concernant les éléments du paysage pouvant constituer une barrière à sa dispersion ou au contraire la facilitant, des rasters de résistance univariés de différentes catégories d'occupation du sol ont été construits (Tab.1), en attribuant la valeur 1 aux pixels correspondant à la catégorie considérée, et la valeur 0 aux autres pixels du raster (Fig.2).



*Figure 2 : Raster binaire de forêts de conifères (1 : présence de conifères, 0 : reste de la matrice)*

Par ailleurs, l'effet des lisières de forêts a été testé en déterminant les contours de forêts dans *ArcGIS*.

En supposant que la forêt est une surface très résistante à la dispersion, la valeur 1 a été attribuée aux pixels correspondant à la lisière, la valeur 10 000 aux pixels de forêt et une valeur 0 aux autres pixels.

En ce qui concerne l'exposition, les données ont été catégorisées en fonction du moment de la journée où l'ensoleillement est maximal et de l'exposition au Nord (Tab.2).

*Tableau 1 : Effets hypothétiques des catégories d'occupation du sol sur la dispersion de l'Apollon*

<b>Catégories d'occupation du sol</b>	<b>Effet hypothétique sur la dispersion</b>
Forêts de feuillus	Résistance
Forêts de conifères	Résistance
Landes arbustives	Facilitation
Zones anthropisées	Résistance
Prairies naturelles	Facilitation
Végétation éparse	Facilitation

*Tableau 2 : Catégorisation des expositions des sites au soleil*

<b>Exposition</b>	<b>Catégorie</b>	<b>Effet hypothétique sur la dispersion</b>
Nord ; Nord-Est ; Nord-Ouest	Exposition au Nord	Résistance
Est ; Sud-Est	Exposition le matin	Facilitation
Sud ; Sud-Ouest	Exposition toute la journée	Facilitation
Ouest	Exposition l'après-midi	Facilitation

Un raster binaire (1 : pixels correspondant à l'exposition ; 0 : autres pixels) pour chaque type d'exposition a été créé. À l'échelle d'un massif, un grain de 100 m x 100 m a été choisi, tandis qu'à l'échelle des Alpes, un grain de 300 m x 300 m a été préféré pour limiter le temps de calcul. Ces valeurs de grain ont été sélectionnées en se basant sur la capacité moyenne de dispersion de l'Apollon, évaluée à environ 300m (Brommer et Fred, 1999).

Afin de déterminer le caractère résistant ou facilitant des différentes surfaces, et d'optimiser la valeur de résistance à lui appliquer, le coefficient de corrélation entre la matrice de distances génétiques normalisées ( $F_{ST}/(1 - F_{ST})$ ) et la matrice de distances paysagères de moindre coût, a été déterminé à l'aide d'un test de Mantel (Epps *et al.*, 2007), en appliquant une gamme de valeurs de résistance discrètes (2, 3, 5, 10, 50, 250, 500, 1000) aux pixels de valeur 1, pour tester l'effet résistant de la variable, puis aux pixels de valeur 0 pour tester l'effet facilitant. Les distances de moindre coût correspondent à la distance que l'organisme doit parcourir au sein

de la matrice paysagère pour relier les différents sites d'échantillonnage, en minimisant le coût de la dispersion en valeur de résistance.

La valeur de résistance pour laquelle la corrélation entre la matrice de distances paysagères et la matrice de distances génétiques est maximale est alors la valeur optimale.

Après optimisation, la part de la variance génétique expliquée par le modèle d'isolement par la résistance correspondant a été évaluée à l'aide d'une régression multiple sur les matrices de distances, avec 1000 permutations.

Par ailleurs, cette analyse a également été réalisée pour tester des modèles d'isolement par la distance, afin de déterminer si la distance génétique entre populations est mieux expliquée par des variables paysagères que par les distances euclidiennes entre les sites d'échantillonnage.

Une analyse de redondance (RDA, package R *vegan* (Oksanen *et al.*, 2007)) a été effectuée pour tester l'effet des variables bioclimatiques, physiques (vitesse du vent, quantité de radiations solaires reçue) et topologiques (pente) sur les fréquences alléliques, calculées avec le package R *adegenet*. Préalablement, la corrélation entre les variables a été contrôlée. De plus, la variable altitude étant corrélée à de nombreuses variables bioclimatiques et afin de tester l'influence de la niche altitudinale de l'Apollon dans la région (entre 800 et 2300 m), l'effet de la présence de vallées profondes (altitude inférieure à 800 m) sur la variabilité génétique a été évalué selon la méthode employé par Rogic *et al.* (2013) : considérant les vallées comme des barrières à la dispersion, les sites situés d'un même côté de la barrière se voit attribuer la valeur 1 et les sites situés de l'autre côté la valeur 0. Une RDA a été réalisée avec cette variable catégorielle entre les massifs de la Chartreuse, du Vercors et des Bauges. L'effet de la présence de montagnes n'a pas été testé car aucun couple de stations n'était séparé par un sommet de plus de 2300 m d'altitude.

### *Inférences démographiques et des temps de divergence entre populations*

Afin d'évaluer les tailles des populations des différents massifs échantillonnés et le temps de leur divergence, et les changements de taille de ces populations au cours du temps, une méthode de calcul Bayésien approché (ABC) de recherche du coalescent, implémentée dans le programme DiyABC (Cornuet *et al.*, 2014) a été employée. Notre hypothèse est que la divergence entre Auvergne et Alpes est liée au réchauffement climatique post-glaciaire qui a fragmenté les populations entre les deux massifs montagneux, et qui a pu également être à l'origine d'une baisse de taille des populations (« bottleneck ») ; l'anthropisation des milieux est une autre cause de déclin des populations, plus récente (Todisco *et al.*, 2010). Nous avons

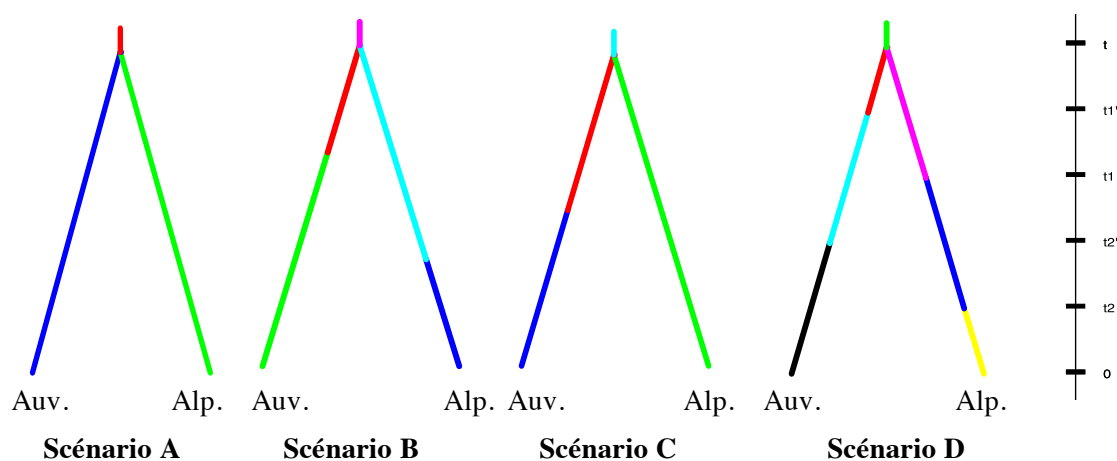
donc testé 3 hypothèses : la divergence a été suivie d'un bottleneck uniquement en Auvergne (où les populations sont actuellement le plus en danger) ; la divergence a été suivie d'un bottleneck dans chacune des deux lignées ; ou de deux bottlenecks successifs, climatique, puis anthropique.

Dans un premier temps, quatre scénarii évolutifs de divergence (suivie ou non de goulots d'étranglement) des populations d'Auvergne et des Alpes ont été comparés (Fig. 4). Les Alpes sont représentées par la population échantillonnée dans le Vercors.

Les gammes de variations des paramètres du modèle ont été définies a priori de la manière suivante :

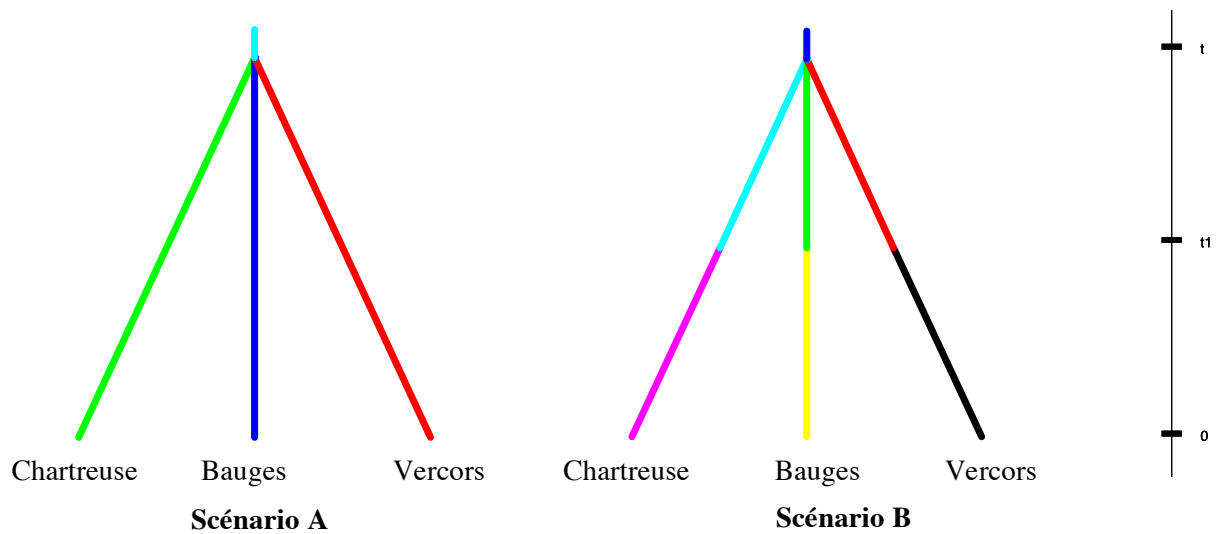
- le paramètre de taille des populations actuelles est compris entre 100 et 10 000 individus, en se basant sur le statut de protection de *Parnassius apollo*, qui suppose des populations de faible effectif ;
- en posant l'hypothèse d'une divergence des populations due au réchauffement climatique qui a suivi le dernier maximum glaciaire (~ -20 000 ans BP), le temps de divergence se situe entre 100 et 30 000 ans ;
- pour les scénarii inférant un goulot d'étranglement, la gamme de temps a été fixée entre 0 et 30 000 ans, afin de ne pas exclure une cause climatique ou une cause anthropique.

Pour chaque run, 200 000 simulations ont été effectuées. Selon leurs résultats, les paramètres a priori ont été ajustés afin d'améliorer la qualité du modèle.



*Figure 4 : Scénarii alternatifs de divergence entre les populations d'Auvergne et des Alpes (A : pas de bottleneck ; B : un bottleneck pour chaque population ; C : un bottleneck pour la population d'Auvergne ; D : deux bottlenecks pour chaque population)*

Ensuite, deux scénarii ont été testés pour la divergence des populations alpines, avec ou sans bottleneck. Les gammes de variations des paramètres du modèle ont été définies a priori en se basant sur les résultats des simulations de la divergence entre les populations auvergnates et



alpines.

*Figure 5 : Scénarii alternatifs de divergence entre les populations des Alpes  
(A : pas de bottleneck ; B : un bottleneck pour chaque population)*

## Résultats

### *Échantillonnage et tri des données*

L'ADN d'un total de 100 individus a été extrait et envoyé au séquençage, avec l'obtention de plus de 70 millions de reads de bonne qualité (environ 750 000 reads par individus). Le tableau 3 et la carte 1 montrent la répartition des échantillons dans les différents massifs.

