



**HAL**  
open science

# Méthodes pour l'étude de l'adaptation locale et application au contexte de l'adaptation aux conditions d'altitude chez la plante alpine *Arabis alpina*

Pierre De Villemereuil

► **To cite this version:**

Pierre De Villemereuil. Méthodes pour l'étude de l'adaptation locale et application au contexte de l'adaptation aux conditions d'altitude chez la plante alpine *Arabis alpina*. Evolution [q-bio.PE]. Université Grenoble Alpes, 2016. Français. NNT : 2016GREAS003 . tel-01322336

**HAL Id: tel-01322336**

**<https://theses.hal.science/tel-01322336>**

Submitted on 27 May 2016

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

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

## THÈSE

Pour obtenir le grade de

### DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

Spécialité : Modèles, méthodes et algorithmes pour la biologie, la santé et l'environnement

Arrêté ministériel : 7 août 2006

Présentée par

**Pierre de VILLEMEREUIL**

Thèse dirigée par **Oscar GAGGIOTTI** et  
codirigée par **Irène TILL-BOTTRAUD**

préparée au sein du **Laboratoire d'Écologie Alpine (LECA)**  
dans l'**École Doctorale Ingénierie pour la Santé, la Cognition et l'Environnement**

## Méthodes pour l'étude de l'adaptation locale et application au contexte de l'adaptation aux conditions d'altitude chez la plante alpine *Arabis alpina*.

Thèse soutenue publiquement le **18 Janvier 2016**,  
devant le jury composé de :

**Dr. Fabrice ROUX**

Chargé de recherche, CNRS (Toulouse), Rapporteur

**Pr. Jérôme GOUDET**

Professeur, Université de Lausanne (Lausanne), Rapporteur

**Dr. Michaël BLUM**

Directeur de recherche, CNRS (Grenoble), Examineur

**Pr. Jane LECOMTE**

Professeure, Université Paris-Sud (Orsay), Présidente

**Pr. Oscar GAGGIOTTI**

Professeur, Université Joseph Fourier (Grenoble), Directeur de thèse

**Dr. Irène TILL-BOTTRAUD**

Directrice de recherche, CNRS (Grenoble), Directrice de thèse





*“Now, these points of data make a beautiful line”*  
GLaDOS, *Still Alive*, Portal

# TABLE DES MATIÈRES

---

<b>AVANT-PROPOS</b>	<b>4</b>
<b>INTRODUCTION GÉNÉRALE</b>	<b>6</b>
1 L'adaptation locale : concepts et mécanismes . . . . .	9
2 Méthodologies associées à l'étude empirique de l'adaptation locale . . . . .	15
3 Intérêt d' <i>Arabidopsis thaliana</i> pour l'étude de l'adaptation locale . . . . .	21
4 Problématique et plan . . . . .	23
<b>PREMIÈRE PARTIE : MÉTHODOLOGIES STATISTIQUES ET EMPIRIQUES LIÉES À L'ÉTUDE DE L'ADAPTATION LOCALE</b>	<b>25</b>
<b>CHAPITRE 1 : Étude et comparaison des méthodes de scan génomique pour détecter la sélection</b>	<b>26</b>
1 Le problème des scans génomiques de sélection . . . . .	26
2 Une nouvelle génération de méthodes de scan génomique . . . . .	27
3 Tests multiples et taux d'erreur . . . . .	28
4 Présentation de l'ARTICLE 1 . . . . .	30
<b>ARTICLE 1</b> . . . . .	<b>32</b>
Introduction . . . . .	32
Material & Methods . . . . .	33
Results . . . . .	37
Discussion . . . . .	43
5 Conclusions principales de l'article . . . . .	47
6 Un mot sur la simulation de sélection polygénique . . . . .	47
<b>CHAPITRE 2 : Une nouvelle méthode de scan génomique pour détecter l'adaptation locale</b>	<b>49</b>
1 Une méthode à $F_{ST}$ plus robuste ? . . . . .	49
2 Le modèle derrière Bayescan : le modèle $F$ . . . . .	50
3 Présentation de l'ARTICLE 2 . . . . .	51
<b>ARTICLE 2</b> . . . . .	<b>53</b>
Introduction . . . . .	53
Statistical model . . . . .	55
Material & Methods . . . . .	56
Results . . . . .	61
Discussion . . . . .	63
4 Caractéristiques de la nouvelle méthode . . . . .	68
5 Pourquoi cette profusion de méthodes ? . . . . .	69
<b>CHAPITRE 3 : L'expérience de jardin commun à l'ère génomique</b>	<b>71</b>
1 Jardins communs et génotypage haut-débit . . . . .	71
2 Présentation de l'ARTICLE 3 . . . . .	71

<b>ARTICLE 3</b> . . . . .	73
Introduction . . . . .	73
High-throughput genotyping in the context of a common garden . . . . .	75
Common gardens 2.0: new markers and new methods . . . . .	76
What is the use of common garden experiment in the genomic era? . . . . .	77
Conclusion . . . . .	79
3 Message principal de l'article . . . . .	81
4 Un mot sur les traits non Gaussiens . . . . .	82
<b>DEUXIÈME PARTIE : ADAPTATION AUX CONDITIONS D'ALTITUDE CHEZ LA PLANTE ALPINE <i>ARABIS ALPINA</i></b>	<b>84</b>
<b>CHAPITRE 4 : Étude démographique <i>in situ</i> d'<i>Arabis alpina</i></b>	<b>85</b>
1 Une plante écologiquement particulière? . . . . .	85
2 <i>Arabis alpina</i> au Laboratoire d'Écologie Alpine . . . . .	86
3 Présentation de l'ARTICLE 4 . . . . .	87
<b>ARTICLE 4</b> . . . . .	90
Introduction . . . . .	90
Material & Methods . . . . .	91
Results . . . . .	95
Discussion . . . . .	100
4 Quelques données plus anecdotiques . . . . .	104
5 Conclusions principales de l'article . . . . .	104
<b>CHAPITRE 5 : Étude de l'adaptation locale chez <i>Arabis alpina</i> à l'aide d'une expérience de jardin commun</b>	<b>106</b>
1 Plasticité phénotypique ou adaptation locale? . . . . .	106
2 Méthodes statistiques et moléculaires utilisées . . . . .	108
3 Présentation de l'ARTICLE 5 . . . . .	109
<b>ARTICLE 5</b> . . . . .	111
Introduction . . . . .	111
Material & Methods . . . . .	113
Results . . . . .	117
Discussion . . . . .	124
4 Adaptation locale le long d'un gradient altitudinal . . . . .	128
5 Du lien entre le phénotype <i>in situ</i> et en jardin commun . . . . .	130
6 Petit retour sur la méthodologie . . . . .	131
<b>DISCUSSION GÉNÉRALE</b>	<b>134</b>
1 Discussion de l'ensemble des résultats sur <i>Arabis alpina</i> . . . . .	135
2 Les obstacles méthodologiques liés à l'étude de l'adaptation locale . . . . .	138
3 Perspectives méthodologiques . . . . .	141
4 Conclusion générale . . . . .	143

---

<b>ANNEXES</b>	<b>145</b>
<b>ANNEXE S1 : Supplementary Information (ARTICLE 1)</b>	<b>146</b>
<b>ANNEXE S2 : Supplementary Information (ARTICLE 2)</b>	<b>156</b>
<b>ANNEXE S3 : Méthodes de génétique quantitative pour les modèles linéaires mixtes généralisés (ARTICLE 6)</b>	<b>163</b>
<b>ANNEXE S4 : Supplementary Information (ARTICLE 4)</b>	<b>180</b>
<b>ANNEXE S5 : Supplementary Information (ARTICLE 5)</b>	<b>183</b>
<b>ANNEXE S6 : Supplementary Information (ARTICLE 6)</b>	<b>186</b>
<b>RÉFÉRENCES BIBLIOGRAPHIQUES</b>	<b>190</b>

# AVANT-PROPOS

---

Avant de rentrer dans le vif du sujet, il convient toujours de donner un peu de contexte. Le contexte scientifique va être exposé en long, en large et en travers dans les pages suivantes, concentrons nous ici sur le contexte humain. Commençons par le commencement : je dois d'avoir connu Oscar à Amaury Lambert, qui a pris beaucoup de temps à m'aider à trouver un directeur de thèse qui correspondait à mes centres d'intérêts. Merci à lui ! J'ai donc rencontré Oscar en 2010 et suis arrivé au Laboratoire d'Écologie Alpine en 2012 (après une année complète à Lyon). Je lui avais demandé s'il était possible de faire avec lui une thèse où je pourrais collecter mes propres données sur le terrain et les analyser, en plus de travailler sur des développements statistiques : vous pourrez constater en lisant cette thèse qu'il a su viser juste ! J'ai rencontré Irène en 2011, quand Oscar m'a proposé une co-direction sur ma thèse. Suite au dépôt d'un projet ANR qui ne sera malheureusement pas financé, Irène m'a proposé de prendre en charge une expérience de jardin commun sur *Arabis alpina* : cette expérience fait désormais complètement parti de ma thèse.

En réalité, le contenu de cette thèse aura été (comme pour beaucoup d'autres) à géométrie variable. À l'origine, une méthode pour l'étude de l'architecture des traits phénotypiques devait en faire partie, elle sera abandonnée pour des raisons de temps et de complexité. Le génotypage, aussi, devait être fait directement sur les individus du suivi démographique : à cause de la faible qualité des échantillons, il ne sera pas mené au bout, après presque un an de tentatives. Nous nous sommes donc réorientés vers le génotypage des individus du jardin commun, ce qui, fort heureusement, nous a permis d'élaborer des analyses (je l'espère) intéressantes et novatrices.

Malgré tout, il me semble, en toute subjectivité, que le résultat final est plutôt cohérent (malgré le manque de temps pour finaliser les dernières analyses...), et qu'en plus de présenter des résultats intéressants sur l'écologie évolutive d'*Arabis alpina*, il permet de dresser un petit bilan méthodologique des problèmes que posent certaines questions évolutives et des approches méthodologiques disponibles pour les surmonter. Quoi qu'il en soit, j'espère que vous trouverez des choses intéressantes dans ces travaux, que vous soyez fans des choux alpins ou non.

Beaucoup de gens ont participé, de manière plus ou moins directe à ces travaux, que je me dois de remercier, bien sûr. Oscar et Irène sont les premiers bien sûr : pour leur support, humain et scientifique, pour leur disponibilité (même à travers la Manche, merci Skype™ !), pour la confiance qu'ils ont eu en moi et surtout pour l'incroyable autonomie qu'ils m'ont laissé durant ces années ! Ensuite, merci aux nombreux gens avec qui j'ai pu collaborer directement sur ces questions : Marco Andrello, Médéric Mouterde, Éric Bazin, Éric Frichot, Olivier François, Michaël Blum, Stéphane Lobréaux, Christelle Mélo-de-Lima et Michel Herzog. Merci aussi à Sandra Lavorel, Philippe Choler et Fabrice Roux pour leur discussions sur *Arabis alpina*, la mise en place de l'expérience et nos résultats. Merci à d'autres personnes avec qui j'ai pu collaborer et qui ont, indirectement, beaucoup apporté à cette thèse notamment Shinichi Nakagawa, Michael Morrissey et Holger Schielzeth.



Absolument rien n'aurait pu être fait sans les innombrables personnes qui ont participé au terrain. Une mention spéciale à Aymeric Pilleux qui a fait un travail incroyable pour mettre en place le jardin, avec l'aide de Pascal Salze et d'Amandine Deschamps, ainsi qu'à Quentin Rousseau, Perrine Augrit et Sofia Karabatsos qui se sont retrouvés avec la tâche ingrate de m'aider durant mes campagnes de phénotypages. Merci à tous les autres pour leur aide ponctuelle mais néanmoins précieuse : Florian Alberto, Charles Pouchon, Tristan Cumer, Mónica Cecilia Corona, Thomas Goeury et Alex Farajdo.

Il n'y a pas que le terrain dans la vie, il y a aussi le labo. Merci donc à tous les gens géniaux qui hantent le 3e étage du LECA (et le caf'litt', sic) dont Céline Mercier, Florian Alberto, Fred Boyer, Aurélie Bonin, Marta De Barba, Thibaut Capblancq, Médéric Mouterde, Charles Pouchon, Tristan Cumer, Marco Andrello, Carla Martins-Lopez... Une mention spéciale pour Fred Boyer et Céline Mercier pour leurs nombreux coups de pouce en bioinformatique, ainsi qu'à Thibaut Capblancq, le maître ès *RAD sequencing* du laboratoire !

Merci aussi beaucoup aux secrétaires du laboratoire (Florence Sagnimorte, Nancy Iacono, Kim Pla, Marie Costarigot...) qui, malgré les difficultés de personnel, font un travail remarquable pour nous aider au quotidien.

Un grand merci aussi à mon ami et ancien coloc, Loïc Chalmandrier qui, non content de me supporter tous les soirs, de faire une bouffe incroyable et d'avoir l'art de m'entraîner dans des débats sans fin, a donné du sien pour le terrain lui-aussi ! Merci aussi à mes autres colocs successifs de m'avoir supporté sans trop broncher : Jérémy Guéré et Coline Byczek.

Une personne en particulier, a beaucoup participé à faire de cette thèse un succès, il s'agit de ma fiancée Justine Bisson, qui aura donné de sa sueur, et surtout prêté son talent incomparable pour l'organisation, afin de faire de mes campagnes de terrain un succès, résolu un nombre incroyable de situations par ses idées brillantes (et apparemment évidentes, d'après elle) et aura toujours été d'un support inconditionnel. Un grand merci à elle pour tout ce qu'elle est.