*Tableau 3 : Nombre de stations et d'individus échantillonnés par massif*

Massif	Nombre de stations	Nombre d'individus
Puy du Sancy	2	10
Chartreuse	9	37
Bauges	6	21
Vercors	5	25
Belledonne	3	7

Quatre individus ont été exclus des analyses, à raison d'un individu par massif (sauf Belledonne), car ils ne présentaient moins de 100 000 reads après séquençage. L'algorithme

STACKS a permis d'obtenir un jeu de 2191 loci polymorphes pour l'ensemble des 96 individus, 2210 pour les 87 individus alpins et 1799 pour les 9 individus auvergnats.

Étant donné le faible nombre d'individus et de stations disponibles dans le massif de Belledonne, les analyses suivantes se sont concentrées sur la Chartreuse, le Vercors et les Bauges, dans les Alpes.

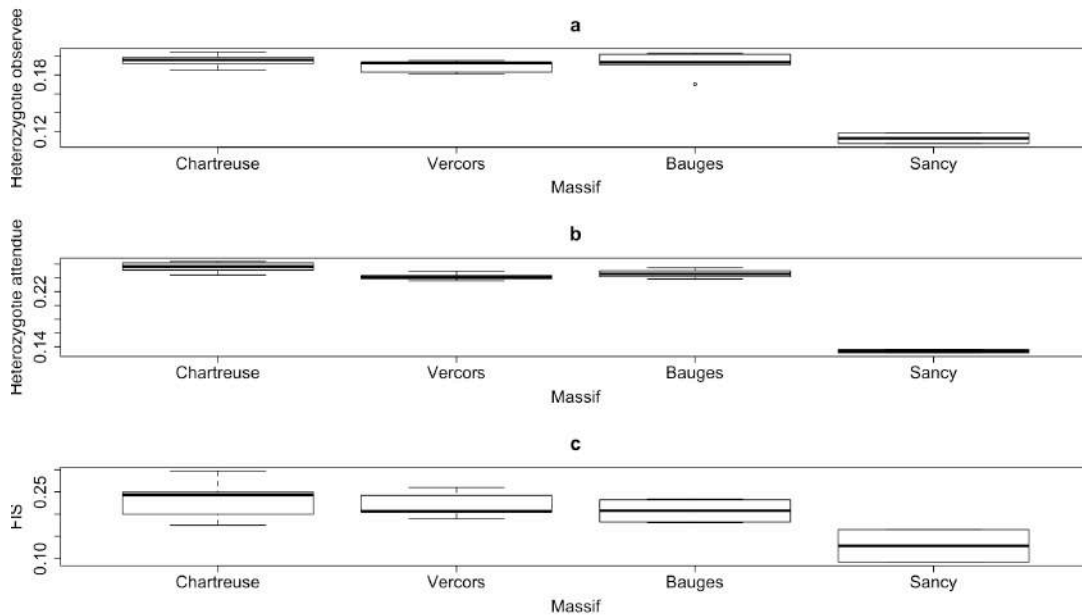
### *Analyse descriptive des données génétiques*

L'hétérozygotie attendue est comprise entre 0,1312 dans le Sancy et 0,2605 en Chartreuse, et par ailleurs significativement plus faible dans le Sancy que dans les autres massifs (ANOVA,  $F_{3,16} = 213,00$ ,  $P < 0,01$ ). En Chartreuse, l'hétérozygotie est plus élevée que dans le Vercors ( $P < 0,01$ ), tandis que le massif des Bauges ne présente pas de différences significatives avec les autres massifs.

*Tableau 4 : Indices de diversité génétique, mesurés par station*

Massif	Code site	Ho	Ho IC 95%	He	He IC 95%	Fis	Fis IC 95%
Chartreuse	CQ	0.1930	0.1827-0.2038	0.2605	0.2487-0.2719	0.2466	0.2204-0.2954
Chartreuse	LS	0.1959	0.1878-0.2053	0.2558	0.2467-0.2645	0.2444	0.2049-0.2613
Chartreuse	VP	0.1850	0.1734-0.1972	0.2632	0.2479-0.2781	0.2974	0.2519-0.3419
Chartreuse	RF	0.1991	0.1897-0.2086	0.2637	0.2531-0.2738	0.2524	0.214-0.2759
Chartreuse	CS	0.1987	0.1895-0.2076	0.2572	0.247-0.2662	0.2425	0.1984-0.2553
Chartreuse	PS	0.1894	0.1724-0.2062	NA	NA	NA	NA
Chartreuse	LR	0.2041	0.194-0.214	0.2524	0.2423-0.2626	0.1750	0.1591-0.2242
Chartreuse	GR	0.1922	0.1831-0.2014	0.2441	0.2351-0.2527	0.1954	0.1862-0.24
Chartreuse	OU	0.1962	0.187-0.2055	0.2504	0.2406-0.259	0.2042	0.1863-0.245
Vercors	CO	0.1831	0.1737-0.1932	0.2358	0.2267-0.2461	0.2421	0.1892-0.2545
Vercors	PG	0.1954	0.186-0.2046	0.2434	0.2342-0.2526	0.1899	0.1653-0.2253
Vercors	MS	0.1928	0.1836-0.2028	0.2407	0.2307-0.2515	0.2047	0.1662-0.2302
Vercors	TU	0.1933	0.1843-0.2026	0.2497	0.2393-0.2589	0.2067	0.1958-0.2546
Vercors	LC	0.1809	0.1718-0.1895	0.2388	0.2296-0.2479	0.2603	0.2151-0.2714
Bauges	MO	0.1701	0.1537-0.1869	NA	NA	NA	NA
Bauges	SE	0.1923	0.1824-0.2017	0.2383	0.2294-0.248	0.1842	0.1619-0.2231
Bauges	CC	0.2031	0.1933-0.2119	0.2462	0.2372-0.2554	0.1806	0.1471-0.2035
Bauges	AR	0.1946	0.1848-0.205	0.2548	0.2455-0.2653	0.2314	0.2033-0.2683
Bauges	LG	0.2018	0.184-0.2203	NA	NA	NA	NA
Bauges	LC	0.1908	0.1808-0.1997	0.2453	0.2348-0.2551	0.2336	0.1899-0.2542
Sancy	S1	0.1186	0.1095-0.127	0.1312	0.1226-0.1398	0.0917	0.0557-0.1344
Sancy	S2	0.1074	0.0983-0.1171	0.1361	0.1265-0.1465	0.1654	0.1638-0.2587

En outre, l'hétérozygotie observée est significativement plus faible dans le Sancy que dans les autres massifs (ANOVA,  $F_{3,16} = 58,00$ ,  $P < 0,01$ ), et est comprise entre 0,1074 dans le Sancy et 0,2041 en Chartreuse, sans qu'il y ait de différences significatives entre les massifs alpins.



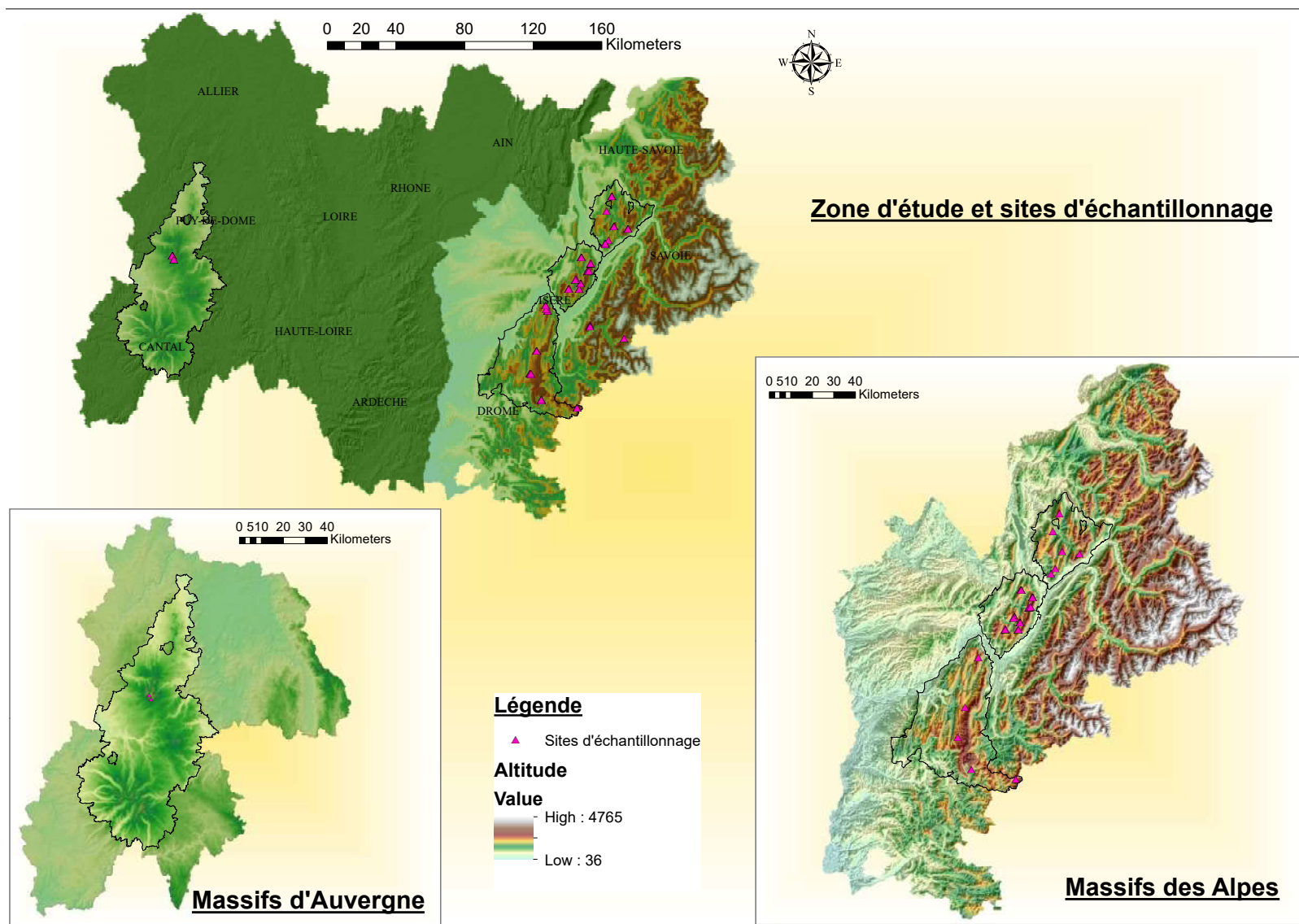
*Figure 6 : (a) Hétérozygotie observée ( $H_o$ ), (b) hétérozygotie attendue ( $H_e$ ) et (c) coefficient de consanguinité ( $F_{IS}$ ) au sein de chacun des massifs échantillonnés*

Le coefficient de consanguinité, compris entre 0,0917 dans le Sancy et 0,2974 en Chartreuse, est plus bas dans le Sancy que dans les massifs du Vercors et de la Chartreuse (ANOVA,  $F_{3,16} = 4,597$ ,  $P < 0,05$ ). Le  $F_{IS}$  n'est pas significativement différent dans les massifs alpins. Ces derniers résultats sont reportés dans le tableau 4 et représentés dans la figure 6.

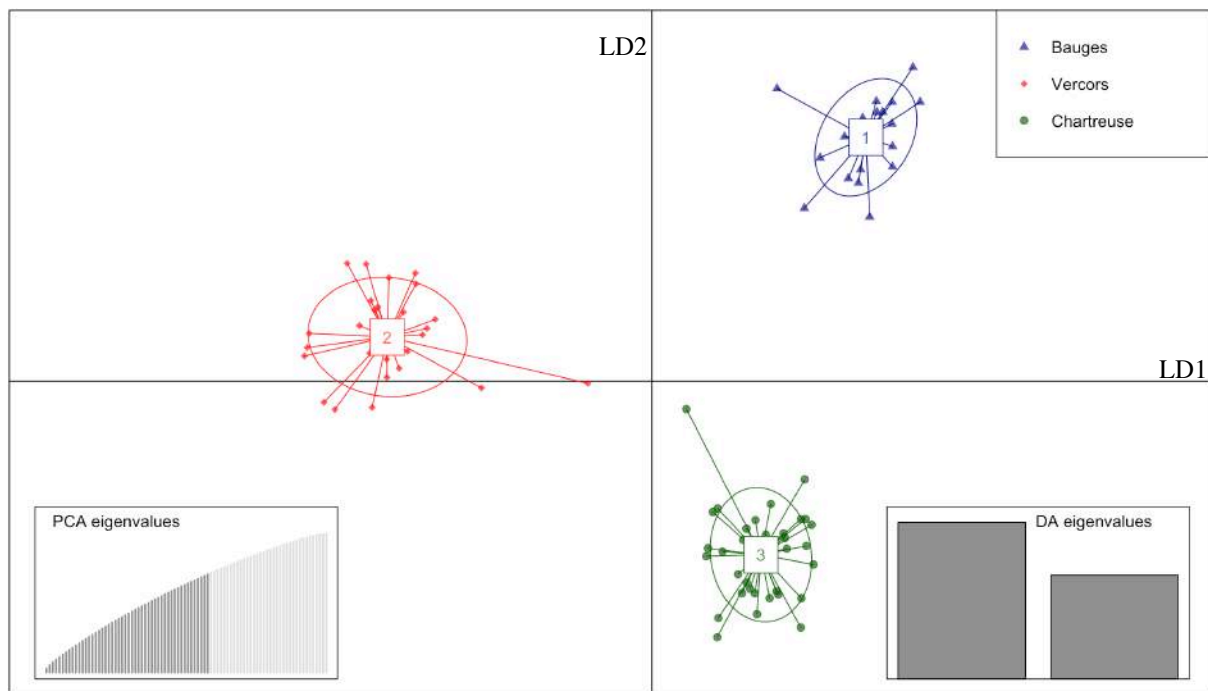
### *Structure génétique des populations*

Le critère d'entropie croisée le plus bas retrouvé par l'algorithme sNMF donne  $K = 2$  comme nombre de clusters génétiques distincts, l'un comprenant les individus échantillonnés en Auvergne et l'autre ceux prélevés dans les Alpes. Étant donnée la distance géographique séparant ces deux zones d'échantillonnage, qui suppose une différenciation génétique importante, la structure des populations alpines a été ensuite étudiée avec le jeu de données comportant 2210 SNPs (individus alpins seulement). En se basant sur le BIC, le résultat de la DAPC de ce jeu de données donne  $K = 1$  optimal, impliquant une absence de structure génétique dans la zone échantillonnée, que ce soit au niveau intra-massifs ou au niveau inter-massifs.





*Carte 1 : Zone d'étude et répartition des sites d'échantillonnage dans la région Auvergne Rhône-Alpes*



*Figure 7 : Représentation graphique de la DAPC pour les individus alpins, pour  $K = 3$*

Néanmoins, la figure 7 montre des clusters de points bien distincts lorsque l'on représente le nuage de points pour  $K = 3$ , ce qui suggère l'existence d'une structure génétique inter-massifs. À noter toutefois que les individus du massif de Belledonne sont inclus dans le cluster du Vercors. En revanche, augmenter la valeur de  $K$  pour représenter le nuage de points des individus ne permet pas de retrouver de structure intra-massifs cohérente avec la répartition géographique des échantillons. Par exemple, pour  $K = 4$ , un individu du col Cochette, situé dans les Bauges, est inclus dans le cluster regroupant tous les individus du massif de la Chartreuse, sauf un individu du site d'Outheran, lui-même inclus dans un cluster d'individus des Bauges.

Les  $F_{st}$  inter-massifs sont compris entre 0,397 entre les massifs du Puy du Sancy et de Belledonne, correspondant à une différenciation génétique très importante, et 0,011 entre les massifs du Vercors et de Belledonne. En moyenne, le  $F_{st}$  est de 0,026 entre les massifs alpins, correspondant à une différenciation faible.

Au niveau intra-massifs, les  $F_{st}$  mesurés entre les sites d'échantillonnage représentent une différenciation faible. Dans le Sancy, les deux sites d'échantillonnage montrent un indice de différenciation de 0,047. La moitié des  $F_{st}$  mesurés pour le Vercors sont non significatifs, donc considérées comme nuls. Les stations se regroupent en deux sous-populations, avec une différenciation de 0,009 (Annexes II et III). Pour la Chartreuse, 38,9% des  $F_{st}$  inter-sites sont non significatifs, impliquant un regroupement des stations en quatre sous-populations, dont les

$F_{st}$  sont compris entre 0,012 et 0,042 (Annexes II et III). Enfin, 40% des indices de différenciation mesurés dans les Bauges sont non significatifs et trois sous-populations sont distinguables, avec des  $F_{st}$  compris entre 0,011 et 0,041 (Annexes II et III).

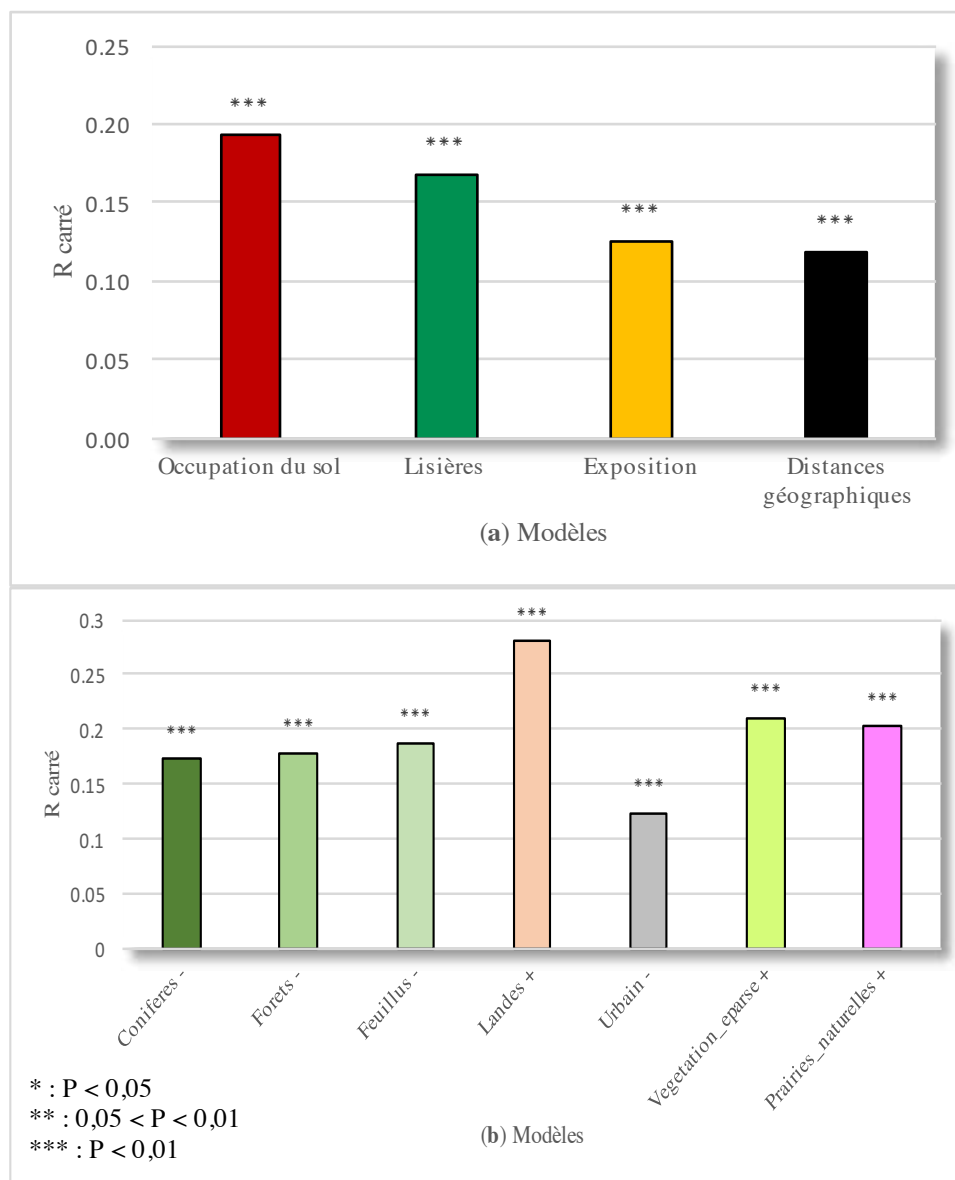
*Modélisation de l'influence du paysage et de la distance géographique sur la variabilité génétique*

Sur l'ensemble de la zone d'étude, un fort signal d'isolement par la distance est détecté ( $R^2 \sim 0,82$ ,  $P < 0,1$ ), principalement influencé par l'importante distance géographique qui sépare la zone auvergnate d'échantillonnage de la zone alpine ( $\sim 350$  km). Cependant, la part de la variance génétique expliquée par les distances euclidiennes à l'échelle des Alpes est beaucoup moins importante ( $R^2 \sim 0,12$ ,  $P < 0,01$ ). À l'échelle des massifs, les distances euclidiennes expliquent environ 17% de la variance génétique, alors qu'elles n'ont pas d'effet significatif dans le Vercors et les Bauges. Par ailleurs, le nombre de stations n'était pas suffisant dans le Puy du Sancy pour effectuer les statistiques.

*Tableau 5 : Effet sur la dispersion et valeur de résistance pour les variables paysagères univariées (les valeurs de résistance sont présentées sous la forme a : b, avec a la valeur attribuée au pixel correspondant à la variable et b, la valeur attribuée au reste de la matrice ; les effets sur la dispersion qui diffèrent d'une échelle à l'autre sont en bleu)*

Variable	Alpes		Chartreuse	
	Valeur de résistance optimale	Influence sur la dispersion	Valeur de résistance optimale	Influence sur la dispersion
Forêts de feuillus	50 :1	Résistance	5 :1	Résistance
Forêts de conifères	1000 : 1	Résistance	100 : 1	Résistance
Landes arbustives	1 : 1000	Facilitation	1 : 1000	Facilitation
Zones anthropisées	3 : 1	Résistance	10 : 1	Résistance
Prairies naturelles	1 : 1000	Facilitation	1 : 1000	Facilitation
Végétation éparse	1 : 1000	Facilitation	100 : 1	Résistance
Exposition au Nord	1 : 3	Facilitation	1 : 10	Facilitation
Exposition le matin	1 : 1000	Facilitation	100 : 1	Résistance
Exposition toute la journée	1 : 10	Facilitation	50 : 1	Résistance
Exposition l'après-midi	1 : 3	Facilitation	1 : 500	Facilitation