Finalement, que serait des remerciements sans la famille ? Merci donc à mes parents, mon frère et ma sœur sans qui (suite à de sombres histoires de canalisation développementale, je suppose) je ne serais probablement pas ce que je suis aujourd'hui.

# **INTRODUCTION GÉNÉRALE**

---

**L'étude de l'adaptation locale :  
exemple des systèmes alpins**

L'adaptation, que le sens commun définit comme une harmonie apparente entre les caractéristiques des êtres vivants et leur mode de vie, est un phénomène qui a toujours fasciné philosophes, naturalistes et scientifiques. Dès les travaux d'Aristote, philosophe et naturaliste (l'un des premiers à tenter une classification des êtres vivants), cette adéquation est particulièrement mise en avant pour justifier une vision finaliste du monde. Cette vision adaptationniste et finaliste du vivant culminera, au sein des philosophies théistes, avec la philosophie optimiste de Leibniz, puis la théologie naturelle au début du XIX<sup>e</sup> siècle (HODGE & RADICK, 2009, p. 155–157). Ces philosophies s'appuyaient sur l'argument de l'harmonie naturelle pour justifier non seulement l'ordre naturel, mais aussi son origine divine. William Paley, en 1802, développe dans son traité de théologie naturelle un argument "adaptationniste" puissant en faveur des thèses créationnistes, qui peut se résumer comme-ci : imaginez que vous trouviez une montre sur un chemin, il ne vous viendrait pas à l'idée que cet artefact puisse être issu d'un quelconque processus aléatoire, mais qu'elle soit bien la création d'un artisan particulièrement doué, en l'occurrence, un horloger. Imaginez encore que vous ne sachiez pas l'utilité de l'objet en question : sa sophistication, à elle seule, vous permet de certifier qu'il existe une intelligence derrière sa création. Cette analogie reste encore un des plus célèbres arguments en faveur des visions créationnistes (SOBER, 2008, p.118–120).

Ironiquement, cette vision adaptationniste (mais fondamentalement fixiste) du vivant influencera beaucoup Charles Darwin lors de ses réflexions sur l'origine des espèces. Cependant, son intérêt ne se focalisera pas tant sur une parfaite adaptation de chaque *espèce* à son milieu et son mode de vie qu'à l'adéquation remarquable entre un *caractère* et l'accomplissement de sa fonction (HODGE & RADICK, 2009, p. 155). En outre, cet intérêt pour les traits adaptatifs s'accompagne d'une vision transformiste (c'est-à-dire évolutive) du vivant. Ce passage d'une vision adaptationniste et fixiste à la vision adaptationniste et transformiste de Darwin requiert donc une légère redéfinition du concept d'adaptation. En effet, en lieu et place d'un concept phénoménologique qui ne fait qu'apprécier l'adéquation entre les êtres vivants et leurs modes de vie, le darwinisme nous offre un concept mécaniste qui, lui, met l'accent sur l'origine sélective, et donc évolutive, des *traits adaptatifs*, et qui n'a que peu de choses à voir avec une vision "harmonique" de l'ordre naturel.

L'idée fondamentale du darwinisme est en effet que les êtres vivants évoluent majoritairement au travers de la sélection naturelle. Au sens stricte, l'adaptation darwinienne est donc définie comme le résultat (dans une population ou une espèce donnée) d'une évolution par sélection naturelle, lors de laquelle les caractères portés par les individus qui possèdent la plus forte valeur sélective (ceux qui laissent le plus de descendants en moyenne) envahissent la population dans laquelle ils résident. Cette définition est donc ontologiquement indépendante de la notion d'adéquation entre les êtres vivants et leurs modes de vie. Ce phénomène ne devient en effet qu'un résultat apparent des mécanismes adaptatifs. Pour autant, la dialectique entre ces deux visions est, comme nous allons le voir, toujours d'actualité. Elle pose particulièrement problème lorsqu'il s'agit de définir l'adaptation pour des cas où les définitions anciennes et modernes s'opposent, comme lorsqu'il s'agit de répondre à la question de savoir si la sélection sexuelle est maladaptative. Alors que la première définition (adéquation au mode de vie) considère clairement la queue du paon comme une maladaptation, la seconde (trait soumis à la sélection) serait clairement plus hésitante vis-à-vis de cette conclusion.

Après une certaine réticence historique vis-à-vis de l'importance de la sélection naturelle (HODGE & RADICK, 2009, p. 286–291), celle-ci est pourtant devenue la clé de voûte d'un paradigme dit "adaptationniste" qui a particulièrement dominé la biologie évolutive au cours du XX<sup>e</sup> siècle. Dans la seconde moitié du siècle cependant, la force de ce paradigme s'est retrouvée affaiblie par la théorie neutraliste de l'évolution moléculaire (KIMURA, 1985) d'une part, et par de sévères critiques à l'encontre d'un panadaptationnisme latent d'autre part. À ce titre, la fameuse critique de GOULD &

LEWONTIN (1979) est intéressante parce qu'elle fait alternativement appel aux définitions anciennes et modernes "d'adaptation"<sup>1</sup>, ce qui montre l'importance encore problématique de la définition ancienne dans la biologie moderne. La réponse de l'adaptationniste DAWKINS (1982, p. 50–51) est tout aussi intéressante en cela qu'elle recadre la notion d'adaptation dans sa version darwinienne (bien que DAWKINS utilise "adaptation" dans les deux sens du terme, il insiste bel et bien sur les notions mécanistiques et sélectionnistes de l'adaptationnisme "néo-darwinienne").

À l'aide de ces controverses, il est possible d'établir les conditions d'un adaptationnisme que l'on pourrait caractériser d'épistémologiquement cohérent :

- Tout d'abord, il est nécessaire que le paradigme adaptationniste reste falsifiable, c'est à dire qu'il soit en compétition avec des théories non adaptatives (exemple : les membres antérieurs réduits du *Tyrannosaurus rex* doivent-ils nécessairement avoir un intérêt adaptatif?),
- Ensuite, les explications adaptationnistes se doivent d'être mécanistes : c'est à dire qu'elles doivent prendre en compte la possibilité évolutive pour qu'une adaptation envahisse la population qui l'accueille (exemple : sans mécanisme spécifique, l'altruisme ne peut pas envahir une population donnée, quand bien même il est bénéfique au niveau de la population),
- Enfin, tout processus de sélection naturelle se fait dans un système éco-biologique donné, qui possède un certain nombre de contraintes développementales, morpho-physiologiques et écologiques (GOULD & LEWONTIN, 1979 ; DAWKINS, 1982).

Mais surtout, la courte analyse historique que nous venons de faire montre qu'il est extrêmement important d'évacuer l'ambiguïté dans laquelle baigne le terme *adaptation*. Par conséquent, afin d'éviter toute confusion sur le sens de ce mot, il est de bon goût de préciser ici que je n'emploierai ce terme que dans son sens moderne. Le cas échéant, j'utiliserai le terme d'*adéquation* lors qu'il sera spécifiquement question du lien entre les traits des organismes vivants et leur mode de vie. Ainsi, par l'expression *adaptation à la température*, je définirai les traits/les locus dont l'évolution a été façonnée (entre autres !) par la sélection naturelle suite à l'action du facteur environnemental *température*. Au même titre, lorsque j'emploierai le terme de *trait adaptatif* j'impliquerai que ce trait est, ou a été, soumis à la sélection naturelle. Finalement, un organisme est dit *adapté* si son phénotype est optimal au sens de la sélection naturelle. L'adéquation entre les caractéristiques des espèces et leur mode de vie qui peut résulter de ce processus est un sous-produit extrêmement intéressant (notamment dans le cadre des changements globaux), mais fonctionnellement indépendant de notre définition d'adaptation.

Jusqu'à nos jours, ces débats sur l'adaptation ont soulevé de nombreuses questions fondamentales concernant l'étude de la sélection naturelle, de l'évaluation de ses mécanismes, de sa prévalence et surtout de son influence relative dans les processus évolutifs. Au cœur de ces nouveaux développements se situe la question de la preuve du lien entre les variations du trait adaptatif étudié et la valeur sélective des individus, de l'importance de la plasticité phénotypique (voir plus bas pour une définition formelle), ainsi que des bases génétiques des traits adaptatifs. Ce dernier domaine de recherche en particulier est en continuelle expansion depuis environ deux décennies. Dans le cadre de l'adaptation intra-spécifique, un phénomène nommé *adaptation locale* est fortement étudié, et notamment fait l'objet de cette thèse.

---

<sup>1</sup>Le néologisme de "panglossisme" forgé par les auteurs est extrêmement révélateur, à ce titre : Pangloss étant un personnage inventé par Voltaire pour caricaturer... la philosophie optimiste de Leibniz !

# 1 L'adaptation locale : concepts et mécanismes

## 1.1 Dynamique d'une méta-population

Une méta-population est un ensemble de populations en contact à travers des flux de gènes<sup>1</sup> (FIG. 1). Ces flux de gènes peuvent être dus à la dispersion soit d'individus, soit de gamètes (pollen, entre autres) entre populations. Au sein de tels systèmes, l'évolution des fréquences alléliques est guidée par quatre forces majeures :

- **La mutation** qui permet l'apparition de nouveaux allèles ;
- **La dérive génétique** qui fait évoluer les fréquences alléliques au hasard, plus particulièrement dans les populations de petite taille ;
- **La dispersion** qui crée du flux de gènes entre populations ;
- **La sélection naturelle** qui guide l'évolution de certains allèles bénéfiques ou délétères.

Lorsqu'il n'y a pas de sélection (on parle alors de *neutralité*), seules les trois premières forces agissent sur l'évolution des fréquences alléliques. Le phénomène de dérive crée une différenciation aléatoire entre populations, qui est contre-balancée par les flux de gènes engendrés par la dispersion. Ainsi, si les populations sont de petites tailles ou si la dispersion entre populations est limitée, la dérive engendrera une forte différenciation aléatoire entre populations ; dans le cas contraire, avec des populations de grande taille ou une forte dispersion, les populations seront plus homogènes entre elles.

On caractérise ce continuum de scénarios à l'aide d'indicateurs de la *différenciation génétique* entre populations, dont le plus connu est la  $F_{ST}$  (WEIR & COCKERHAM, 1984). La  $F_{ST}$  peut être définie généralement comme le rapport de la variabilité des fréquences alléliques entre populations sur la variabilité des fréquences alléliques au niveau de la méta-population. Une  $F_{ST}$  proche de 0 témoigne ainsi d'une faible différenciation entre populations, tandis qu'une forte  $F_{ST}$  (avec un maximum, rarement atteint, de 1) est le signe d'une forte différenciation génétique entre populations. Dans la FIG. 1, par exemple, les populations B et C sont différenciées au hasard, par la suite de l'action de la dérive génétique.

Cette description simple de la dynamique de la méta-population masque en réalité l'extrême diversité des contingences historiques et démographiques qui interviennent dans la dynamique d'une méta-population. En effet, les fréquences alléliques sont affectés par un grand nombre de processus stochastiques en lien avec les modalités précises de migration (migration asymétrique ou préférentielle entre populations), les variations anciennes et récentes de taille de population (dont l'exemple le plus extrême est le goulot d'étranglement, c'est à dire une ancienne et très importante réduction de la taille de population) ou des colonisations récentes de nouveaux milieux. La prise en compte de ces scénarios particuliers, notamment de l'incertitude et de la stochasticité qui leurs sont associées, est un des plus grand défi de la biologie évolutive.

Lorsque la sélection naturelle agit sur l'évolution des fréquences alléliques, la description de la situation se complique davantage, car tout dépend de la nature du régime sélectif. L'un des régimes sélectifs le plus simple possible consiste en la sélection du même allèle dans toutes les populations : dans ce cas, on s'attend à ce que la fréquence de cet allèle augmente dans toutes les populations, jusqu'à atteindre la fixation (c'est à dire une fréquence de 1) dans toute la méta-population. Le scénario d'adaptation locale représente un régime sélectif légèrement plus compliqué.

---

<sup>1</sup>Notons que, historiquement, *méta-population* renvoie à une dynamique plus précise d'extinction/colonisation. En utilisant cette définition plus large, je poursuis un abus de langage devenu classique en génétique des populations.

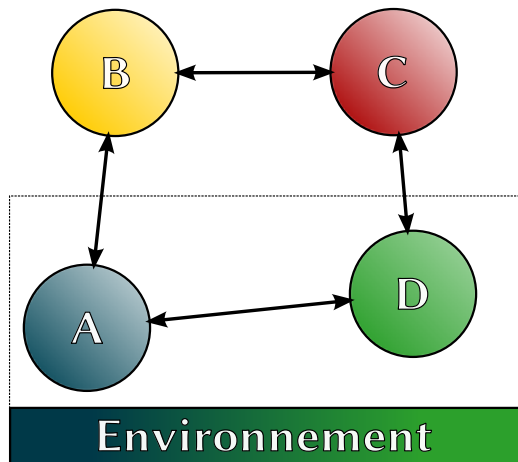


FIG. 1 : Une méta-population. Les populations sont figurées par des ronds, les flux de gènes par des flèches. Un environnement sélectif crée une sélection différentielle entre la population A et la population D, entraînant un phénomène d'adaptation locale : la population A est adaptée à un environnement "bleu", tandis que la population D est adaptée à un environnement "vert". Les différences entre les populations B et C sont, elles, simplement dues aux forces neutres (dérive génétique et dispersion).