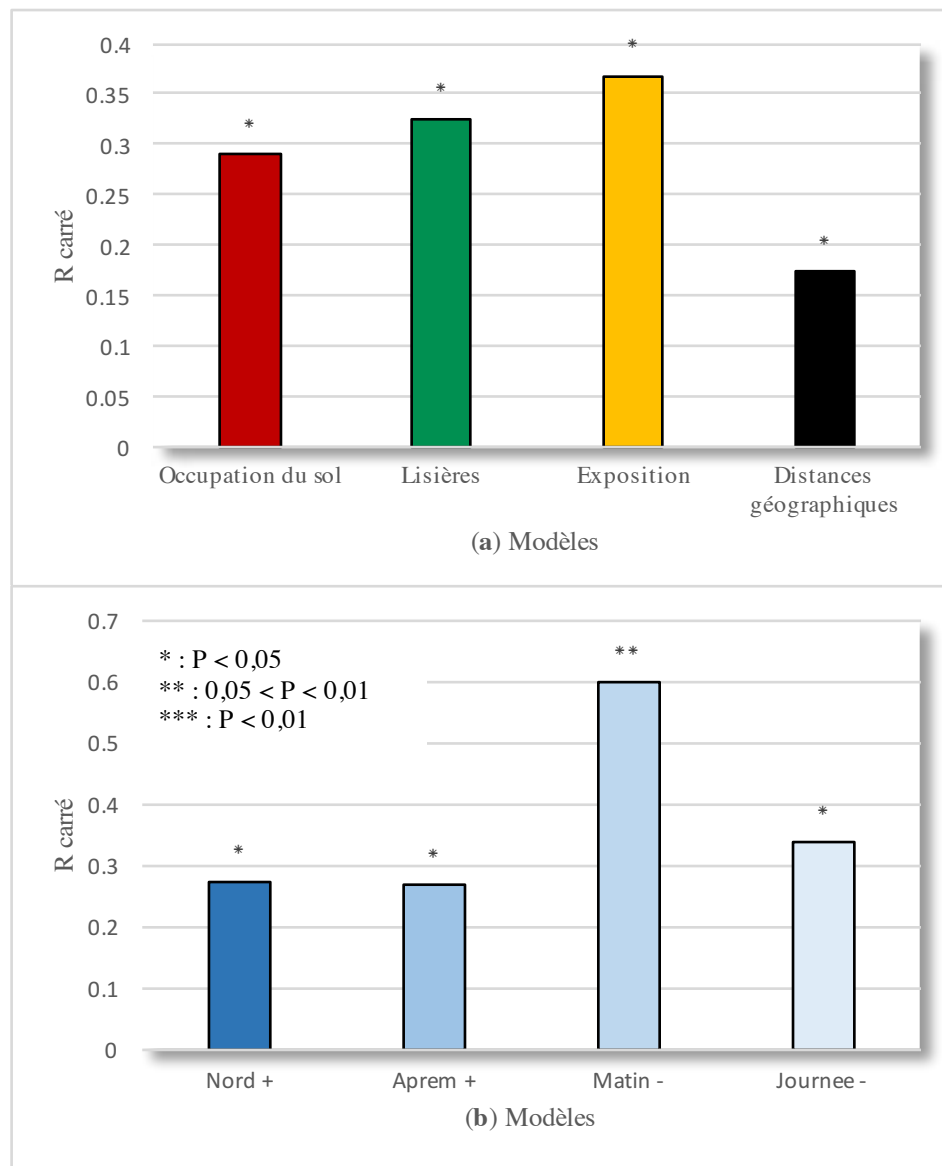
L'optimisation des valeurs de résistance à appliquer aux différentes variables considérées est présentée dans le tableau 5. L'absence de résultats significatifs dans les massifs des Bauges et du Vercors explique leur absence de ce tableau. De même, les plantes-hôtes n'ont pas pu être prises en compte. Le caractère lacunaire des données du CBNA dans la zone d'étude et l'imprécision de nombreuses occurrences de *Gbif* n'a pas permis d'obtenir un raster de présence fiable.



*Figure 8 : Part de la variance génétique expliquée par (a) les catégories de modèles d'isolement par la résistance et (b) les modèles basés sur les catégories d'occupation du sol, dans les Alpes*

Les résultats diffèrent d'une échelle à l'autre, le plus souvent en termes de valeurs de résistance mais également parfois au niveau du caractère résistant ou facilitant de la dispersion, à l'image de la variable « végétation éparse ».

En moyenne à l'échelle des Alpes, ce sont les modèles d'isolement par la résistance (IBR) basés sur les différents types d'occupation du sol qui expliquent le mieux les distances génétiques, bien que faiblement en proportion ( $R^2 \sim 0,19$ ,  $P < 0,01$ ) (Fig. 8a).



*Figure 9 : Part de la variance génétique expliquée par (a) les catégories de modèles d'isolement par la résistance et (b) les modèles basés sur les catégories d'exposition, en Chartreuse*

Parmi ces catégories d'occupation du sol, les landes arbustives, facilitant la dispersion, expliquent environ 28% de la variance ( $P < 0,01$ ), contre 12% environ pour les zones anthropisées, résistantes à la dispersion ( $P < 0,01$ ) (Fig. 8b). En outre, toutes les catégories d'exposition facilitent la dispersion ( $R^2$  moyen  $\sim 0,12$ ,  $P < 0,01$ ). En revanche, à l'échelle de la Chartreuse, le modèle d'IBR basé sur la présence de lisières de forêts de conifères est le plus explicatif ( $R^2 \sim 0,50$ ,  $P < 0,05$ ) (Fig. 9a). Suivent ensuite les modèles basés sur les quatre

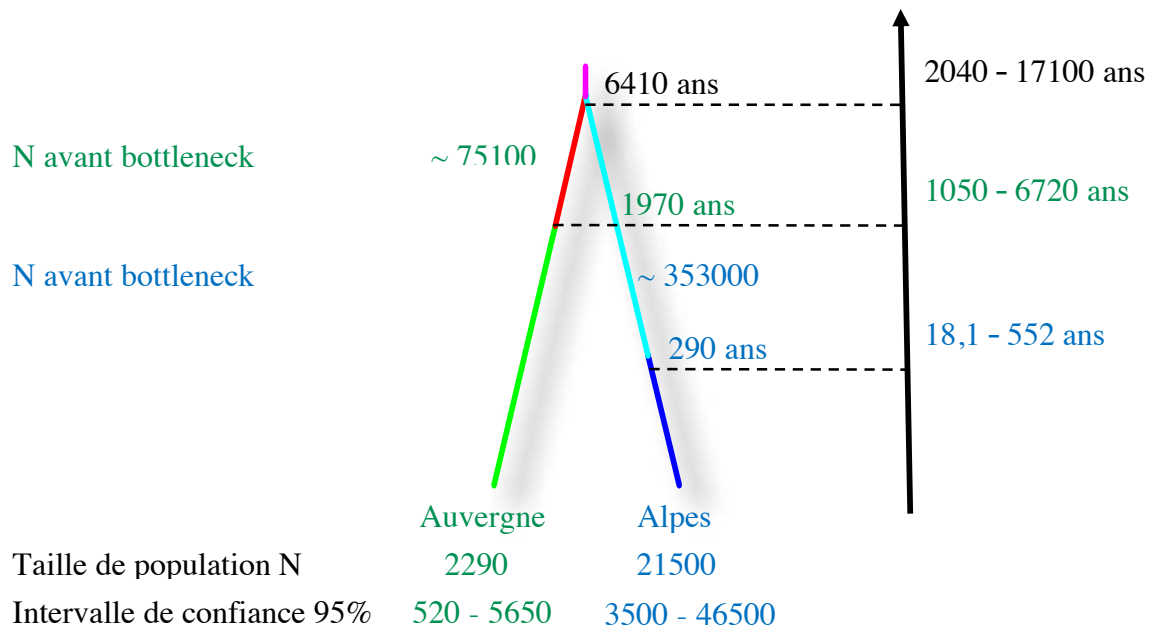
catégories d'exposition, qui expliquent environ 37% de la variance ( $P < 0,05$ ) (Fig. 9b). Toutefois, toutes les catégories ne sont pas facilitantes, comme une exposition entraînant un ensoleillement le matin et toute la journée. L'occupation du sol arrive en troisième position ( $R^2$  moyen  $\sim 0,29$ ,  $P < 0,05$ ) (Fig. 8a).

Les résultats de la RDA ne montrent pas d'influence significative des variables bioclimatiques, physiques et topologiques sur les distances génétiques. Toutefois, un effet de la présence de vallées profondes est retrouvé à l'échelle des Alpes, avec une part de variance génétique expliquée de 10,14% ( $P < 0,01$ ). En décomposant par couple de massifs, cette part passe à 2,24% ( $P < 0,05$ ) entre les Bauges et la Chartreuse, 6,67% ( $P < 0,01$ ) entre la Chartreuse et le Vercors et 3,28% ( $P < 0,05$ ) entre le Vercors et les Bauges.

#### *Inférences démographiques et des temps de divergence entre populations*

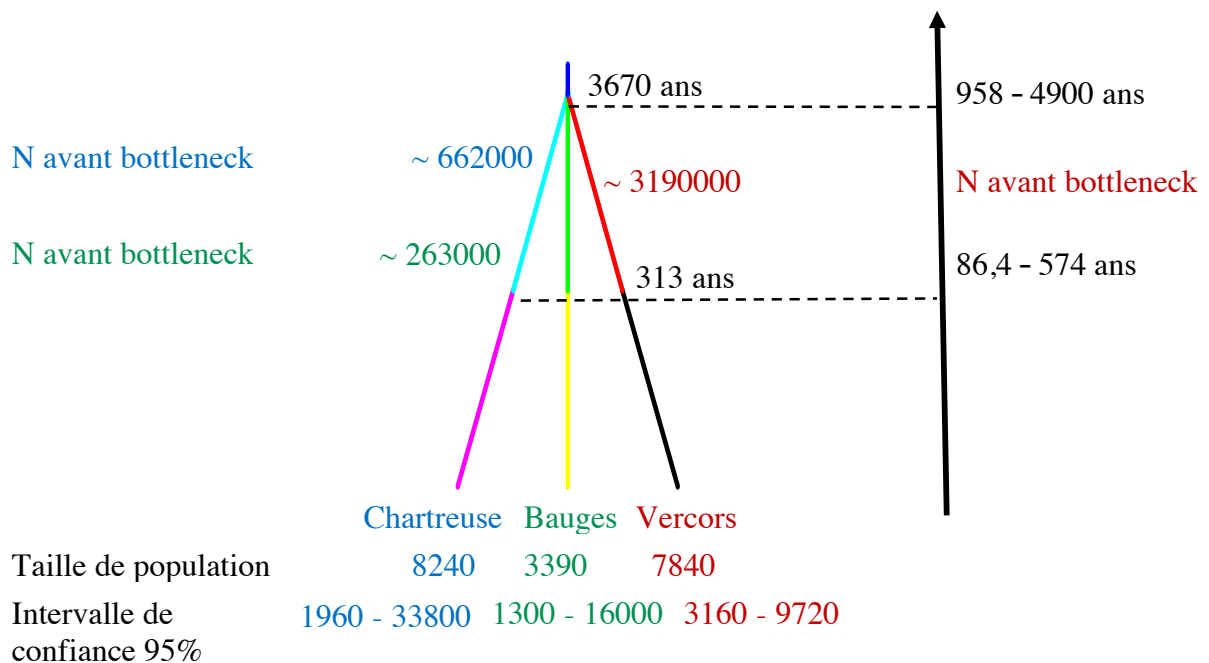
Afin d'inférer les paramètres démographiques de la population des Alpes, et son temps de divergence avec la population auvergnate, les génotypes des individus échantillonnés dans le Vercors ont été utilisés, car composant la population la plus homogène génétiquement. Le meilleur scénario (Fig. 4 ; scénario A) infère un temps de divergence moyen de 6410 ans, avec un intervalle de confiance à 95% (IC95) compris entre 2040 et 17100 années. Des signaux de forts goulots d'étranglement sont retrouvés pour les deux populations, avec une date plus ancienne en Auvergne ( $t$  moyen = 1970 ans [IC95 : 1050 – 6720 années]) que dans les Alpes ( $t$  moyen = 290 ans [IC95 : 18,1 – 552 années]). L'effectif de la population d'Auvergne a été divisé par 30. Par ailleurs, la population alpine est 10 fois plus importante que la population auvergnate (Fig. 10).