## 1.2 Qu'est-ce que l'adaptation locale ?

L'adaptation locale intervient lorsque les populations qui constituent la méta-population se retrouvent dans des environnements (biotiques ou abiotiques) différents. Si ces environnements exercent des pressions de sélection différentes dans les différentes populations, alors il est possible que chaque population évolue vers un optimum adaptatif local. Dans la FIG. 1, par exemple, la population A se retrouve dans un environnement "bleu", tandis que la population D se retrouve dans un environnement "vert" : il en résulte une différenciation entre ces deux populations dirigée par l'environnement, plutôt que par le hasard. Dans cet exemple hypothétique, puisque l'environnement n'influe pas sur les populations B et C, la différenciation avec ces deux populations se fait, elle, au hasard.

Cette dynamique d'adaptation locale amène au résultat suivant : on s'attend à ce que les individus possèdent une meilleure valeur sélective dans leur environnement local que les immigrants (KAWECKI & EBERT, 2004). Dans l'exemple de la FIG. 1, si l'on devait placer des individus de chaque population dans un environnement "bleu", on s'attendrait, en moyenne, à ce que les individus de la population A s'en "sortent mieux", c'est à dire qu'ils se reproduisent ou survivent mieux que les autres. Ce principe est le fondement de l'expérience de *transplantation réciproque* (KAWECKI & EBERT, 2004) qui sera développée plus bas (voir Section 2).

Les considérations générales ci-dessus cachent en réalité plusieurs caractéristiques complexes du phénomène que nous allons étudier. Tout d'abord, la sélection agit sur des phénotypes, dont la variabilité est contrôlée par un certain nombre de locus au sein du génome. Ainsi, alors que l'essentiel du génome est impacté par la dynamique neutre de la méta-population, seul un certain nombre de locus est impacté par celle de l'adaptation locale. Pis encore, la dynamique neutre vient se superposer à celle de l'adaptation locale pour ces locus, ce qui tend à bruite le signal adaptatif.

Ensuite, malgré la présence d'un différentiel environnemental, il est possible que l'on n'aboutisse pas à une situation d'adaptation locale (KAWECKI & EBERT, 2004). C'est notamment le cas lorsque les flux de gènes entre populations sont tels que toute adaptation est immédiatement noyée parmi des migrants non localement adaptés ; lorsque la dérive est tellement forte (par exemple dans de petites populations) qu'elle prend le pas sur la dynamique sélective ; ou enfin si, malgré une forte pression de sélection, le potentiel adaptatif de la population (c'est à dire sa variabilité génétique) n'est pas suffisante pour atteindre un optimum adaptatif.

Enfin, l'environnement est en fait un ensemble composite. Afin d'appréhender sa complexité, il est usuel d'utiliser une approche réductionniste consistant à séparer cet ensemble en différentes va-

riables environnementales. Bien que cette démarche soit essentielle pour décrypter les contributions relatives de ces variables à la sélection naturelle et à l'adaptation locale, elle pose de nombreux problèmes. Premièrement, il est impossible de décrire totalement un environnement et ses variations fines avec un nombre fini (et souvent réduit) de variables. Il est donc probable que certaines variables environnementales fondamentales soient tout simplement absentes de l'analyse. Deuxièmement, le découpage en variables environnementales génère régulièrement des variables dont la nature est complètement différente, mais qui peuvent être extrêmement corrélées entre elles. Par exemple, les milieux alpins sont caractérisés par une diminution de la température, mais aussi une augmentation du rayonnement solaire, avec l'altitude (KÖRNER, 2003). Étudier l'effet sélectif de la température indépendamment des radiations solaires peut donc s'avérer complexe, voire impossible. Troisièmement, la variabilité de l'environnement est caractérisée très différemment à différentes échelles. Par exemple, alors que la température moyenne diminue avec l'altitude, elle est aussi influencée par le climat à l'échelle régionale, la topologie du massif, ainsi que par des facteurs micro-topologiques (cailloux, trous, etc...). L'étude de l'adaptation locale à ces différentes échelles (régionale, locale, micro-locale) peut donner des résultats très différents (BRACHI et al., 2013). Dans mes travaux, je me suis concentré principalement sur l'échelle locale.

### 1.3 Pourquoi étudier l'adaptation locale ?

L'adaptation est un concept clé de la biologie évolutive, et notamment du paradigme darwinien. Pourtant, l'étude fine des processus adaptatifs n'est pas toujours aisée. Dans l'exemple archétypal du long cou de la girafe, par exemple, la démonstration de la nature adaptative de ce caractère requiert plus que l'observation qu'il permet à la girafe de se nourrir. Ce serait retomber dans le panglossisme dénoncé par GOULD & LEWONTIN (1979). Il faudrait en plus montrer qu'une évolution graduelle de la taille du cou est possible, ainsi que préciser le mécanisme sélectif qui entraînerait cette évolution adaptative. Même à ce stade, la *sous-détermination* de l'observation ("les girafes ont un long cou") par les nombreuses théories évolutives possibles rendrait les conclusions spéculatives, puisque plusieurs théories concurrentes peuvent mener au même résultat observable.

Deux approches complémentaires se distinguent quant à l'étude de l'évolution des traits adaptatifs. L'approche macro-évolutive consisterait, dans l'exemple ci-dessus, à comparer les girafes actuelles avec leurs espèces les plus apparentées, ou avec les registres fossiles, afin de montrer le caractère non aléatoire, déterminé, de l'évolution de la taille du cou des girafes. L'approche micro-évolutive consisterait à rechercher des girafes mutantes possédant un cou de plus petite taille pour montrer que celles-ci possèdent une valeur sélective moindre par rapport aux girafes "sauvages". De manière générale, la macro-évolution s'intéresse plutôt aux grands motifs évolutifs (entre espèces, à long terme), tandis que la micro-évolution s'intéresse plutôt aux processus mesurables (au sein d'une espèce, à court terme). Les processus micro-évolutifs sont parfois difficiles à mesurer à cause de temps de génération trop grands ou de difficultés à manipuler les espèces (notamment comment obtenir une girafe mutante à petit cou ?).

Parce que l'adaptation locale crée des situations où une sélection différentielle est mise en place dans différents environnements, elle agit comme une expérience "naturelle" de micro-évolution, où des individus d'une même espèce sont soumis à différents régimes évolutifs. Étudier les conséquences de l'adaptation locale, c'est donc étudier la façon dont les phénomènes micro-évolutifs fonctionnent. Cela permet, par conséquent, d'obtenir des informations sur la nature et les bases génétiques des traits adaptatifs, sur la nature des variables environnementales sélectives, ainsi que sur la façon dont ces différents acteurs interagissent.

L'étude de l'adaptation locale possède aussi un intérêt pratique plus immédiat dans le contexte des changements globaux (PIMM et al., 1995 ; PARMESAN, 2006 ; BELLARD et al., 2012). La survie d'une espèce face aux pressions que font peser sur elle les changements globaux passe en effet par trois principaux mécanismes non exclusifs (AITKEN et al., 2008) :

- **La dispersion**, qui permet de suivre l'évolution géographique de son habitat (par exemple, se déplacer vers le Nord pour fuir le réchauffement climatique) ;
- **La plasticité phénotypique**, qui permet aux organismes de s'acclimater rapidement à de nouvelles conditions environnementales (voir plus bas pour une description plus poussée) ;
- **L'adaptation**, qui permet une réponse évolutive, donc génétique et pérenne, aux changements environnementaux.

Chacun de ces mécanismes possède ses limites, en fonction de la biologie de l'espèce. Le premier, par exemple, est limité bien sûr par la capacité de dispersion des espèces, mais aussi par la présence d'une zone "refuge" dans laquelle disperser. Dans le cas des plantes alpines, par exemple, les massifs ont une altitude maximale finie, qui limite la présence de ces refuges pour les plantes de plus haute altitude. Le second présuppose l'existence de mécanismes de plasticité, qui même lorsqu'ils existent, peuvent être relativement coûteux et limités (DEWITT et al., 1998). Le dernier mécanisme, étant évolutif, est caractérisé par sa lenteur. Pourtant, il s'agit probablement du mécanisme le plus souhaitable, puisqu'il s'agit d'une réponse durable et efficace aux changements environnementaux. Il est cependant limité, en plus de sa lenteur, par le *potentiel adaptatif* de l'espèce, c'est à dire la variabilité génétique associée aux traits phénotypiques existant dans la population soumise à la sélection.

Une propriété intéressante de l'adaptation locale est qu'elle augmente substantiellement ce potentiel adaptatif au niveau de la méta-population, puisqu'elle maintient des niveaux de polymorphisme génétique élevés entre populations pour les traits localement adaptés. Par exemple, dans le cas d'une plante alpine localement adaptée à différents niveaux d'altitude, l'adaptation de la méta-population à un réchauffement climatique serait grandement facilitée par le flux de gènes entre des populations de basse altitude vers des populations de plus haute altitude. Il s'agit en quelque sorte d'un scénario intriquant la dispersion et l'adaptation. D'un point de vue de la conservation, il est supérieur à un scénario de pure dispersion, qui lui, pourrait conduire à une certaine maladaptation suite à l'influence de facteurs micro-locaux.

Pourtant, malgré cet intérêt théorique et pratique assez fort, l'étude de l'adaptation locale reste un exercice difficile. Il l'est d'autant plus que de nombreux phénomènes stochastiques (dérive, migration) ou non (plasticité phénotypique) viennent complexifier l'interprétation des observations. La plasticité phénotypique, notamment, mais aussi sa forme plus complexe, l'interaction génotype-par-environnement ( $G \times E$ ), sont des phénomènes particulièrement intéressants qui méritent de s'y attarder plus avant.

## 1.4 La génétique quantitative, la plasticité phénotypique et l'interaction $G \times E$

### ► La génétique quantitative : modèle de base

La *génétique quantitative* est la discipline de la biologie qui s'intéresse aux bases génétiques des caractères *complexes*, c'est à dire des caractères influencés par un grand nombre de gènes. Elle doit son existence à la formulation du modèle infinitésimal de FISHER (1918), qui permet d'inclure définitivement les traits continus dans une logique mendélienne de l'hérédité. Le principe de ce modèle est que l'effet cumulé d'un grand nombre de gènes à ségrégation mendélienne converge rapidement



vers une distribution normale. L'utilisation, avec succès, de ce cadre conceptuel dans l'amélioration des espèces domestiques animales et végétales a permis l'essor scientifique de cette discipline (HILL & KIRKPATRICK, 2010), en plus d'être en partie responsable de la Révolution Verte au cours du xx<sup>e</sup> siècle.

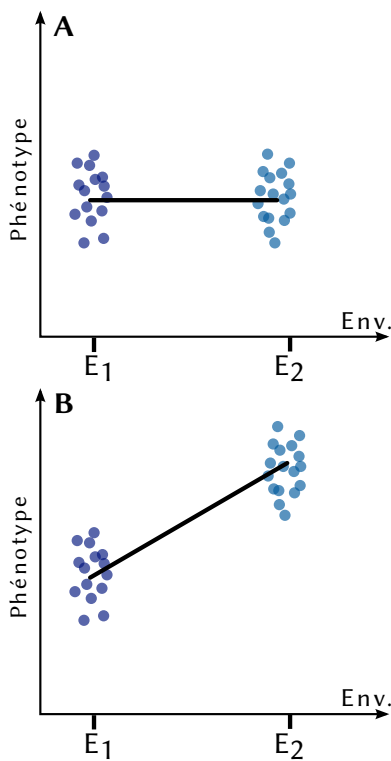


FIG. 2 : Différences entre variabilité *aléatoire* et *systématique* d'un phénotype sous l'influence de l'environnement. Ici, le même génotype est exposé à deux environnements ( $E_1$  et  $E_2$ ). Si l'influence de l'environnement est totalement aléatoire (A), les phénotypes moyens entre  $E_1$  et  $E_2$  sont égaux (A), sinon, ils diffèrent (B).

la moyenne phénotypique calculée en  $E_1$  et celle calculée en  $E_2$  devraient être égales (FIG. 2, A). Dans le cas contraire, il existe un effet *systématique* de l'environnement (FIG. 2, B). On parle alors de *plasticité phénotypique* et la variation systématique du phénotype entre environnements est nommée *norme de réaction*.

Lorsqu'il y a plasticité phénotypique (FIG. 2, B et FIG. 3, A), la valeur moyenne du phénotype dépend donc de l'environnement dans lequel il se trouve et les normes de réaction ne sont pas horizontales. Par exemple, un arbre situé dans une région à fort vent dominant aura tendance à se développer dans le sens de ce vent, alors qu'un arbre tout à fait identique planté en pleine forêt, à l'abri du vent développera un port classique (à symétrie radiale). Notons ici que la plasticité phénotypique ne suppose pas forcément un mécanisme pro-actif de la part de l'organisme, mais peut simplement résider en une limitation en ressources : un individu moins bien nourri que les autres

Le principe de base de la génétique quantitative est donc que l'on peut séparer un phénotype observé  $P$  en deux composantes : l'influence du génotype d'un individu  $G$  et celle de son environnement  $E$  :

$$P = G + E. \quad (1)$$

La composante  $G$  rassemble l'influence de tous les gènes (ainsi que leurs allèles et l'interaction entre toutes ces composantes) qui participent à la variabilité du phénotype. La composante  $E$ , elle, rassemble toute variable environnementale qui participe à cette variabilité. L'équation 1 peut paraître simpliste, mais, comme expliqué ci-dessus, le modèle infinitésimal sous-jacent est en réalité très précis et plutôt complexe. Il prédit par ailleurs que la distribution de  $P$ ,  $G$  et  $E$  devrait être normalement distribuée avec des variances respectives  $V_P$ ,  $V_G$  et  $V_E$ . Cette décomposition de la variance phénotypique  $V_P$  en plusieurs composantes interviendra régulièrement au sein de ce mémoire.

L'intérêt de la génétique quantitative dans l'étude de l'adaptation locale est que la grande majorité des traits adaptatifs sont des traits complexes. Il est donc nécessaire de faire appel à ce cadre théorique si l'on veut explorer les bases génétiques de ce type de traits. Ainsi, de nombreux outils, que nous allons explorer par la suite, ont été développés à partir de ceux existant en génétique quantitative.

### ► Plasticité phénotypique, interaction $G \times E$

Considérons des individus partageant le même génotype (des clones, par exemple) placés dans des environnements contrastés  $E_1$  et  $E_2$ . Si l'effet de l'environnement est purement *aléatoire* (micro-variations entre individus par exemple), alors la

aura tendance à être plus petit, par exemple. Seul compte l'aspect systématique de l'influence de l'environnement.

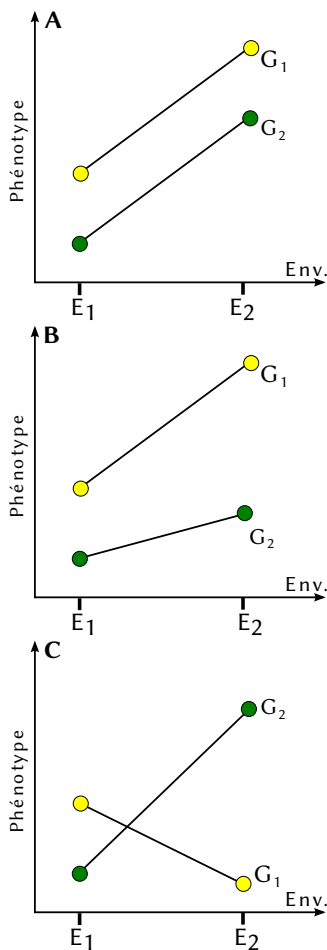


FIG. 3 : Normes de réactions pour 2 génotypes ( $G_1$  et  $G_2$ ) dans deux environnements ( $E_1$  et  $E_2$ ). La plasticité phénotypique pure (A) entraîne la même réponse pour tous les phénotypes (normes parallèles). Une interaction  $G \times E$  donne des normes divergentes (B) ou même croisées (C).

soyons intéressés à comparer un phénotype adaptatif quelconque entre la population A (environnement "bleu") et la population D (environnement "vert") : un échantillonnage *in situ* montre que le phénotype est différent entre ces deux populations, peut-on en déduire qu'il y a adaptation locale ? La réponse, évidemment, dépendra de l'existence ou non d'une plasticité phénotypique sur le trait en question. Si l'on s'attend à ce qu'elle existe, au moins partiellement, la majeure partie du temps, alors on voit bien qu'une différence phénotypique entre A et D ne peut pas être considérée comme

Un phénomène apparenté, mais plus complexe que la plasticité phénotypique apparaît lorsque l'influence systématique de l'environnement dépend du génotype des individus. C'est ce qu'on appelle l'interaction génotype-par-environnement ou  $G \times E$ . Tout se passe alors comme si nous levions une hypothèse dans l'équation 1 en admettant que l'effet de l'environnement puisse être différent en fonction du génotype des individus (et vice-versa) :

$$P = G + E + G \times E. \quad (2)$$

Dans le cas de la plasticité phénotypique "seule", notons que les normes de réactions ne sont pas horizontales, mais sont parallèles (FIG. 3, A). Dans le cas de l'interaction  $G \times E$ , ces normes de réaction ne sont plus parallèles (FIG. 3, B), voire même elles se croisent (FIG. 3, C, c'est-à-dire que la "hiérarchie" entre les génotypes n'est plus respectée).

Dans la littérature, comme à certains endroits dans la suite de cette thèse de doctorat, le terme "plasticité phénotypique" est abusivement utilisé pour englober tout phénomène où l'environnement possède une influence systématique sur l'expression d'un phénotype, qu'il y ait interaction avec le génotype ou non. Ainsi, lorsque le terme "plasticité phénotypique" sera utilisé seul dans un contexte (ou précisé comme étant "au sens large"), il englobera les deux phénomènes, et n'englobera que sa définition stricte lorsqu'il sera explicitement différencié de l'interaction  $G \times E$ .

### ► Plasticité phénotypique et adaptation locale

La plasticité phénotypique (au sens large, donc) est un phénomène omniprésent lorsque l'on étudie l'adaptation locale<sup>1</sup>. En effet, par définition, cette étude porte sur des populations se trouvant dans des environnements contrastés, ce qui rend la comparaison directe des phénotypes entre populations impossible. En effet, en utilisant la FIG. 1, imaginons par exemple que nous

<sup>1</sup>De manière intéressante, l'adaptation locale peut elle-même être considérée comme une interaction  $G \times E$  sur la valeur sélective (KAWECKI & EBERT, 2004) ! C'est-à-dire qu'un "génotype" issu d'une population donnée aura une meilleure valeur sélective dans son environnement local que les "génotypes" immigrés. Si l'on utilise la valeur sélective comme phénotype, on voit qu'on se retrouve dans le cas de la FIG. 3, C, avec des normes de réaction croisées.

une preuve solide d'adaptation locale. Pire, les effets de l'adaptation locale et de la plasticité phénotypique sont totalement confondus au sein de la méta-population, et il est virtuellement impossible de séparer l'un de l'autre sans une approche expérimentale, que je détaillerai dans la section suivante.

## 2 Méthodologies associées à l'étude empirique de l'adaptation locale

Je détaillerai dans cette section la diversité des approches expérimentales, moléculaires et statistiques disponibles pour étudier l'adaptation locale, et les bases génétiques des traits adaptatifs. L'ensemble de ces méthodes fera l'objet d'un examen plus approfondi dans la première partie de cette thèse de doctorat, et la complémentarité entre ces méthodes sera souligné lors de l'étude empirique qui fait l'objet de la seconde partie. Il s'agit donc ici plutôt de présenter succinctement l'état de l'art en ce qui concerne cette méthodologie, les problèmes qu'elle permet de surmonter et ceux dont elle souffre encore.

### 2.1 Méthodologie expérimentale

Comme expliqué ci-dessus, une approche expérimentale est nécessaire lorsque l'on s'intéresse aux traits adaptatifs, et spécifiquement à leur comparaison entre populations. Pour rappel, le problème vient du fait que la plasticité phénotypique est indissociable de l'adaptation locale lorsque l'on mesure les phénotypes *in situ*. À cause de cet effet confusionnel<sup>1</sup> de l'environnement<sup>2</sup>, il est impossible d'étudier les bases génétiques des traits adaptatifs mesurés dans des populations différentes.

Une approche expérimentale, nommée *jardin commun*, permet cependant de compenser ce problème en plaçant des individus (idéalement des zygotes ou des juvéniles) issus de populations différentes dans le même environnement. S'il y a plasticité phénotypique au sens strict, on s'attend à ce qu'elle s'applique de la même manière à tous les individus. Ainsi, toute différence systématique observée entre populations sur le phénotype étudié serait nécessairement d'origine génétique. Ce point est valide même en présence d'interaction G×E, puisque tous les individus sont dans le même environnement. Dans ce cas, la présence d'interaction G×E n'est plus tout à fait un effet confusionnel, mais limite simplement la portée de l'analyse dans un environnement particulier : un réplicat de l'expérience dans un autre environnement pourrait donner des résultats quantitativement et qualitativement différents. Il peut donc être intéressant d'effectuer plusieurs expériences de jardin commun dans des environnements différents.

La meilleure preuve qu'il existe une adaptation locale, quels que soient les traits phénotypiques concernés, est une expérience apparentée au jardin commun, et nommée *transplantation réciproque* (KAWECKI & EBERT, 2004). L'idée de cette expérience est de faire croître, dans plusieurs environnements de la méta-population, des individus locaux et des individus provenant d'autres populations. S'il y a adaptation locale, on s'attend à ce que, au moins dans certains environnements, les individus locaux possèdent une meilleure valeur sélective que les individus immigrés.

Le principal intérêt de l'expérience de jardin commun est qu'elle permet d'étudier la différenciation génétique entre populations pour une série de traits (potentiellement) adaptatifs. Pour ce faire, encore une fois, il est nécessaire d'utiliser le cadre théorique de la génétique quantitative. Ici, le but est

<sup>1</sup>C'est à dire qu'il est confondu avec l'effet que nous voulons tester et donc "brouille" toute analyse des données.

<sup>2</sup>Notons ici que si l'effet de l'environnement était *aléatoire* plutôt que *systématique* (c'est à dire s'il n'y avait pas de plasticité phénotypique), alors il ne serait plus confusionnel !

de décomposer la variance phénotypique en 3 principaux composants : la variance génétique entre populations  $V_{\text{pop}}$ , la variance dite "génétique additive"<sup>1</sup>  $V_A$  au sein des populations et la variance résiduelle  $V_R$  (qui renferme toute la variance génétique non-additive et environnementale).

Le calcul de  $V_{\text{pop}}$  est assez simple, puisqu'il s'agit de la variance des phénotypes moyens entre populations. Rappelons en effet que, grâce à l'environnement commun, toute différence systématique observée entre population doit être d'origine génétique. Ce n'est malheureusement pas aussi simple pour le calcul de  $V_A$ , pour lequel il est nécessaire d'utiliser un protocole expérimental particulier. Ce paramètre ne peut être mesuré qu'en présence d'apparentés (notamment frères/sœurs, ou demi-frères/demi-sœurs) dans le jardin expérimental. L'utilisation de l'apparentement entre individus permet alors d'isoler la composante génétique de la variance phénotypique intra-population (FALCONER & T. F. MACKAY, 1996 ; LYNCH & WALSH, 1998), puisque le lien statistique entre le phénotype de deux individus apparentés passe par les effets additifs des gènes qu'ils ont en commun<sup>2</sup>. Afin de contrôler l'apparentement entre individus dans les expériences de jardin commun, il est courant d'utiliser des descendants de plusieurs parents de chaque population d'origine.

La part relative de chacun de ces trois composants de la variance nous renseigne sur la variabilité génétique observée entre et au sein des populations. À ce titre, plusieurs indices peuvent être calculés à partir de chacun de ces composants. L'héritabilité,  $h^2$ , mesure la proportion de la variabilité d'origine génétique additive (c'est à dire transmissible) au sein des populations :

$$h^2 = \frac{V_A}{V_A + V_R}. \quad (3)$$

C'est un paramètre clé en génétique quantitative, puisque sa valeur possède une influence très forte sur la dynamique évolutive d'un trait soumis à sélection. En effet, si l'héritabilité d'un trait est proche de 0 (variabilité génétique transmissible quasiment nulle), alors toute pression sélective sur ce trait restera sans effet aux générations suivantes. Au contraire, plus la valeur de l'héritabilité sera forte, plus la sélection sera efficace pour faire évoluer le trait en question. Dans le contexte d'une méta-population, il mesure la proportion de la variabilité intra-population qui est d'origine génétique, et ainsi donc, le potentiel évolutif de ce trait dans un contexte sélectif.

Le second indice d'intérêt dans le contexte d'une expérience de jardin commun est la  $Q_{\text{ST}}$  (SPITZE, 1993). Il s'agit d'un équivalent de l'indice de différenciation génétique entre population, la  $F_{\text{ST}}$ , mentionné plus haut, mais calculé à l'aide de données phénotypiques plutôt que de fréquences alléliques. Le principe du calcul, pourtant, reste le même, il s'agit de la variabilité génétique entre populations ramené à la variabilité génétique totale dans la méta-population :

$$Q_{\text{ST}} = \frac{V_{\text{pop}}}{V_{\text{pop}} + 2V_A}. \quad (4)$$

Le 2 au dénominateur de cette équation vient du fait que nous nous intéressons ici à une espèce diploïde. La  $Q_{\text{ST}}$  s'interprète donc comme la  $F_{\text{ST}}$ , c'est à dire qu'elle mesure simplement la différenciation génétique entre populations (pour le trait étudié, et uniquement pour celui-ci, cependant). En fait, s'il n'existe aucune pression de sélection particulière sur le trait en question, on s'attend à ce que la  $Q_{\text{ST}}$  soit égale à la  $F_{\text{ST}}$  (mesurée, par exemple, à partir de données moléculaires).

<sup>1</sup>La variance génétique additive est la part de la variance génétique qui est transmise de générations en générations au sein d'une population.

<sup>2</sup>En absence d'effets de confusion, bien sûr.

La définition de ces indices, ainsi que leur interprétation, sera revue au CHAPITRE 3 de cette thèse. Il sera juste nécessaire ici de mentionner que ces indices, notamment la  $Q_{ST}$ , sont extrêmement utilisés pour étudier l'adaptation locale. Notamment, la comparaison entre  $Q_{ST}$  et  $F_{ST}$  (sur laquelle nous reviendrons amplement par la suite) est un test relativement répandu pour savoir si un trait donné est soumis à une adaptation locale. L'idée est que si la différenciation génétique au niveau du trait (mesurée par la  $Q_{ST}$ ) est significativement supérieure à la différenciation génétique neutre (mesurée par la  $F_{ST}$ , à partir de marqueurs moléculaires, voir ci-dessous), alors la scénario explicatif le plus vraisemblable est celui de l'adaptation locale.