À l'échelle des Alpes, le meilleur scénario (Fig. 5 ; scénario B) infère un temps de divergence entre les populations des trois massifs de 3670 ans en moyenne [IC95 : 958 – 4900 années]. Un bottleneck récent est retrouvé dans les trois massifs ( $t$  moyen = 313 ans [IC95 : 86,4 – 574 années]).



*Figure 10 : Résultats des inférences démographiques par ABC pour le meilleur scénario de divergence des populations auvergnate et alpine*

Les effectifs actuels moyens sont du même ordre de grandeur pour les trois massifs, compris entre 3390 individus pour les Bauges et 8240 pour la Chartreuse. Les populations des massifs des Bauges et de la Chartreuse ont été divisées par 100 environ après le bottleneck, alors que la population du massif du Vercors a été divisée par 1000 environ (Fig. 11).



*Figure 11 : Résultats des inférences démographiques par ABC pour le meilleur scénario de divergence des populations alpines*

## Discussion

Les objectifs de cette étude étaient, d'une part, d'identifier la structure génétique des populations de l'Apollon en Auvergne Rhône-Alpes et les facteurs qui l'influencent, et, d'autre part, d'inférer les paramètres démographiques dans les massifs échantillonnés.

La quantité et la qualité de l'ADN obtenu lors du séquençage confirme que le prélèvement d'une seule patte est largement suffisant pour l'obtention du nombre souhaité de marqueurs génétiques et pose la question de la possibilité de réduire encore la quantité de matériel biologique prélevé sur les individus, ce qui probablement réduirait encore davantage l'impact sur la survie et la reproduction.

L'hétérozygotie observée est très faible dans le Sancy en comparaison avec les massifs des Alpes. Ce résultat est cohérent avec le statut de protection de l'espèce en Auvergne, plus préoccupant que dans les Alpes, supposant une taille de population plus faible. C'est également ce que suggère l'inférence Bayésienne, avec une taille de population 10 fois plus faible dans le Sancy que dans le Vercors. En outre, peu d'individus ont été échantillonnés dans le Sancy du fait de la difficulté de les localiser, conformément à ce que suggère le statut de protection, et bien qu'ils aient tous été prélevés dans une zone protégée, situé dans le Parc Naturel Régional des Volcans d'Auvergne, créé en 1977. Pour autant, la diversité génétique n'est pas conservée dans ce secteur. Le faible nombre de sites en Auvergne ne fournit pas d'indices sur les causes potentielles de cette faible diversité génétique car il est impossible de réaliser des statistiques robustes sur l'influence du paysage avec un nombre de stations aussi faible. Néanmoins, bien que les deux points soient séparés de seulement 2 km environ, les individus des sites constituent bien deux sous-populations distinctes, suggérant que la connectivité paysagère n'est pas idéale. Un complément d'échantillonnage pourrait permettre d'analyser plus finement la structure des populations en Auvergne. Cependant, les inférences démographiques confirment les faibles effectifs dont parlait déjà Henri Descimon en 1994. Il notait également que les quelques populations subsistantes se trouvaient très près des crêtes, ce qui rend leur capture difficile. Un échantillonnage et séquençage sur plusieurs saisons de vol, peut donc compléter les données. Étant donné les résultats pour cette zone, il paraît crucial pour la conservation de l'Apollon en Auvergne de réaliser un diagnostic plus étendu de l'état des populations, d'identifier les facteurs qui contribuent à alimenter leur déclin et de donner des préconisations de gestion concrètes afin de tenter d'endiguer le phénomène.

La diversité génétique dans les Alpes est, en revanche, globalement bonne ( $H_o$  moyenne = 0,1926), en comparaison avec un autre rhopalocère menacé, *Coenonympha oedippus*, qui



présente une hétérozygotie observée moyenne de 0,1425 dans les Alpes (Després *et al.*, 2019) et est comparable à la diversité génétique observée chez d'autres papillons alpins non menacés (Capblancq *et al.*, 2019). Toutefois, une faible proportion des Alpes a été couverte par cette étude. L'état de la diversité génétique de l'Apollon dans les Alpes reste donc à confirmer.

D'autre part, la structure génétique a été particulièrement difficile à étudier. L'analyse des coefficients d'admixture avec sNMF ne permet pas de détecter de structure dans les Alpes, bien qu'une connectivité entre les massifs, et donc un flux de gènes, soit peu probable. En effet, bien que l'Apollon ait de bonnes capacités de vol du fait de sa taille (Brommer et Fred, 1999), il utilise peu le vent comme source de propulsion et vole généralement à ras du sol plutôt qu'en altitude (comm. pers. Yann Baillet, FlaviaAPE). L'étude de la variance génétique permet d'obtenir trois clusters distincts pour les massifs alpins de la Chartreuse, des Bauges et du Vercors mais ne parvient toutefois pas à séparer les individus du Vercors de ceux du massif de Belledonne. Seul le calcul des  $F_{ST}$  permet véritablement de différencier les massifs entre eux et d'observer une structuration intra-massifs. Ces résultats suggèrent d'une part une bonne connectivité au sein des massifs pour lesquels les analyses ont pu être faites, en particulier dans le Vercors où la configuration des hauts plateaux semble favorable à la dispersion de l'Apollon, et sont d'autre part assez étonnant étant donné le statut de protection global de l'espèce et ses caractéristiques héliophile et sédentaire.

La différenciation génétique très importante détectée entre le Sancy et les massifs alpins, du même ordre de grandeur que la différenciation génétique observée entre espèces proches d'un complexe d'espèces de papillons alpins (Capblancq *et al.* 2019), est en faveur de l'existence d'une sous-espèce distincte en Auvergne. Il pourrait s'agir de *P. apollo planeixi*, décrite auparavant dans le Sancy sur la base de différences morphologiques (Eisner, 1964). Une étude phylogénétique pourra confirmer ou infirmer cet élément. Capdeville (1978) a fait l'inventaire d'un certain nombre de sous-espèces, déterminées sur des critères morphologiques, notamment dans le Massif Central et les Alpes. Leur présence reste à confirmer par des analyses génétiques. En revanche, la faible différenciation génétique entre les massifs alpins échantillonnés tend à infirmer la présence de sous-espèces en leur sein, et notamment dans le Vercors où *P. apollo vercorcius* avait été décrite par H. Testout en 1943.

Toutefois, une partie des résultats de structure génétique peuvent être vérifiée en opérant un tri supplémentaire des marqueurs génétiques. Si certains loci présentent un  $F_{ST}$  très supérieur à la moyenne, ces derniers sont potentiellement sous sélection et ne peuvent pas être considérés comme neutres. Or, la structure génétique neutre des populations repose sur l'équilibre

mutation-dérive génétique (Veuille, 2018) et son obtention implique le retrait des loci sous l'influence d'une autre force évolutive du jeu de données. En outre, le sexage des individus lors de la capture permettrait de faire la part de l'influence de la dispersion des mâles et des femelles sur cette structuration, les mâles dispersant globalement plus souvent que les femelles, qui elles sont capables de se déplacer sur des distances importantes une fois débarrassées de leurs œufs (Capdeville, 1978 ; Deschamps-Cottin, 1998). En étudiant comment se répartissent les différents loci dans une analyse en composantes principales pour les mâles et les femelles d'un complément d'échantillonnage, il pourrait même être possible d'attribuer les individus du présent jeu de données à leurs sexes respectifs (comm. pers. Claire Mérot, University of Québec).

La différenciation entre les individus d'Auvergne et des Alpes s'expliquent très majoritairement par la distance géographique. Mais à l'échelle des Alpes seules, les modèles d'IBR sont plus explicatifs, en particulier ceux qui impliquent l'occupation du sol, et notamment la présence de landes arbustives. Quelle que soit l'échelle, cette variable facilite énormément la dispersion. En effet, ce type de végétation ne gêne pas le vol de *P. apollo* du fait de sa hauteur modérée mais peut être simplement traversé sans arrêt de nourrissage, les plantes nourricières des imagos n'étant pas présentes ou en grande quantité. À l'échelle de la Chartreuse, c'est plutôt la présence de lisières de forêts de conifères qui explique le mieux la variance. Cette variable facilite la dispersion, ce qui a déjà été documenté : Haddad et Tewksbury (2005) ont montré que ce type de corridor, constituant un habitat de faible qualité, favorise la dispersion et le flux de gènes chez le Papillon Ocellé (*Junonia coenia*) et la Fritillaire Panachée (*Euptoieta claudia*), deux papillons inféodés aux milieux ouverts. Cependant, la part de la variance expliquée par le meilleur modèle dans les Alpes n'est que de 28% environ (contre 50% environ en Chartreuse). De même, la présence de vallées profondes n'explique qu'environ 10% de la variance. Le caractère univarié des modèles limite leur puissance et la complexification de la modélisation avec des surfaces de résistance multivariées est à explorer. L'influence du grain et de l'étendue choisis est aussi à considérer puisqu'une échelle plus locale favorise une meilleure résolution et permet de mieux expliquer les distances génétiques. Par ailleurs, modéliser la présence des plantes hôtes à large échelle est complexe : l'utilisation d'un large grain biaise fortement cette variable et l'emploi de grain de faible taille augmente considérablement le temps de calcul. De plus, les données d'occurrences de ces espèces restent lacunaires et/ou peu précises. Une prise en compte de ce facteur écologique majeur des lépidoptères serait plus aisée à l'échelle locale mais nécessiterait une caractérisation précise de la végétation des localités échantillonnées. Et il pourrait aussi être très intéressant d'étudier l'adaptation aux différentes espèces de plantes-

hôtes, et surtout du point de vue de la conservation, puisque ces résultats donneraient des indications sur la faisabilité des réintroductions.

Les inférences démographiques confirment des effectifs plus faibles en Auvergne que dans les Alpes. Les dates de divergence des populations, que ce soit entre les populations auvergnate et alpine ou entre les populations alpines correspondent à la période de réchauffement qui a suivi le dernier maximum glaciaire (LGM). Les modèles de niche climatique de l'espèce et les projections à ces périodes, réalisées lors de mon stage de M1 (Kebaili, 2018), montrent des probabilités d'occurrence de plus de 0,5 de l'Apollon sur une zone continue entre les Alpes et le Massif Central au LGM, qui disparaît à l'Holocène (~6000 ans BP), au profit d'une vallée avec une faible probabilité de présence. En outre, la date du goulot d'étranglement inférée en Auvergne est ancienne et correspond également à cette période. En effet, la montée en altitude à la recherche de refuges est plus limitée dans cette zone que dans les Alpes et les populations ont tout de même réussi à se maintenir à de faibles altitudes jusqu'à récemment (Descimon, 1994). Le goulot d'étranglement inféré dans les Alpes est beaucoup plus récent, ce qui suggère une origine anthropique. L'afforestation due à l'abandon de pâturages est documentée comme étant l'une des sources de pertes d'habitat de l'Apollon (Todisco *et al.*, 2010) mais d'autres changements d'utilisation des terres ont pu jouer un rôle, au gré de l'exploitation forestière, incluant l'essartage, qui appauvrit les sols (Fourchy, 1966) et de l'exploitation agricole. L'histoire de l'utilisation des terres dans les différents massifs est à étudier pour mieux comprendre les sources de ces pertes d'effectifs, en particulier dans le Vercors, où elles ont été particulièrement importantes.

## **Conclusion**

La situation de l'Apollon semble être plutôt favorable dans les Alpes, avec une structuration génétique peu marquée. Les résultats suggèrent une bonne connectivité et la diversité génétique n'est pas aussi érodée qu'attendue. En revanche, la situation en Auvergne semble préoccupante et doit être confirmée. Les facteurs qui influencent la variabilité génétique ne sont pas complètement identifiés, nécessitant une poursuite du travail entrepris, par le biais d'un complément d'échantillonnage et d'une complexification des modèles. Cela permettra de confirmer l'« état de santé » de l'espèce dans la région Auvergne Rhône-Alpes, et de préciser les facteurs qui le déterminent, afin de pouvoir donner des préconisations de gestion utiles à la conservation de l'Apollon.