## 2.2 Méthodologie moléculaire

### ► Notion de marqueurs moléculaires, génotypage haut-débit

Un marqueur moléculaire est le produit d'un protocole expérimental de biologie moléculaire visant à capter du polymorphisme dans la séquence ADN des individus. Le type de marqueur obtenu dépend fortement du protocole utilisé. En ce qui concerne l'étude de l'adaptation locale, il existe trois types de marqueurs moléculaires principalement utilisés dans la littérature (dont les propriétés sont résumées en TAB. 1) :

- Les **bandes AFLP** (Vos et al., 1995), issues du protocole nommé *Amplified Fragment Length Polymorphism* qui consiste à digérer de l'ADN par des enzymes de restriction, puis à amplifier les fragments obtenus grâce à la PCR<sup>1</sup>. Le polymorphisme est enregistré comme une présence/absence de fragment d'une taille donnée, correspondant à une présence/absence de mutation au site de restriction. Pour un hétérozygote (dont seul un des sites de restriction est muté), il y a tout de même présence d'une bande : le marqueur est dit *dominant*.
- Les **micro-satellites**, qui sont des séquences ADN caractérisées par une répétition du même motif un nombre variable de fois. Le polymorphisme détecté est formé par les variations du nombre de répétitions entre individus. Le taux de mutation de ce type de marqueurs est particulièrement élevé.
- Les **SNP** (*Single Nuclear Polymorphism*), qui consistent en un polymorphisme au niveau d'un locus nucléotidique spécifique (un A remplacé par un T dans la séquence ADN, par exemple). Tout protocole basé sur du séquençage permet d'extraire cette information, donc de nombreux protocoles produisent ce type de marqueurs. De fait, il s'agit d'un marqueur moléculaire standard de référence.

**TAB. 1** : Différents types de marqueurs moléculaires et certaines de leurs caractéristiques (dominance, nombre moyen de locus différents) et des protocoles qui permettent de les produire. Les nombres de ce tableau sont des ordres de grandeur moyens.

Type de marqueurs	Dominant	Taux de mutation	Nombre d'allèles	Nombre de locus	Protocoles
Bandes AFLP	Oui	$10^{-6}$	2	100	AFLP
Micro-satellites	Non	$10^{-2} - 10^{-5}$	5 - 30	10	Micro-satellites
SNP	Non	$10^{-6}$	2	$100 - 10^6$	Séquençage complet/ciblé, GBS <sup>2</sup> , RADseq <sup>3</sup> , puces à ADN
Haplotype RADseq	Non	$10^{-5} - 10^{-6}$	2 - 8	$10^3 - 10^5$	RADseq <sup>3</sup>

<sup>1</sup>Polymerase Chain Reaction.

<sup>2</sup>Genotyping By Sequencing, ELSHIRE et al. (2011).

<sup>3</sup>voir section suivante

Outre les caractéristiques du protocole utilisé, les caractéristiques d'un marqueur peuvent aussi dépendre de sa localisation dans le génome. Ainsi, lorsqu'un marqueur sera dans une région non soumise à la sélection naturelle (souvent, des régions intergéniques), on parlera de *marqueur neutre*. On considère, notamment depuis les travaux sur la théorie neutraliste de KIMURA (1985), que la majorité des marqueurs pris aléatoirement dans un génome sont neutres.

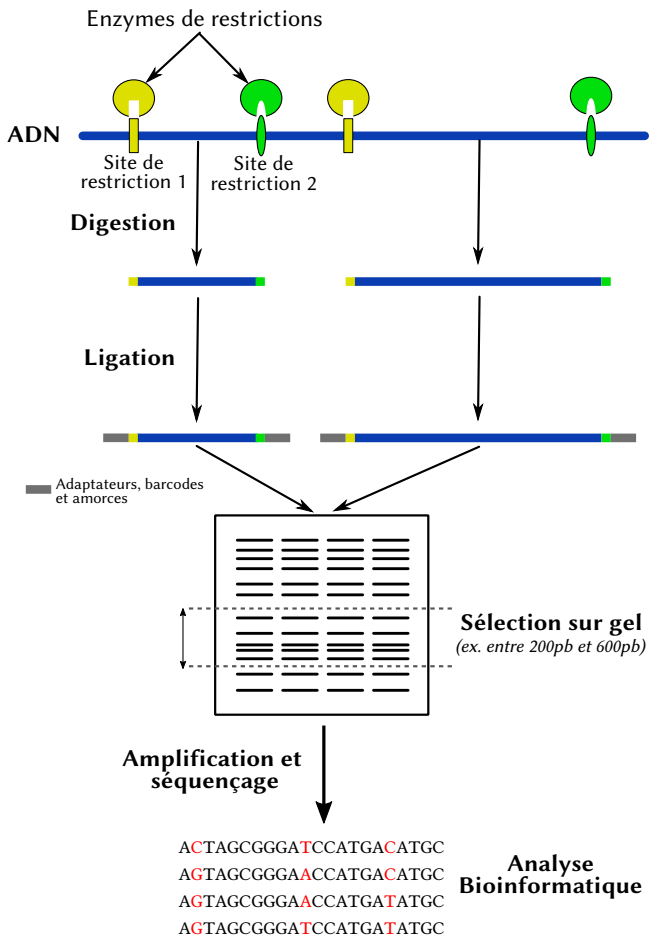


FIG. 4 : Schéma représentant le principe du protocole *double digest RAD sequencing*. L'ADN est d'abord digéré par deux enzymes de restriction (en jaune et en vert). Après une étape de ligation d'adaptateurs contenant des barcodes et des amorces choisis, les fragments sont sélectionnés sur gel d'électrophorèse puis amplifiés. Ils sont ensuite envoyés à séquençer.

séquencer qu'une fraction aléatoire du génome des individus : dans le cas du RADseq, il s'agit de digérer l'ADN à l'aide d'une ou deux enzymes de restriction (dans ce cas, on parle de *double digest RADseq*), et de séquençer les régions flanquantes des sites de restriction. Afin de limiter la quantité d'ADN nécessaire, le protocole RADseq ajoute une étape de sélection sur la taille des fragments, suivie d'une PCR (voir FIG. 4) Il s'agit d'un protocole de génotypage haut-débit qui peut produire de l'ordre de la dizaine de milliers de marqueurs.

Ce protocole possède de nombreuses propriétés intéressantes. Tout d'abord, à condition d'externaliser l'étape de séquençage, il ne requiert que des équipements classiques de biologie moléculaire. Ensuite, puisqu'une petite fraction du génome est séquençée, le coût du séquençage par individu

Le paysage des marqueurs moléculaires a considérablement changé depuis une dizaine d'année : à cause de la très forte chute des coûts et du temps de séquençage, les SNP se sont définitivement installés comme des marqueurs de références, et le nombre de SNP utilisés dans les études récentes a considérablement augmenté durant ce temps (d'une centaine à plusieurs millions de SNP). Ce changement au niveau des marqueurs moléculaires a, en retour, considérablement changé les pratiques. Le séquençage de génome complet, ou les protocoles de génotypage haut-débit (qui seront plus abondamment abordés dans la suite de cette thèse), ont permis de passer d'études génétiques sur quelques locus à des études dites *génomiques*, utilisant des marqueurs densément distribués au sein des génomes. Nous verrons plus loin l'impact de ces changements sur l'étude de l'adaptation locale, mais tout d'abord intéressons-nous à un protocole spécifique dont il sera beaucoup question au cours de ce mémoire.

### ► Le protocole RADseq

Le RADseq (*Random Amplified DNA sequencing*, MILLER et al., 2007 ; BAIRD et al., 2008) est un protocole spécifique qui appartient à la famille des protocoles de "séquençage d'une représentation du génome". Le principe de ce type de protocole est de ne

est considérablement diminué. Cette fraction, en outre, se situe dans des régions certes aléatoires, mais extrêmement répétables entre individus, ce qui simplifie considérablement les analyses bio-informatiques (CATCHEN et al., 2011; CATCHEN et al., 2013). Enfin, puisqu'il s'agit d'un protocole utilisant du séquençage, ce protocole produit des SNP, qui ne sont pas dominants (contrairement aux bandes AFLP).

Comme indiqué dans le TAB. 1, il est possible d'obtenir à partir du protocole RADseq un type de marqueurs que nous appellerons ici "haplotype RADseq". Il s'agit en fait d'utiliser ensemble tous les SNP qui appartiennent à la portion lue lors du séquençage d'un fragment. À partir de ces lectures, il est possible de lier ces SNP le long de très courts haplotypes (125pb). Ces mini-haplotypes forment ainsi des allèles, dont le polymorphisme peut être étudié. Alors que les marqueurs SNP ont généralement un polymorphisme de 2 allèles, ces marqueurs haplotypiques permettent de combiner l'information de plusieurs SNP pour arriver à un polymorphisme entre 2 et 8 allèles (TAB. 1). Cela permet donc de tirer profit au maximum de l'information générée par le protocole RADseq en limitant les problèmes de liaison entre marqueurs.

### ► Les marqueurs moléculaires pour l'étude de l'adaptation locale

L'ensemble des marqueurs moléculaires ci-dessus sont abondamment utilisés dans le domaine de la micro-évolution. Puisque les fréquences alléliques de ces marqueurs sont généralement fortement affectées par les processus stochastiques cités en section 1.1, leur étude permet des inférences sur les événements micro-évolutifs passés de la méta-population (S. WRIGHT, 1949; WEIR & COCKERHAM, 1984; PRITCHARD et al., 2000; CORNUET et al., 2008). Plus le nombre de marqueurs est important, plus la précision (et possiblement l'exactitude !) de ces inférences est grande. L'information obtenue sur la dynamique de la méta-population est extrêmement importante, puisqu'il s'agit de la toile de fond sur laquelle se reposent les inférences sur l'adaptation locale.

Comme abordé plus haut (sections 1.1 et 2.1), un indice extrêmement utilisé pour résumer les propriétés démographiques et historiques d'une population est la  $F_{ST}$ . Celle-ci est calculée à l'aide des fréquences alléliques d'un ou plusieurs locus, pour un type de marqueurs moléculaires donné. Pour rappel, plus cet indice est fort, plus la différenciation génétique entre populations est importante. D'autres approches, dont certaines<sup>1</sup> seront détaillées dans la section suivante et dans la première partie de cette thèse permettent d'affiner encore les inférences sur les processus à l'œuvre dans la méta-population.

Afin que ces inférences sur les processus historiques et démographiques soient valides, il est nécessaire que les marqueurs utilisés soient neutres (voir ci-dessus), c'est à dire qu'ils ne soient pas affectés par les processus de sélection naturelle qui agissent sur leurs fréquences alléliques.

Bien que l'apport des marqueurs moléculaires dans l'étude des processus neutres soit primordiale, leur utilité ne se limite pas à cela. Lorsqu'ils sont en densité suffisante dans le génome (donc issus d'un génotypage haut-débit), ils permettent d'étudier plus précisément les bases génétiques de l'adaptation locale. L'intérêt se porte alors non pas sur les locus neutres, mais précisément sur la détection des locus impliqués dans le processus de l'adaptation locale. On parle alors de *scans génomiques*, et les méthodes statistiques associées seront abordées dans la section suivante.

---

<sup>1</sup>La liste des méthodes abordées ici n'est pas exhaustive des méthodes disponibles pour étudier les processus neutres d'une méta-population, tant la diversité de ces méthodes reflète celle des processus existants.

## 2.3 Méthodologie statistique

Il existe *grosso modo* deux grands types d'approches auxquelles le terme *scan génomique* se réfère. Les premières, les scans génomiques de sélection, tentent de détecter les locus, au sein du génome, qui pourraient être soumis à la sélection naturelle, sans aucune hypothèse concernant le phénotype soumis à la sélection. Les secondes approches, que l'on pourrait appeler ici scans d'association (plus connues sous le nom de *genome wide association studies*, ou GWAS dans la littérature), partent de phénotypes identifiés comme adaptatifs, et tentent de détecter les locus responsables de la variabilité de ces phénotypes. Des méthodes à la frontière entre ces types d'approches commencent à apparaître (BERG & COOP, 2014).

Dans tous les cas, les méthodes de scan génomique exploitent le phénomène de *déséquilibre de liaison*, qui veut que les événements de recombinaison entre deux positions proches au sein du génome sont moins fréquentes qu'entre deux positions éloignées. Ainsi, les marqueurs moléculaires proches du locus recherché (soit parce qu'il est sous sélection, soit parce qu'il a un effet sur la variabilité du phénotype étudié) renvoient un signal similaire à celui-ci (mais atténué par la recombinaison, d'autant plus que le marqueur est loin du locus). La détection d'un signal de sélection est donc maximisée par l'étendue du déséquilibre de liaison et la densité des marqueurs dans le génome. En effet, même si le locus sous sélection ne fait pas parti du panel des marqueurs, on maximise ainsi la chance d'avoir tout de même un marqueur en déséquilibre de liaison avec ce locus sous sélection.

### ► Scan génomique de sélection

Puisque les méthodes de scan génomique de sélection n'utilisent pas de données phénotypiques, elles ne comptent que sur l'analyse des données de marqueurs moléculaires, ainsi que, parfois, de données environnementales. Lorsqu'il y a sélection à un endroit particulier du génome, son effet se fait ressentir dans les régions alentours à cause du déséquilibre de liaison : notamment, l'augmentation en fréquence de l'allèle sélectionné du locus entraîne une augmentation de la fréquences des allèles des marqueurs moléculaires auxquels il est lié par un effet d'*autostop génétique*. C'est cet effet sur les fréquences des marqueurs moléculaires que cherchent à détecter les méthodes de scan génomique de sélection.

Le principe commun à toutes les méthodes de scan génomique est qu'en cas d'adaptation locale, les fréquences alléliques des locus sélectionnés (et donc des marqueurs proches) vont être différentes dans différentes populations et environnements.

### ► Scan d'association

Les méthodes d'association des GWAS cherchent à détecter des marqueurs proches de locus responsables de la variabilité d'un phénotype ou QTL (pour *Quantitative Trait Locus* en anglais). Il s'agit généralement d'effectuer, pour chaque marqueur, une regression entre le génotypes des individus à ce marqueur et la valeur de leur phénotype. Lorsqu'une de ces régressions est significative, on considère que le marqueur se situe à proximité d'un locus variable génétiquement et impliqué dans l'expression du phénotype en question. Ces méthodes permettent d'étudier l'architecture génétique des traits phénotypiques, notamment ceux qui sont adaptatifs.

### ► Structure de population

La difficulté commune à toutes les méthodes ci-dessus vient de leur sensibilité générale aux *structures de population*. On appelle structure de population l'effet de tout mécanisme qui entraîne une



reproduction non complètement aléatoire des individus : flux de gènes limités ou asymétriques entre populations, histoire démographique, auto-fécondation/consanguinité, choix du partenaire, etc... Tous ces phénomènes créent de la structure dans les données, qu'elles soient phénotypiques ou génétiques. Or cette structure, neutre, peut être très problématique lorsqu'on tente d'isoler des variants génétiques sélectionnés ou impliqués dans un phénotype.

Imaginons par exemple que vous cherchiez à faire une étude d'association sur la couleur des yeux chez l'Homme. Pour comparaison, vous échantillonnez des individus suédois et turques, et utilisez une méthode d'association sur le phénotype "couleur des yeux". Il est fort probable que vous trouviez des marqueurs proches de QTL impliqués dans la couleur des yeux comme significatifs. Seulement ces deux populations sont différentes à la fois pour ce phénotype, mais sur beaucoup d'autres (couleur des cheveux, couleur de peau...) : beaucoup de marqueurs sortiront comme significatifs parce qu'ils sont associés, par exemple, à la couleur de cheveux, non des yeux.

Bien que l'exemple ci-dessus soit relativement caricatural, la structure de population pose des problèmes parfois très subtils. Parce que les populations naturelles sont souvent complexes, et que souvent peu de connaissances sur l'histoire démographique de celles-ci sont disponibles, il est très difficile d'échantillonner des données sans structure. Ceci engendre des problèmes de *faux positifs*, c'est à dire de résultat statistiquement significatifs, mais qui n'appartiennent pas aux signaux biologiques recherchés. Les diverses méthodes statistiques de scan génomique ont donc été développées de manière à prendre cette structure en compte. Pourtant, les problèmes de faux positifs persistent. Ceci est dû à l'extrême diversité des phénomènes et scénarios possibles qui peuvent structurer les données. Une seule méthode ne peut, à elle seule, couvrir l'ensemble de la stochasticité évolutive de ces phénomènes. Il convient alors de bien évaluer les méthodes disponibles, et de choisir la méthode adéquate vis-à-vis de son sujet d'étude.

### 3 Intérêt d'*Arabis alpina* pour l'étude de l'adaptation locale

Nous avons maintenant fait le tour des considérations conceptuelles et méthodologiques qui concernent l'étude de l'adaptation locale. L'objet de cette thèse de doctorat, pour la partie empirique, est l'étude de l'adaptation locale chez la plante alpine *Arabis alpina*. Il convient donc d'introduire l'intérêt du système alpin pour l'étude de l'adaptation locale (et vice-versa) et plus particulièrement celui de la plante que nous allons étudier.

#### 3.1 Le système alpin et l'adaptation locale

Les systèmes alpins sont caractérisés par des variations climatiques locales et régionales très importantes (KÖRNER, 2003). À l'échelle de quelques centaines de mètres, par exemple, les gradients d'altitude peuvent être très importants, et s'accompagnent de forts gradients de température, de précipitations, de couverture neigeuse et de rayonnement solaire. Cette tendance générale est en outre très fortement influencée par des facteurs locaux comme l'exposition (notamment adret ou ubac), la topologie, voire la micro-topologie et la nature du socle rocheux. En outre, les milieux alpins sont caractérisés par une variabilité temporelle importante du milieu : les contrastes entre les températures diurnes et nocturnes sont d'autant plus forts avec l'altitude, ainsi que les écarts saisonniers (KÖRNER, 2003). Ces écarts saisonniers sont cependant généralement tamponnés par la présence d'une couverture neigeuse (KÖRNER, 2003).

En résumé, nous pouvons dire que les milieux alpins sont à la fois caractérisés par une forte contrainte environnementale (qui augmente globalement avec l'altitude) et une variabilité spatio-

temporelle extrême, même à petite échelle. Parce que la contrainte topologique est forte, une grande partie des espèces alpines sont en outre caractérisées par une faible dispersion. Cette relative *viscosité* (bien entendu dépendante de l'espèce en question), entraîne des flux de gènes modérés qui, combinées à la variabilité environnementale du milieu, engendrent des conditions favorables à l'émergence d'une adaptation locale.

### 3.2 La plante alpine *Arabis alpina*

*Arabis alpina*, autrement appelée l'Arabette des Alpes, est une plante arctico-alpine pérenne de la famille des *Brassicacées* qui pousse dans les milieux montagnards, sub-alpins et alpins. Elle possède une aire de répartition relativement large allant du Sud au Nord de l'Europe, ainsi que le Moyen-Orient et l'Afrique du Nord et de l'Est. Sa répartition actuelle résulte probablement d'une émergence de l'espèce au Moyen-Orient il y a moins de 2 millions d'années suivi d'une colonisation progressive notamment de l'Europe il y a environ 500 000 ans (KOCH et al., 2006). En Europe, *A. alpina* se retrouve dans la majeure partie des massifs montagneux (Pyrénées, Alpes, Carpates et Balkans) ainsi que dans les régions arctiques (bouclier Scandinave notamment). Elle vit typiquement dans des milieux ouverts et instables, de type éboulis. Elle est notamment caractéristique du front de colonisation post-glaciation (WHITTAKER, 1993).

Dans les Alpes, sa gamme altitudinale est considérable, de 1000m à 3000m d'altitude, partant ainsi de l'étage montagnard pour arriver à la limite supérieure de l'étage alpin. La question se pose donc de savoir ce qui permet à cette espèce d'affronter des environnements si différents, et notamment si l'adaptation locale n'est pas responsable de cet état de fait.

Le scénario d'adaptation locale est vraisemblable pour plusieurs raisons. Tout d'abord à cause de cette forte gamme altitudinale. Ensuite parce qu'*A. alpina* forme des populations denses mais de tailles modestes. Elle est ensuite caractérisée aussi par une dispersion à courte distance (BUEHLER et al., 2012) : sa pollinisation est zoochore et surtout ses graines ont une dispersion autochore (siliques déhiscentes).

Outre ces caractéristiques, *A. alpina* est un modèle scientifique intéressant. Bien qu'elle soit phylogénétiquement plus proche du chou (*Brassica oleracea*, BAILEY et al., 2006), elle possède une orthologie génétique très forte avec la plante modèle *Arabidopsis thaliana* (LOBRÉAUX et al., 2014). Ceci permet l'utilisation des abondantes ressources disponibles sur *A. thaliana* (notamment les annotations génomiques) pour l'étude d'*A. alpina*. Pour autant, *A. alpina* possède de nombreuses différences avec *A. thaliana*. Il s'agit d'une plante pérenne, ce qui permet l'étude pluri-annuelle du déterminisme de la floraison (R. WANG et al., 2009 ; R. WANG et al., 2011 ; ALBANI et al., 2012). Elle possède aussi des stratégies différentes concernant l'auto-fécondation : les populations du Sud de l'Europe étant complètement allofécondes, tandis que les populations du Nord pratiquent une auto-fécondation partielle (ANSELL et al., 2008 ; TEDDER et al., 2015). Ces caractéristiques ont fait émerger *A. alpina* comme une plante modèle pour l'étude de l'adaptation aux conditions alpines (MANEL et al., 2010 ; PONCET et al., 2010 ; BUEHLER et al., 2012 ; BUEHLER et al., 2013 ; WINGLER et al., 2015 ; TORÄNG et al., 2015), bien que ce type d'analyses ait récemment été faites aussi sur *A. thaliana* (LUO et al., 2015).

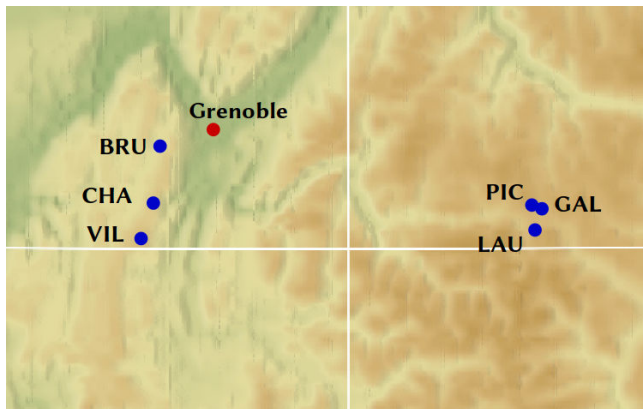


FIG. 5 : La plante alpine *Arabis alpina* ou Arabette des Alpes.

TAB. 2 : Liste des 6 populations d'*Arabis alpina* étudiées depuis 2008.

Site	Code	Massif	Latitude	Longitude	Altitude	Exposition	Milieu
Gorges du Bruyant	BRU	Vercors	45,15065	5,61112	930m	Sud-Ouest	Éboulis calcaire
Combes de Chaulange	CHA	Vercors	45,07117	5,59267	1480m	Nord	Éboulis calcaire, prairie
Côte 2000	VIL	Vercors	45,01809	5,57083	1980m	Sud-Ouest	Rochers, éboulis calcaires
Col du Lautaret	LAU	Lautaret	45,02846	6,39034	2090m	Nord	Ruisseau schisteux
Col du Galibier	GAL	Lautaret	45,06049	6,40375	2590m	Nord	Éboulis calcaire
Pic Blanc	PIC	Lautaret	45,06385	6,38426	2930m	Sud	Schistes

Au sein du Laboratoire d'Écologie Alpine, *A. alpina* est étudié depuis moins d'une dizaine d'années, dans le but de comprendre les mécanismes adaptatifs de cette plante aux conditions de vie alpines. Les études menées sur cette espèce vont des analyses de génétique du paysage à différentes échelles, à l'aide de marqueurs AFLP (MANEL et al., 2010 ; PONCET et al., 2010), à l'étude de la réponse au gel (KOLAKSAZOV et al., 2013 ; KOLAKSAZOV et al., 2014) en passant par le développement de ressources génomiques (LOBRÉAUX et al., 2014). Les travaux précédents ont permis de retrouver des signaux d'adaptation locale (MANEL et al., 2010 ; PONCET et al., 2010), mais ont été limité par le relativement faible nombre de marqueurs utilisés et le manque d'outils d'analyse appropriés, qui sont venus par la suite.

FIG. 6 : Carte des six populations d'*Arabis alpina* étudiées.

Des travaux de suivi démographiques *in situ* portant sur six populations (TAB. 2 et FIG. 6), démarrés en 2008 par Marco Andreollo ont permis d'obtenir des informations intéressantes sur l'écologie d'*A. alpina*. Ces travaux portent sur 3 populations du Vercors (de 900m à 2000m) et 3 populations autour du col du Lautaret (2000m à 3000m). Je présenterai l'analyse de ce suivi démographique dans la second partie de cette thèse. Ces populations ont servi comme populations sources pour nos expériences de jardin commun et d'étude de l'adaptation locale.

## 4 Problématique et plan

L'objectif de cette thèse de doctorat est double. Dans un premier temps, je ferais une analyse des méthodes disponibles pour étudier les problématiques d'adaptation locale. Cette analyse portera principalement sur les analyses de scan génomique pour détecter la sélection, faisant appel à des marqueurs moléculaires denses dans le génome, mais aussi sur l'expérience de jardin commun. Dans un second temps, je présenterai les résultats que nous avons obtenus sur *Arabis alpina*. Je montrerai, entre autre, qu'il existe une adaptation de cette plante à la température moyenne et que cette adaptation se présente sous la forme d'un syndrome phénotypique comprenant la taille et le port des plantes, ainsi que leur croissance, leur reproduction et potentiellement leur survie.

La première partie, qui concerne l'étude méthodologique, se compose de 3 chapitres. Le CHAPITRE 1 s'intéresse à l'évaluation des méthodes de scan génomique de détection de sélection existantes. Il présente une analyse par simulation des performances d'une partie de ces méthodes qui montre

qu'elles sont particulièrement sujettes aux faux positifs lorsque confrontées à des jeux de données complexes. Le CHAPITRE 2 porte sur le développement d'une nouvelle méthode de scan génomique, BayeScEnv, basée sur une méthode existante nommée Bayescan. Cette méthode est basée sur la notion de  $F_{ST}$ , mais utilise en plus une information environnementale. Elle permet d'éviter de nombreux faux positifs, comparée à Bayescan, et de détecter des patrons de fréquences alléliques que les méthodes d'association ne peuvent pas détecter. Enfin le CHAPITRE 3 consiste en une perspective méthodologique reliant les expériences de jardin commun, le génotypage haut-débit et de nouvelles méthodologies statistiques. Elle montre qu'il existe beaucoup de possibilités d'analyse encore sous-exploitées dans le cadre de ces expériences.