## Références bibliographiques

- Ashton, S., Gutiérrez, D., & Wilson, R.** (2009). Effects of temperature and elevation on habitat use by a rare mountain butterfly : implications for species responses to climate change. *Ecological Entomology*, (34), 437–446.
- Barnosky, A. D.** (2011). Has the Earth's sixth mass extinction already arrived? *Nature*, 471, 51–57.
- Brommer, J. E., & Fred, M. S.** (1999). Movement of the Apollo butterfly *Parnassius apollo* related to host plant and nectar plant patches. *Ecological Entomology*, 24, 125–131.
- Capblancq, T. et al.** (2019). Speciation with gene flow: Evidence from a complex of alpine butterflies (Coenonympha, Satyridae). *Ecology and Evolution*, 00, 1–14.
- Capdeville, P.** (1978). Les races géographiques de *Parnassius apollo*. *Editions Sciences Nat.*
- Catchen, J., & et al.** (2011). Stacks : building and genotyping loci de novo from short-read sequences. *G3: Genes, Genomes, Genetics*, 1, 171–182.
- Deschamps-Cottin, M.** (1998). Facteurs écologiques de la distribution de *Parnassius Apollo* L. (Lepidoptera : papilionidae), du maintien et de l'extinction des populations en France. Conséquences pour l'établissement d'une politique de conservation efficace. *Aix-Marseille I*.
- Deschamps-Cottin, M., Roux, M., & Descimon, H.** (1997). Valeur trophique des plantes nourricières et préférence de ponte chez *Parnassius apollo* L. (Lepidoptera, Papilionidae). *Comptes Rendus de l'Académie Des Sciences - Series III - Sciences de La Vie*, 320(5), 399–406.
- Descimon, H.** (1994). Les *Parnassius* français : Écologie, Génétique, Perspectives pour leur conservation. *Insectes*, (93), 2–6.
- Després, L., & et al.** (2019). Inferring the biogeography and demographic history of an endangered butterfly in Europe from multilocus markers. *Biological Journal of the Linnean Society*, 126, 95–113.
- Eisner, C.** (1964). *Parnassiana Nova : Nachträgliche betrachtungen zu der revision der subfamilie parnassiinae.*
- Epps, C. W., & et al.** (2007). Optimizing dispersal and corridor models using landscape genetics. *Journal of Applied Ecology*, 44(4), 714–724.
- Fourchy, P.** (1966). Déboisement et reboisement : les débuts de la lutte contre l'érosion au XIX<sup>ème</sup> siècle dans les Alpes Françaises. *Revue Forestière Française*, 467–487.

- Frichot, E.** (2014). Fast and Efficient Estimation of Individual Ancestry Coefficients. *Genetics*, 196, 973–983.
- Goudet, J.** (2004). hierfstat, a package for r to compute and test hierarchical F-statistics. *Molecular Ecology*, 5(1), 184–186.
- Haddad, N. M., & Tewksbury, J. J.** (2005). Low-quality habitat corridors as movement consults for two butterfly species. *Ecological Applications*, 15(1), 250–257.
- Jombart, T.** (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24, 1403–1405.
- Kebaili, C.** (2018). *Reconstitution de la phylogéographie de Parnassius apollo à partir d'un marqueur mitochondrial* (p. 19) [Rapport de stage]. Saint Martin d'Hères: Université Grenoble-Alpes.
- Lafranchis, T. et al.** (2015). *La vie des Papillons : Écologie, Biologie et Comportement des Rhopalocères en France*. Barcelone: Diatheo.
- Marschalek, D. A., Jesu, J. A., & Berres, M. E.** (2013). Impact of non-lethal genetic sampling on the survival, longevity and behaviour of the Hermes copper (*Lycaena hermes*) butterfly. *Insect Conservation and Diversity*, 6(6), 658–662.
- Nakoneczny, M., kedziorski, A., & Michalczyk, K.** (2007). Apollo Butterfly (*Parnassius apollo* L;) in Europe - its History, Decline and Perspectives of Conservation. *Functional Ecosystems Anc Communities*, 1(1), 56–79.
- Oksanen, J., & et al.** (2007). The vegan package. *Community Ecology Package*, 10, 631–637.
- Testout, H.** (1943). Études lépidoptélogiques (VII). Contributions à la connaissance des *Parnassius*. *Publications de La Société Linnéenne de Lyon*, 12(1), 6–15.
- Todisco, V. et al.** (2010). Phylogeography of *Parnassius apollo*: hints on taxonomy and conservation of a vulnerable glacial butterfly invader. *Biological Journal of the Linnean Society*, (101), 169–183.
- Veulle, M.** (2018). Polymorphisme génétique et sélection. In *Encyclopédie de l'environnement* (Université Grenoble-Alpes). Retrieved from <https://www.encyclopedie-environnement.org/vivant/polymorphisme-genetique-suivi-populations-naturelles/>

## Logiciels utilisés

**Cornuet, J.-M.** (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), 1187–1189.

**Environmental Systems Research Institute (ESRI).** (2018). *ArcGIS Release 10.6*. Redlands, CA.

**R Development Core Team.** (2005). R: A language and environment for statistical computing. (Version 1.1.447).

Massif	Station	Code	X RGF93	Y RGF93	X WGS84	Y WGS84
Chartreuse	Granier	GR	929020	6487480	5.9304951	45.4485072
Chartreuse	Outheran	OU	923650	6491050	5.8634984	45.4824057
Chartreuse	Col du Coq	CQ	922370	6472650	5.8386587	45.3172463
Chartreuse	La Scia	LS	923180	6475910	5.8504914	45.3463191
Chartreuse	Charmant Som	CS	916297.72	6472739.44	5.76124718	45.3199896
Chartreuse	Le Racapé	LR	920555.625	6478457.225	5.8181617	45.37008873
Chartreuse	Pas de la Suiffière	PS	920185.55	6477967.5	5.81321425	45.36580065
Chartreuse	Roche de Fitta	RF	927423.725	6482914.4	5.90793765	45.40795515
Chartreuse	Vallon de Pratcel	VP	927761	6483073.6	5.9123214	45.4092758
Belledonne	Col de la Botte	CB	928683.1	6451503.7	5.9092352	45.1248977
Belledonne	Col de l'Infernet	CI	928755.05	6450859.475	5.90984785	45.11907718
Belledonne	Lac Lovitel	LV	948558.1	6443909.8	6.1579581	45.0496996
Vercors	Correncon	CO	897617.4	6436641.3	5.5083103	45.0007501
Vercors	Pichet-Gerland	PG	893869	6422990.75	5.45631814	44.87962782
Vercors	Molière-Sornin	MS	903310	6461289.04	5.59077404	45.22090798
Vercors	Tussac	TU	900412.86	6407845.16	5.53205176	44.74081604
Vercors	Lus-la-croix-haute	LC	921183.9	6403191.5	5.79219	44.6926399
Auvergne	S1	S1	685273.5714	6491848.429	2.811343814	45.52595703
Auvergne	S2	S2	686165	6489884	2.82282	45.5082865
Bauges	Montagny	MO	938369.0888	6518149.47	6.0652335	45.7212838
Bauges	Semnoz	SE	941376.5683	6526593.303	6.10813942	45.79620308
Bauges	Col Cochette	CC	942656.8633	6508950.261	6.11564888	45.6369991
Bauges	Armenaz	AR	950796.8451	6507781.842	6.2194388	45.62354662
Bauges	La Galopaz	LG	939524.9302	6501027.945	6.0715218	45.5668198
Bauges	Le Crouzat	LC	937609.9849	6498913.255	6.04594465	45.54845965

*Annexe I : Coordonnées géographiques des stations d'échantillonnage*

	LV	MO	SE	CC	AR	LG	LC	CB	CI	CQ	LS	VP
LV												
MO	0.0372											
SE	0.0564	0.0459										
CC	0.0302	0.0214	0.0395									
AR	0.0097	-0.0042	0.0324	0.0157								
LG	0.0108	0.0000	0.0297	0.0069	-0.0179							
LC	0.0403	0.0456	0.0357	0.0188	0.0011	-0.0023						
CB	0.0151	0.0000	0.0862	0.0654	0.0193	0.0000	0.0688					
CI	-0.0092	0.0454	0.0620	0.0510	0.0341	0.0106	0.0493	0.0036				
CQ	0.0112	0.0181	0.0533	0.0400	0.0167	-0.0059	0.0241	0.0138	0.0344			
LS	0.0139	0.0019	0.0446	0.0284	0.0156	-0.0078	0.0220	0.0204	0.0266	-0.0085		
VP	0.0219	-0.0014	0.0677	0.0483	0.0271	-0.0214	0.0483	0.0194	0.0471	-0.0179	0.0054	
RF	0.0215	0.0265	0.0559	0.0366	0.0193	-0.0016	0.0289	0.0308	0.0297	-0.0066	0.0062	-0.0577
CS	0.0228	0.0416	0.0656	0.0511	0.0310	0.0195	0.0421	0.0374	0.0429	0.0041	0.0048	0.0321
PS	0.0258	0.0000	0.0639	0.0426	0.0072	0.0000	0.0389	0.0000	0.0184	-0.0274	-0.0128	-0.0347
LR	0.0466	0.0633	0.0708	0.0516	0.0317	0.0404	0.0419	0.0592	0.0515	0.0018	0.0259	0.0392
GR	0.0510	0.0650	0.0679	0.0522	0.0376	0.0409	0.0406	0.0672	0.0633	0.0256	0.0244	0.0497
OU	0.0555	0.0402	0.0852	0.0527	0.0376	0.0552	0.0561	0.0628	0.0624	0.0433	0.0388	0.0411
CO	0.0127	0.0567	0.0725	0.0424	0.0299	0.0198	0.0558	0.0205	0.0252	0.0465	0.0304	0.0599
PG	-0.0046	0.0500	0.0590	0.0399	0.0284	0.0214	0.0412	0.0106	0.0216	0.0223	0.0239	0.0560
MS	-0.0003	0.0426	0.0645	0.0426	0.0287	0.0283	0.0508	0.0360	0.0243	0.0348	0.0212	0.0544
TU	0.0019	0.0487	0.0621	0.0373	0.0261	0.0365	0.0422	0.0228	0.0139	0.0285	0.0326	0.0492
LC	0.0067	0.0610	0.0631	0.0430	0.0317	0.0274	0.0400	0.0195	0.0171	0.0298	0.0341	0.0476
S1	0.4219	0.4847	0.3438	0.3468	0.3435	0.4551	0.3494	0.4779	0.3515	0.3783	0.3311	0.4213
S2	0.3810	0.4544	0.2990	0.3058	0.2917	0.4178	0.3017	0.4464	0.3026	0.3247	0.2797	0.3664