La deuxième partie est composée de 2 chapitres, portant sur l'analyse de l'adaptation aux conditions d'altitude chez *Arabis alpina*. Le CHAPITRE 4 porte sur l'analyse des données démographiques récoltées lors du suivi *in situ*. Il présente les analyses de Marco Andrello et moi-même montrant quelques caractéristiques démographiques de cette plante encore inconnues (faible espérance de vie, survie et reproduction extrêmement différentielle entre individus...), ainsi qu'une analyse montrant que la croissance et la survie des plantes dépendent de la température moyenne dans la population locale. Le CHAPITRE 5 montre que ces variations phénotypiques observées en milieu naturel comportent une base génétique adaptative. Notamment, j'y montrerai que la taille et le port des plantes, leur croissance et leur reproduction mesurés en jardin commun dépendent de la température moyenne de la population d'origine. J'y fournis aussi une courte liste de gènes candidats pouvant être impliqués dans ce phénomène d'adaptation locale.

# **PREMIÈRE PARTIE**

---

**Méthodologies statistiques et empiriques liées à  
l'étude de l'adaptation locale**

# CHAPITRE 1

## Étude et comparaison des méthodes de scan génomique pour détecter la sélection

---

Les obstacles méthodologiques en ce qui concerne l'étude de l'adaptation locale ont été présentés dans l'Introduction. Le but de la première partie de cette thèse est de détailler et discuter certains de ces obstacles. Ces discussions vont principalement se concentrer sur les conséquences du passage à l'ère de la génomique des populations (LUIKART et al., 2003), suite aux développements de génotypage haut-débit décrits dans l'Introduction.

Dans ce chapitre, le terme *scan génomique*, sans précision, fera référence aux méthodes de *scan génomique de sélection* décrites dans l'Introduction. Pour rappel, ces méthodes tentent de détecter des locus du génome qui sont, ou ont été, sous sélection.

### 1 Le problème des scans génomiques de sélection

L'idée de tester si certains locus sont sous sélection différentielle entre populations (donc liés à une adaptation locale) est assez ancienne et peut être retracée au moins jusqu'à LEWONTIN & KRAKAUER (1973) qui proposèrent l'idée de comparer les valeurs de  $F_{ST}$  estimées pour plusieurs locus à une distribution théorique de ces valeurs sous un modèle neutre. Détecter ces locus sélectionnés est un objectif alléchant pour le biologiste évolutif, étant donné l'intérêt que ces analyses peuvent avoir pour étudier les bases génétiques de l'adaptation. Très rapidement pourtant, la robustesse et la généralité du modèle neutre proposé par LEWONTIN & KRAKAUER ont été remises en cause (NEI & MARUYAMA, 1975), des effets biologiques pouvant générer des faux positifs lors de ce test.

En effet, comme tout test statistique ne renseigne que sur la plausibilité d'une hypothèse (nommée hypothèse nulle), ces tests ne permettent que de rejeter une hypothèse de neutralité. Pour pouvoir conclure que le signal détecté est dû à un phénomène d'adaptation locale, il est donc crucial que cette hypothèse de neutralité soit correctement définie, et reflète effectivement la démographie historique de l'espèce, ainsi que ses caractéristiques génétiques. Toute déviation à ce modèle, même neutre, est susceptible de générer des faux positifs (des locus significatifs, mais pourtant neutres). Or, la diversité des histoires démographiques, ainsi que des caractéristiques écologiques et génomiques, des espèces étudiées est telle que la recherche d'un modèle neutre unique capable de toutes les modéliser est nécessairement une quête perdue d'avance.

Ces débats ont mené au développement de nouveaux modèles pour prendre en compte au mieux possible la structure de population, afin de réduire les faux positifs (BALDING & NICHOLS, 1995 ;

BEAUMONT & NICHOLS, 1996 ; NICHOLSON et al., 2002). Parallèlement, l'avènement du séquençage haut-débit a augmenté l'intérêt de ces approches en permettant un scan de plus en plus exhaustif du génome pour détecter des locus sous sélection. À tel point que la logique a été complètement renversée : l'idée n'étant plus de tester si un locus particulier est sous sélection, mais directement de détecter les locus qui peuvent être soumis à sélection à l'aide de données génomiques.

## 2 Une nouvelle génération de méthodes de scan génomique

Depuis les années 1990, un grand nombre de méthodes de scan génomique ont émergé suite à de nombreux efforts pour développer des modèles de génétique des populations assez robustes pour servir de modèles neutres.

Les méthodes de scan génomique de sélection se divisent globalement en deux catégories. La première utilise le concept de  $F_{ST}$ , c'est à dire qu'elle cherche à détecter les locus des marqueurs moléculaires qui sont fortement différenciés entre populations, ou plus précisément, qui sont plus différenciés que ne l'est le reste du génome (supposé neutre). La seconde cherche à corrélérer les fréquences alléliques des marqueurs à une variable environnementale. Une relation significative entre ces deux quantités serait le témoin d'une sélection différentielle entre populations.

### 2.1 Les méthodes à $F_{ST}$

Une des premières méthodes assez populaire et utilisant le concept de  $F_{ST}$  a été FDIST (BEAUMONT & NICHOLS, 1996). Cependant, cette méthode souffrait de quelques limitations concernant la rigidité du modèle utilisé, ainsi que dans le calcul de la distribution des  $F_{ST}$ . Un modèle de vraisemblance développé par BEAUMONT & BALDING (2004) et plus tard conceptualisé sous le terme de *modèle F* par GAGGIOTTI & FOLL (2010) a permis l'émergence de nouvelles méthodes plus robustes et plus efficaces, dont les méthodes Bayescan (FOLL & GAGGIOTTI, 2008) et SelEstim (VITALIS et al., 2014). De nouveaux développements ont aussi été fait directement à partir des travaux de LEWONTIN & KRAKAUER (1973), mais en corrigeant plus finement la structure de population (BONHOMME et al., 2010). Enfin, les derniers développements méthodologiques sur la détection de différenciation génétique entre population abandonnent même la notion de  $F_{ST}$  pour utiliser l'analyse en composante principale (PCAdapt, DUFORET-FREBOURG et al., 2014).

Dans l'article ci-dessous, seule la méthode Bayescan représente cette famille des méthodes à  $F_{ST}$ , pour une raison simple : elle a déjà été montrée comme plus performante que FDIST (VILAS et al., 2012) et les autres méthodes plus récentes n'étaient pas publiées lorsque notre étude a démarré en 2012.

### 2.2 Les méthodes d'association environnementale

Une des premières méthodes d'association qui possède un modèle d'évolution neutre est basée sur le modèle de NICHOLSON et al. (2002), qui approxime l'évolution neutre des fréquences alléliques par une loi Normale tronquée en 0 et 1. Cette approximation permet de calculer une matrice de variance-covariance entre populations. Cette méthode, BayEnv (COOP et al., 2010), teste donc si, en prenant en compte de cette matrice de variance-covariance, l'effet de l'environnement sur les fréquences alléliques est significatif, ce qui est ensuite interprété comme un signe d'adaptation locale. Cette approche a ensuite été raffinée par GAUTIER (2015). Certaines méthodes n'utilisent pas de modèles explicites d'évolution neutre, mais corrigent statistiquement pour la structure de population : c'est

le cas de la méthode LFMM (FRICHOT et al., 2013), qui utilise une approximation de la structure des génotypes à l'ordre  $K$  (voir plus de détails dans l'article ci-dessous) et de la méthode gINLAnd (GUILLOT et al., 2014), qui modélise la structure de population par une auto-régression spatiale.

Ces méthodes d'association environnementale ne seront représentées dans l'article suivant que par BayEnv et LFMM qui étaient les principales méthodes disponibles en 2012.

## 2.3 Le point commun entre ces nouvelles méthodes

Malgré la très forte diversité des méthodes présentées ci-dessus, celles-ci possèdent un point commun fondamental : elles supposent toutes que la vaste majorité des locus génotypés suivent le même modèle de neutralité. L'ensemble des locus est donc utilisé pour adapter un modèle neutre, contre lequel est testé une hypothèse alternative (généralement d'adaptation locale). Ce mode de raisonnement s'appuie sur l'idée que les événements évolutifs neutres (principalement la démographie) impactent l'ensemble des locus de la même façon.

Ainsi, pour que les méthodes de scan génomique fonctionnent correctement, deux conditions doivent être validées. La première, c'est qu'il existe un nombre négligeable de locus sous sélection, ou qu'il soit possible, en utilisant une source d'information indépendante (par exemple, une annotation du génome, ou la différence entre mutation synonyme et non-synonyme) d'isoler un ensemble de locus nécessairement neutres. La seconde, c'est que le modèle neutre puisse effectivement s'adapter aux données : c'est à dire qu'il soit capable de prendre en compte les particularités de la structure de populations, ou à défaut, qu'il y soit robuste.

Si la première condition est facilement vérifiable pour certaines espèces pour lesquelles on possède les ressources génomiques suffisantes, elle peut être difficile à vérifier dans le cas d'espèces non modèles. La seconde condition est extrêmement difficile à vérifier sans une étude par simulation. C'est ce que se propose de faire l'article ci-dessous : vérifier la robustesse de ces méthodes à des modèles de démographie dont on s'attend à ce qu'ils les mettent en difficulté.

## 3 Tests multiples et taux d'erreur

### 3.1 Classification des résultats d'un test statistique

Lorsque l'on effectue un test statistique, celui-ci ne peut avoir que deux résultats : le test est soit significatif, soit il ne l'est pas. Dans le cas d'un test de sélection, on peut considérer que le locus peut avoir deux statuts possibles : soit il est (ou a été) soumis à sélection, soit il est neutre. À l'aide de ces deux critères, nous pouvons classer le résultat d'un test statistique en 4 catégories, comme indiqué dans le TAB. 3 ci-contre. Bien sûr, le but de tout test statistique est de tendre vers le maximum de vrai positifs et de vrai négatifs.

TAB. 3 : Classification des résultats d'un test statistique

	Significatif	Non Significatif
Neutre	Faux Positif	Vrai Négatif
Sélectionné	Vrai Positif	Faux Négatif

Cependant, il existe un compromis fondamental entre la quantité de vrai positifs et de faux positifs. En effet, à cause des fluctuations stochastiques, le risque de faux positif est toujours présent. En d'autres termes, l'hypothèse nulle (neutralité) contient nécessairement des configurations telles que la relation recherchée, par exemple une relation linéaire entre l'environnement et les fréquences alléliques apparaisse comme vraisemblable. Puisqu'il est impossible d'éliminer complètement les



faux positifs, la question se pose en terme de quantité acceptable. Éliminer les faux positifs avec certitude revient en gros à éliminer aussi tous les vrai positifs, tandis qu'accepter le plus de vrai positifs possible peut revenir à être submergé de faux positifs.

Ce compromis complexe, fondamental dans le test d'hypothèse en statistiques, est généralement contrôlé par deux quantités : la *puissance* du test (son taux de vrai positif) et son *taux d'erreur*.

### 3.2 Taux de faux positif, taux de fausse découverte

Le taux d'erreur d'un test peut être contrôlé à partir de deux perspectives différentes selon que l'on veuille contrôler l'erreur d'un seul test ou d'une série entière de tests.

**Taux de faux positif** Le taux de faux positif (TFP) de notre test de sélection est la probabilité qu'un locus neutre soit considéré comme significatif par le test. En terme de la classification en TAB. 3, c'est le rapport entre le nombre de faux positifs et le nombre de locus neutres :

$$\text{TFP} = \frac{\text{FP}}{\text{FP} + \text{VN}} \quad (1.1)$$

Lorsque l'on calcule la  $p$ -valeur d'un test statistique, la distribution attendue de celle-ci sous l'hypothèse nulle (ici, que le locus soit neutre) est uniforme par construction. Cette propriété fait que la  $p$ -valeur est utilisé pour contrôler le taux de faux positif : sous l'hypothèse de neutralité (et en considérant le modèle correctement calibré), seulement 5% des tests doivent avoir une  $p$ -valeur inférieure à 5%.

**Taux de fausse découverte** Le taux de fausse découverte (TFD) se définit pour un ensemble de tests, et non un seul test. C'est la probabilité qu'un locus significatif soit en fait neutre. En terme de la classification en TAB. 3, il se définit comme le rapport entre le nombre de faux positifs et le nombre de locus positifs :

$$\text{TFD} = \frac{\text{FP}}{\text{FP} + \text{VP}} \quad (1.2)$$

Bien que sa construction soit légèrement plus complexe que celle de la  $p$ -valeur, il existe une quantité, nommée  $q$ -valeur, qui contrôle le taux de fausse découverte exactement de la même manière. L'algorithme de STOREY & TIBSHIRANI (2003) permet de transformer une distribution de  $p$ -valeur en  $q$ -valeur.

### 3.3 Quelle erreur choisir ?

Considérons le cas suivant. Nous faisons le test de sélection sur 1 million de locus différents. Faisons l'hypothèse que le modèle est parfaitement calibré et qu'il renvoie, quand les locus sont neutres, une distribution totalement uniforme de  $p$ -valeurs. En prenant le seuil (courant en biologie) de 5%, on s'attend à obtenir 50 000 faux positifs. S'il existe 100 locus sous sélection dans le génome, ça donne un total de 50 100 positifs, soit un taux de fausse découverte de plus de 99% : la très vaste majorité des positifs sont des faux positifs.

Il est souvent proposé de prendre en compte le nombre de tests dans des corrections de type *Bonferroni* par exemple. Dans ce cas, il faudrait diviser le seuil de significativité de notre  $p$ -valeur par 1 million. À moins d'avoir une puissance de test considérable (un très grand nombre de populations par exemple), atteindre ce genre de significativité est relativement difficile. Littéralement, ce genre de corrections cherche à réduire l'espérance du nombre de faux positifs à 0.05. Ici, nous tombons donc

dans l'excès inverse : le taux de fausse découverte serait autour de  $5 \cdot 10^{-4}$ . Ce taux est tellement drastique que la puissance du test ne peut qu'être considérablement réduite.