*Annexe II : Tableau des indices de différenciation ( $F_{ST}$ ) inter-sites*

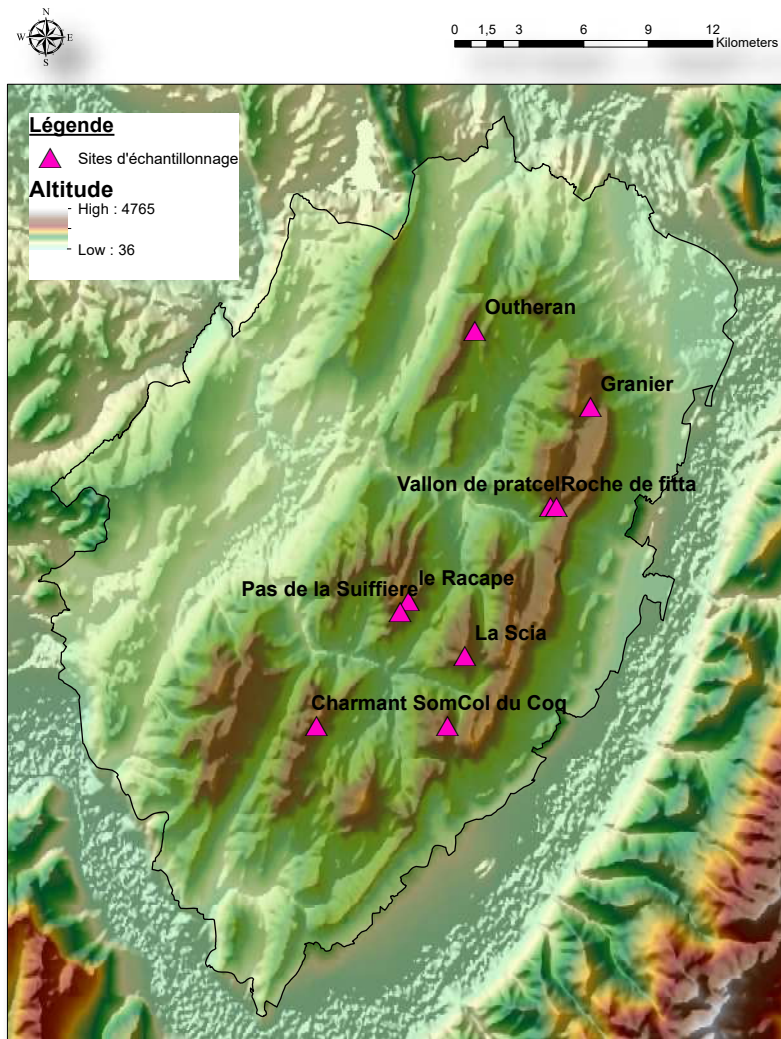
Légende $F_{ST}$	
	$F_{ST}$ significatif
	$F_{ST}$ non significatif

Légende Sites	
	Belledonne
	Bauges
	Chartreuse
	Vercors
	Sancy

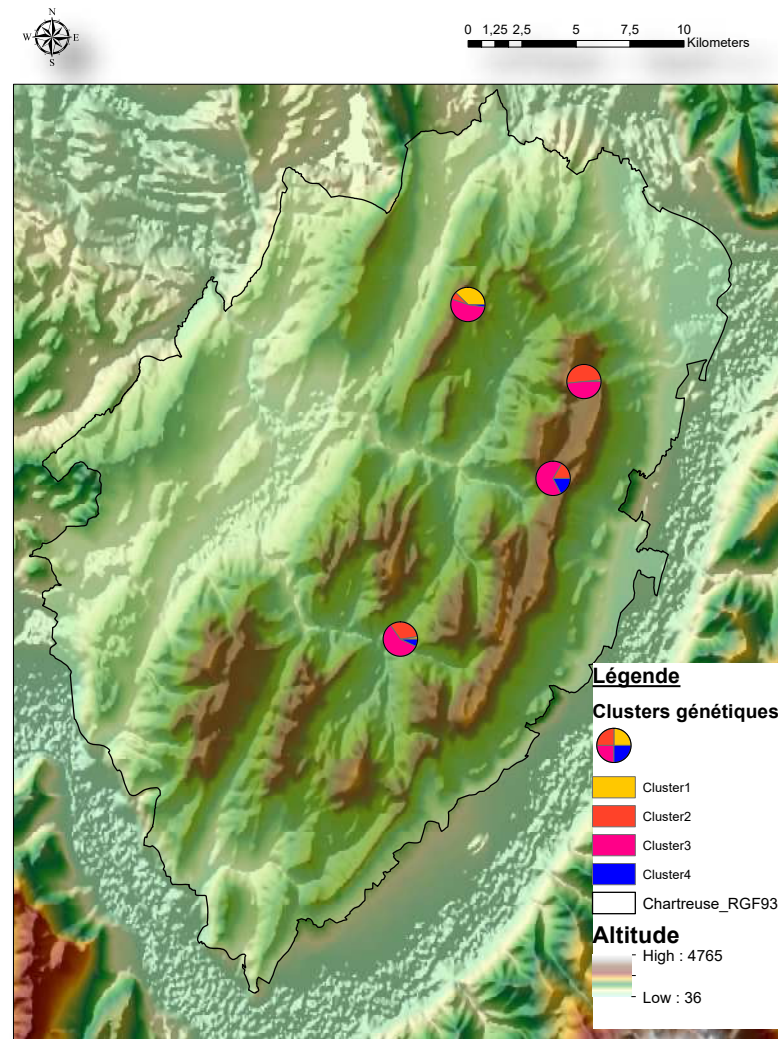


	RF	CS	PS	LR	GR	OU	CO	PG	MS	TU	LC	S1	S2
LV													
MO													
SE													
CC													
AR													
LG													
LC													
CB													
CI													
CQ													
LS													
VP													
RF													
CS	0.0181												
PS	0.0007	0.0025											
LR	0.0178	0.0265	0.0294										
GR	0.0251	0.0405	0.0325	0.0476									
OU	0.0189	0.0459	0.0483	0.0564	0.0497								
CO	0.0378	0.0425	0.0498	0.0509	0.0697	0.0694							
PG	0.0349	0.0368	0.0439	0.0438	0.0563	0.0523	0.0105						
MS	0.0373	0.0372	0.0438	0.0445	0.0606	0.0629	0.0076	0.0053					
TU	0.0322	0.0356	0.0333	0.0436	0.0517	0.0522	0.0126	0.0047	0.0130				
LC	0.0351	0.0419	0.0452	0.0454	0.0547	0.0622	0.0176	0.0011	0.0266	0.0071			
S1	0.3483	0.3320	0.4109	0.3780	0.3430	0.3419	0.3658	0.3438	0.2840	0.2608	0.2669		
S2	0.2913	0.2861	0.3610	0.3235	0.2968	0.2949	0.3237	0.3071	0.2497	0.2291	0.2408	0.0476	

*Annexe II (suite) : Tableau des indices de différenciation (FST) inter-sites*

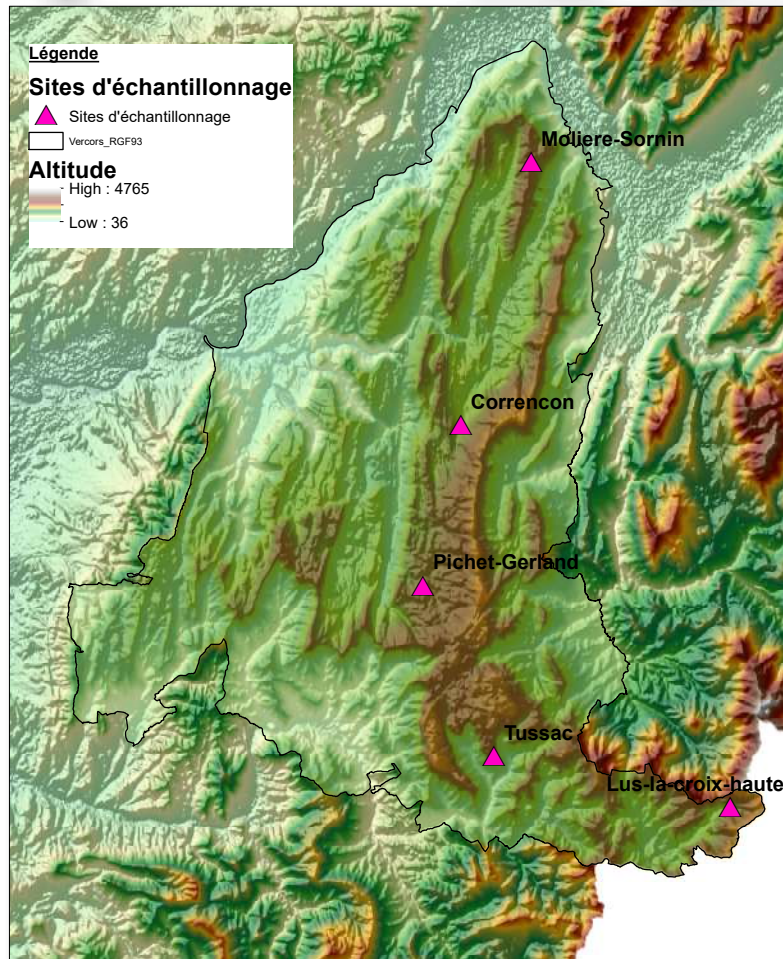


**Sites d'échantillonnage dans le massif de la Chartreuse**

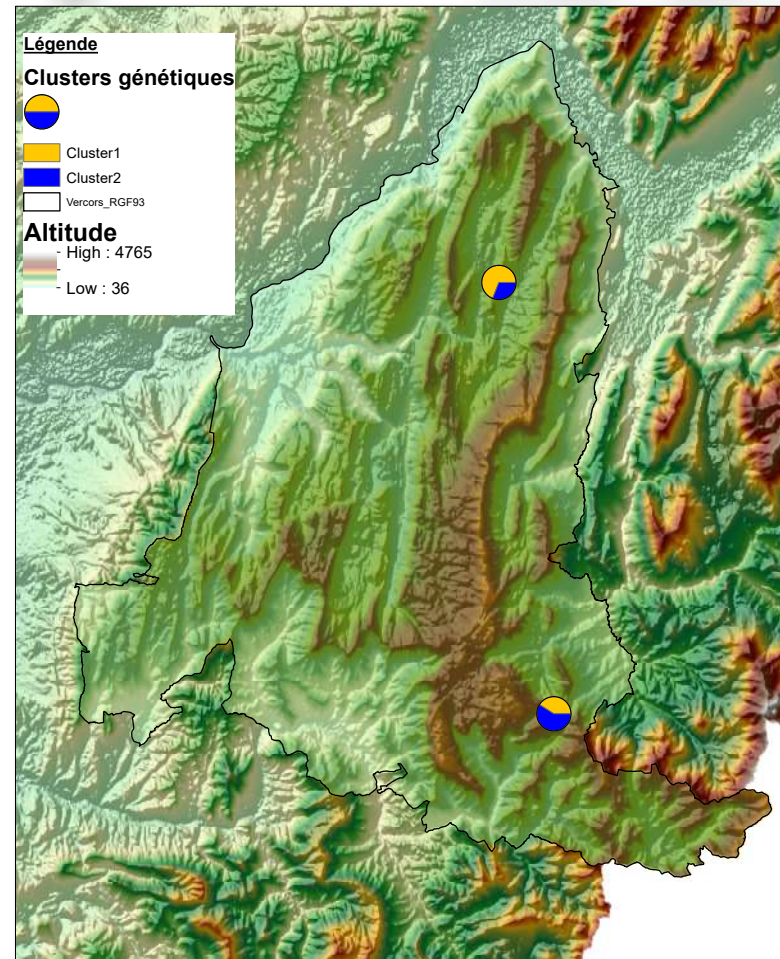


**Clusters génétiques dans le massif de la Chartreuse**

*Annexe III : Cartes des sites et des sous-populations dans les massifs (les sous-populations sont représentées par leur coefficients d'admixture)*

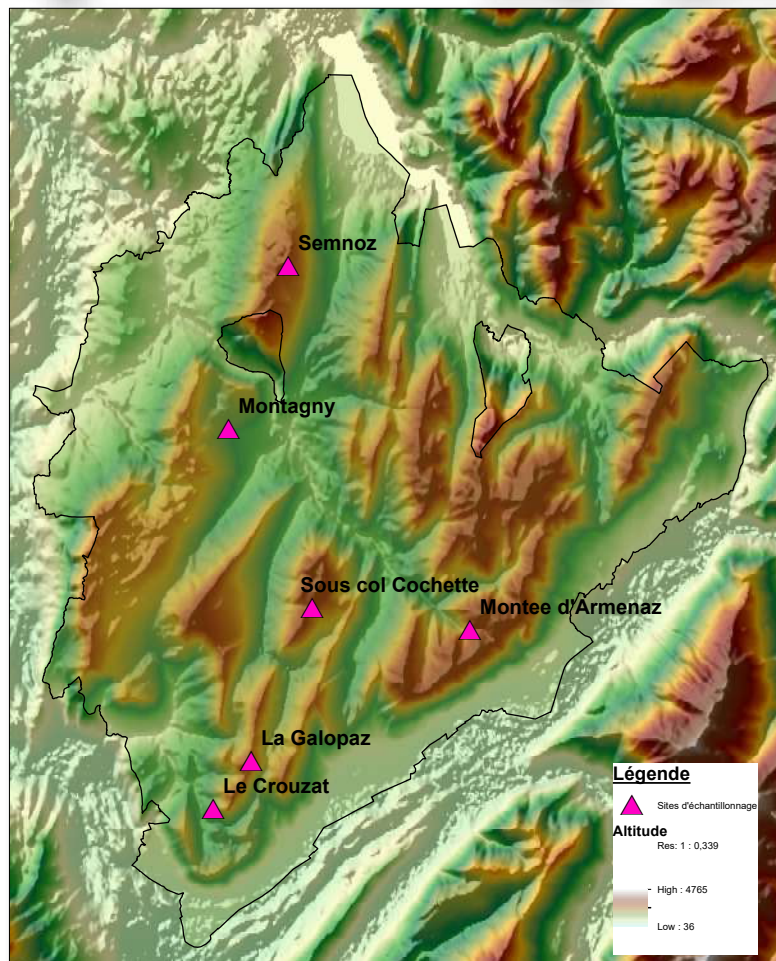


**Sites d'échantillonnage dans le massif du Vercors**

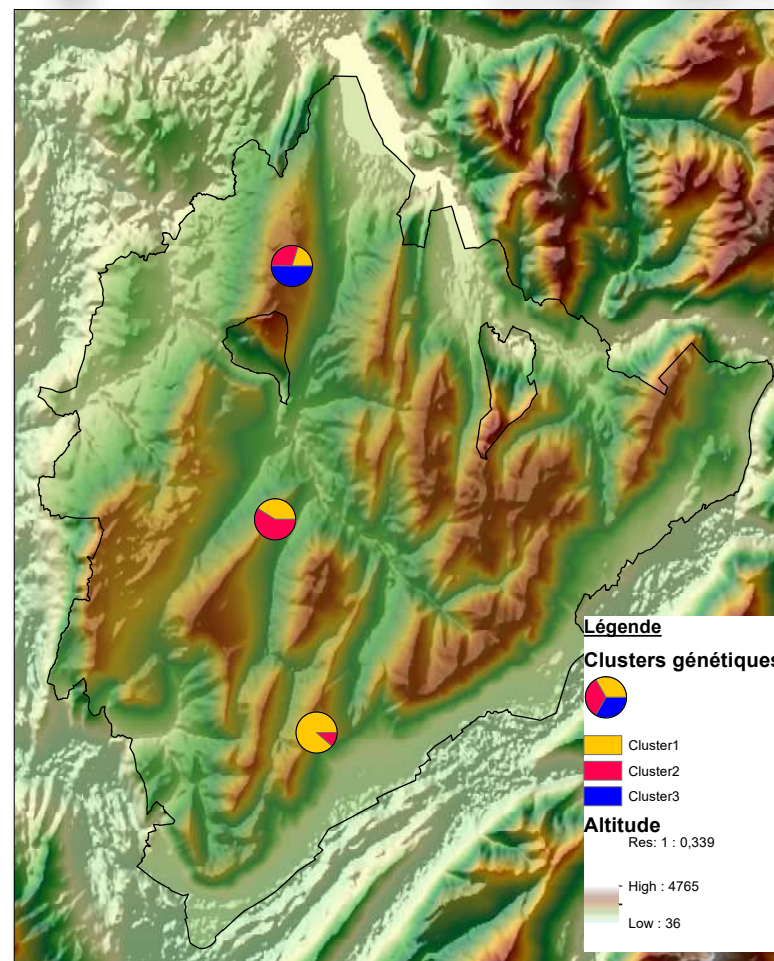


**Clusters génétiques dans le massif du Vercors**

*Annexe III : Cartes des sites et des sous-populations dans les massifs (les sous-populations sont représentées par leur coefficients d'admixture)*

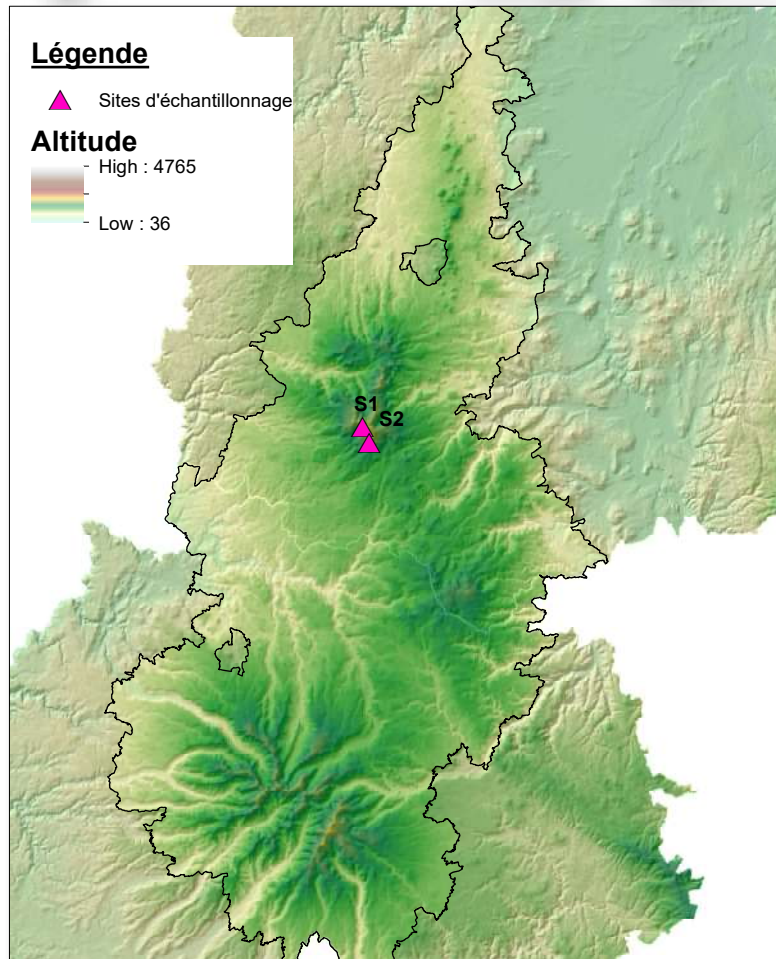


**Sites d'échantillonnage dans le massif des Bauges**

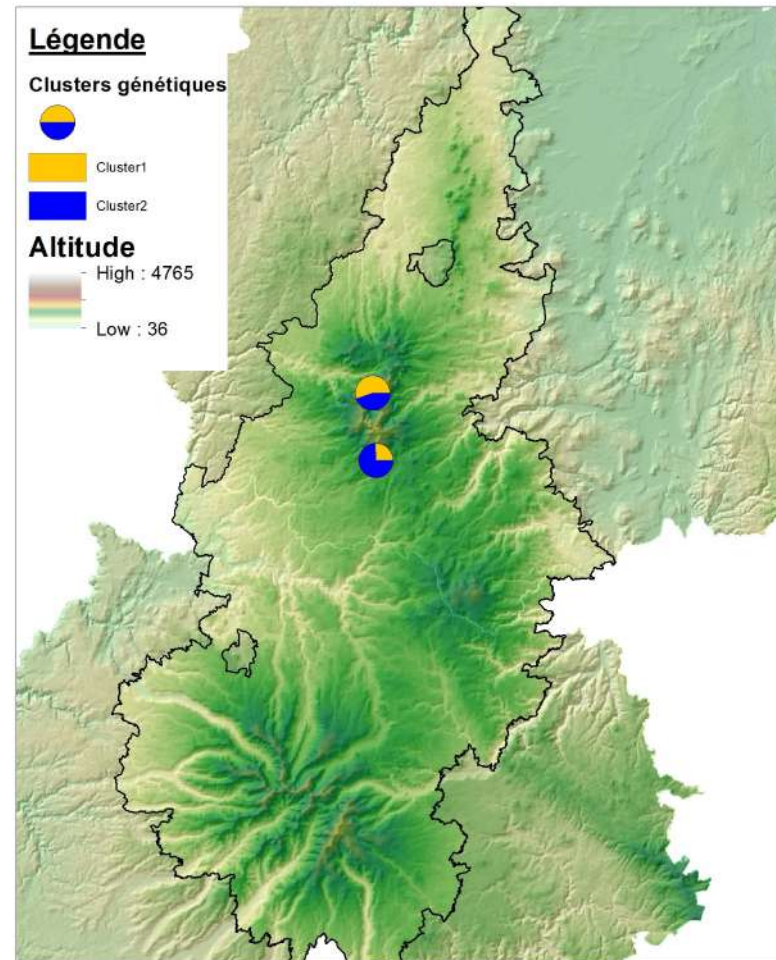


**Clusters génétiques dans le massif des Bauges**

*Annexe III : Cartes des sites et des sous-populations dans les massifs (les sous-populations sont représentées par leur coefficients d'admixture)*



**Sites d'échantillonnage dans le massif du Sancy**



**Clusters génétiques dans le massif du Sancy**

*Annexe III : Cartes des sites et des sous-populations dans les massifs (les sous-populations sont représentées par leur coefficients d'admixture)*

L'Apollon est un lépidoptère, emblématique des montagnes. Cette espèce globalement menacée, est classée en danger d'extinction en Auvergne par l'IUCN. L'emploi de SNPs, obtenu par ddRAD Sequencing, a permis d'étudier la structure génétique des populations de la région Auvergne Rhône-Alpes et les facteurs paysagers et environnementaux qui l'influencent. Les populations des Alpes et de l'Auvergne sont très différenciées génétiquement, ce qui est en faveur de l'existence d'une sous-espèce en Auvergne. La différenciation génétique faible dans les Alpes suggère une bonne connectivité au sein des massifs. L'occupation du sol et les lisières de forêts de conifères influence ce résultat. La diversité génétique est faible en Auvergne, en comparaison avec les Alpes. De plus, les inférences démographiques Bayésiennes montrent des effectifs auvergnats dix fois plus faible que les effectifs alpins, et une divergence entre ces populations en lien avec le réchauffement climatique post-LGM. Un complément d'échantillonnage et d'autres analyses aideront à fournir des préconisations de gestion précises pour la conservation de l'Apollon.

The Apollo is an iconic butterfly of mountains. This globally threatened species is classified as endangered in Auvergne by IUCN. The use of SNPs, obtained by ddRAD Sequencing allowed us to study the genetic structure of populations from the Auvergne Rhône-Alpes region and landscape and environmental factors influencing it. Populations from the Alps and from Auvergne are very differentiated genetically, which is in favour of the existence of a subspecies in Auvergne. The weak genetic differentiation in the Alps suggests a good connectivity within the massifs. Land use and coniferous forests boundaries influences this result. The genetic diversity is weak in Auvergne, in comparison with the Alps. Furthermore, the Bayesian demographic inferences show that the population size in Auvergne is ten times lower than those in the Alps, and a divergence of these populations in connection with the climate warming post-LGM. Additional sampling and analyses will help to give management recommendations for the conservation of the Apollo butterfly.