Une façon plus raisonnable de contrôler le taux d'erreur lorsque l'on réalise un grand nombre de tests est donc d'utiliser le taux de fausse découverte. Ceci a plusieurs avantages. D'abord, c'est un taux qui est facilement interprétable : c'est la proportion de faux positifs dans l'ensemble des résultats positifs. Si notre test nous renvoie 100 locus significatif avec un taux de fausse découverte à 5%, il y a 5 faux positifs (en moyenne) dans le lot. Ensuite, ce taux d'erreur ne dépend pas du nombre de tests effectués. Enfin, il a été montré que le taux de fausse découverte permet une classification optimale des résultats en terme de faux et vrais positifs lors de tests multiples (STOREY, 2003).

## 4 Présentation de l'ARTICLE 1

L'article ci-dessous se propose de tester la robustesse de différentes méthodes de scan génomique appartenant aux deux familles principales (méthodes à  $F_{ST}$  et méthodes d'association environnementale). Des scénarios démographiques complexes ont été simulés pour tester ces méthodes dans des conditions semi-réalistes. Étant donné que la sélection n'agit que très rarement sur un seul gène en même temps, nous avons comparé ces scénarios avec un modèle de sélection monogénique (1 seul locus sélectionné) ou polygénique (50 locus sélectionnés).

Une difficulté pour comparer les différentes méthodes vient du fait qu'elles utilisent des statistiques de test différentes ( $p$ -valeur,  $q$ -valeur, Bayes Factor...). Partant du constat ci-dessus que le taux de fausse découverte est ce qui importe dans ce type d'analyses, nous avons décidé, contrairement à ce qui avait été fait auparavant (PÉREZ-FIGUEROA et al., 2010 ; DE MITA et al., 2013) de transformer toutes ces statistiques en  $q$ -valeur afin de ramener toutes ces statistiques sur une échelle pertinente.

### 4.1 Contributions

Dans cet article, j'ai conçu l'analyse conjointement avec Oscar Gaggiotti. J'ai développé et mis au point les scénarios démographiques avec l'aide d'Oscar Gaggiotti et d'Éric Bazin. Les analyses ont été faites par moi-même et Éric Frichot pour la partie LFMM. L'interprétation des résultats a été menée par moi-même, Oscar Gaggiotti, Éric Frichot et Olivier François. Enfin, j'ai écrit l'article, avec l'aide d'Oscar Gaggiotti.

### 4.2 Résumé de l'article

La récente disponibilité du séquençage haut-débit a permis l'utilisation de marqueurs génétiques denses pour l'identification de régions du génome qui pourraient être sous l'influence de la sélection. Plusieurs méthodes statistiques ont été développées récemment dans ce but.

Ici, nous présentons les résultats d'une étude par simulations individu-centrées s'intéressant à la puissance et au taux d'erreur de méthodes de scan génomique populaires ou récentes : la régression linéaire, Bayescan, BayEnv et LFMM. Contrairement à certaines études précédentes, nous nous concentrons sur des structure de population complexes et hiérarchiques, ainsi que sur de la sélection polygénique. Nous utilisons le taux de fausse découverte pour comparer les méthodes fréquentistes et bayésiennes. Enfin, nous inspectons l'influence de la façon de traiter les données : fréquences alléliques ou génotypes individuels, pour LFMM et la régression.

Le classement relatif des méthodes est impacté par le type de sélection (monogénique ou polygénique). Pour les scénarios fortement hiérarchiques, avec en plus des effets de confusion entre la

démographie et l'environnement, la puissance des méthodes peut être très faible. À part pour un scénario, Bayescan présente une puissance et un taux d'erreur modérés. Les performances de BayEnv sont bonnes avec un scénario non hiérarchique, tandis que LFMM possède le meilleur compromis entre puissance et taux d'erreur à travers l'ensemble des scénarios. Nous démontrons qu'il est possible de réduire fortement le taux d'erreur en combinant les résultats des trois différentes méthodes pour identifier les locus sous sélection.

**Journal:** *Molecular Ecology*, Volume 23, Issue 8 (2014)  
Les *Supplementary Information* de cet article sont disponibles dans l'Annexe S1

## ARTICLE 1

# Genome scan methods against more complex models: when and how much should we trust them?

Pierre de Villemereuil, Éric Frichot, Éric Bazin, Olivier François  
& Oscar E. Gaggiotti

### Abstract

The recent availability of Next Generation Sequencing (NGS) has made possible the use of dense genetic markers to identify regions of the genome that may be under the influence of selection. Several statistical methods have been developed recently for this purpose. Here, we present the results of an individual-based simulation study investigating the power and error rate of popular or recent genome-scan methods: linear regression, Bayescan, BayEnv and LFMM. Contrary to previous studies, we focus on complex, hierarchical population structure and on polygenic selection. Additionally, we use a False Discovery Rate (FDR) based framework, which provides a unified testing framework across frequentist and Bayesian methods. Finally, we investigate the influence of population allele frequencies *versus* individual genotype data specification for LFMM and the linear regression. The relative ranking between the methods is impacted by the consideration of polygenic selection, compared to a monogenic scenario. For strongly hierarchical scenarios with confounding effects between demography and environmental variables, the power of the methods can be very low. Except for one scenario, Bayescan exhibited moderate power and error rate. BayEnv performance was good under non-hierarchical scenarios while LFMM provided the best compromise between power and error rate across scenarios. We found that it is possible to greatly reduce error rates by considering the results of all three methods when identifying outlier loci.

### Introduction

The detection of signatures of selection has been a long-standing interest of population geneticists and evolutionary biologists. However, until recently, the paucity of molecular markers available limited the power of statistical methods to detect selection because other biological processes such as structure and migration have confounding effects on polymorphism and linkage disequilibrium. This situation has changed radically with the advent of Next Generation Sequencing (NGS, see SHENDURE & JI, 2008), which can generate dense arrays of markers, typically Single Nucleotide Polymorphisms (SNPs), spread across the genome. These new data can be used to distinguish between neutral processes that have a genome-wide effect (e.g. demographic history) and processes that have a local effect, particularly selection (LUKART et al., 2003). Several so-called genome-scan methods have been developed for this purpose (reviewed in DE MITA et al., 2013).

One of the most popular types of methods is based on an idea first proposed by LEWONTIN & KRAKAUER (1973). The underlying rationale is that loci influenced by directional selection will show larger genetic differentiation than neutral loci while the opposite is true for loci subject to balancing selection. Thus, loci that exhibit unusually high or low  $F_{ST}$  are good candidates for being influenced by selection. Several variants of this test exist (e.g. BEAUMONT & NICHOLS, 1996; VITALIS et al., 2001; BEAUMONT & BALDING, 2004; FOLL & GAGGIOTTI, 2008) and have been frequently applied to non-model species. Another recent group of genome-scan methods is based on the idea that many selected loci should be correlated with the environmental factors underlying the selective pressure (JOOST et al., 2007; COOP et al., 2010; FRICHOT et al., 2013). Genotype-environment association methods identify loci that show strong correlations with one or more environmental variables, and those loci are interpreted as potentially under selection.

All genome scan methods are based on the premise that it is possible to clearly distinguish between the genetic signals left by neutral and non-neutral processes. However, this assumption is frequently violated in real life scenarios (HERMISSON, 2009). Several demographic processes such as allele surfing (EDMONDS et al., 2004) and bottlenecks can leave signatures that mimic those left by positive selection. Moreover, complex spatial structuring can increase the variance of genetic parameters across the genome leading to high false positive rates (EXCOFFIER et al., 2009). Sensitivity analyses published thus far (PÉREZ-FIGUEROA et al., 2010; DE MITA et al., 2013) focus on these confounding effects of demographic covariance among populations arising through migration. The overall pattern that emerges from these studies is rather positive. Although all evaluated methods suffer from either low power (differentiation-based methods) or high false-positive rates (genotype-environment association methods), a strategy based on the use of both types of methods seems to lead to reliable identification of outlier loci. Nevertheless, one limit of existing studies is that they consider the effect of selection on a single locus. This is a quite unjustified assumption because selection for a specific quantitative phenotypic trait will influence several regions across the genome (ROCKMAN, 2012). There are only two studies (NARUM & HESS, 2011; VILAS et al., 2012) that considers several selected loci. However, the first is limited both by the number of loci (only 5) and the number of replicates of the simulated data, while the second focuses on the question of whether or not detected outlier markers are physically close to selected loci.

In this study we focus attention on more realistic scenarios than those considered in previous analyses. In particular, we investigate biases that may arise when selection acts upon traits determined by several genes. Indeed, as ROCKMAN (2012) recently pointed out, there is a paucity of empirical and theoretical support for the abundance of large-effect Quantitative Trait Nucleotides (QTNs) in the wild. Instead it is likely that “alleles that matter for evolution” are numerous small-effect loci. It is unclear if current genome-scan methods will simply have low power or if they will also have a high false discovery rate when applied to these situations. Another important consideration about real populations and species is that they are unlikely to be at migration-drift equilibrium. Thus, we evaluate scenarios where they have experienced recent divergence from an ancestral population, a process that may also affect power and false discovery rates of existing methods.

Instead of evaluating the performance of a very large number of methods we focus on a few that have proven popular or that are very recent and tested only under some restricted scenarios. More precisely, we focus on two genotype-environment association Bayesian methods that explicitly take into account the covariance of allele frequencies across populations (COOP et al., 2010; FRICHOT et al., 2013) and we compare these methods to one of the most frequently used genome-scan methods based on population differentiation (FOLL & GAGGIOTTI, 2008). We did not include more population differentiation methods as they have been shown to be less efficient than this particular one (PÉREZ-FIGUEROA et al., 2010; VILAS et al., 2012; DE MITA et al., 2013). We further consider a naive frequentist regression approach without any correction for population structure. The comparison is done using a rigorous statistical framework based on false discovery rates (FDR, see BENJAMINI & HOCHBERG, 1995) and *q-values* (STOREY & TIBSHIRANI, 2003; STOREY et al., 2004), which allow for a unified comparison of the performance of the methods.

## Material & Methods

### Simulation model

We carried out simulations using the SimuPop package for Python (PENG & KIMMEL, 2005). We focused on highly structured population scenarios where selection acts on a multigenic trait. For the sake of clarity we describe each component of the simulation model separately and also present the main attributes of each scenario in TAB. A1.1. We simulated 100 replicates for each scenarios (but only used 50 for Bayescan, see below).

### ► Demographic process

Our main scenario is a dichotomous process of population fission in which an ancestral population of 500 individuals gives birth to two descendant populations after 50 generations of drift. The fission is instantaneous with local populations reaching carrying capacity of 500 individuals in a single generation. This dichotomous fission process is repeated until 16 populations are obtained (see dendrogram, FIG. A1.1.A).

Migration occurs all along the process and preferentially between historically close populations: two populations issued from the same fission event will exchange twice as many migrants as two populations issued from two distinct fission events. In other words, the proportion of migrants between two populations is determined by phylogeographic distance. We aimed at capturing the main features of a spatial expansion in a heterogeneous habitat. For example, a post-glaciation colonisation scenario, where new valleys and sub-valleys are progressively reached. The further apart two populations are along the population tree, the lower the migration rate between them. This model corresponds to a highly structured Isolation with Migration model (noted HsIMM). We assume a recent demographic origin for all populations (500 generations in total since the initial fission event). In addition, we consider two simpler scenarios: an isolation with migration (IMM) model where the sixteen populations are issued from a single fission event and a stepping stone scenario (SS) where all sixteen populations are issued from a single fission event. For these two models the length of the runs was 400 generations. These settings allow us to stop the simulation at a near-equilibrium situation. In all scenarios, each population consists of 500 individuals.

The proportion of individuals in a local population that do not migrate,  $(1 - m)$  is the same under all three scenarios but the proportion of individuals that migrate between pairs of populations differ. Under the HsIMM it is  $m/2^{(i+1)}$  where  $i$  is the number of fission events between each local population and the most recent common ancestral population (SuppInf S1 eq. S1.1). Under the IMM it is  $m/15$  for all pair of populations (SuppInf S1 eq. S1.2). Under the SS model it is equal to  $m/2$  for neighbouring populations and zero for all other pairs of populations (SuppInf S1 eq. S1.3). For all simulations we chose  $m = 0.0045$ , which yielded pairwise  $F_{ST}$  roughly equal to 0.1.

More information about the simulation process can be found in the supplementary information (SuppInf S1.1). The python code used can also be found online in the data accessibility section.

### ► Genetic process

We simulated 5000 SNP regularly spread along 10 chromosomes. The recombination rate between adjacent pairs of SNPs is set to 0.002 in order to have, on average, one recombination event per population per generation. This amounts to spacing 500 SNPs uniformly along each chromosome. The mutation rate is set to  $10^{-7}$  per generation at each SNP. We consider two genetic architectures: either a single locus case, or 50, randomly distributed, loci influencing a phenotypic trait directly linked to fitness. In each case, we assume co-dominance.

We use a multiplicative fitness function to describe the ‘cumulative’ effect of all loci on fitness :

$$W = (1 + s_p)^{n_{11}} (1 - s_p)^{n_{00}} \quad (\text{A1.3})$$

where  $s_p$  is the local coefficient of selection (depending on the local value of the environment, see next paragraph) and  $n_{11}$  and  $n_{00}$  are the number of (1, 1) and (0, 0) homozygous loci, respectively. Note that fitness is normalized such that the relative fitness of any heterozygous locus is 1. For small  $s$ , this multiplicative fitness function is equivalent to an additive one.

### ► Environmental variable underlying the selective pressure

In the case of a highly structured model (HsIMM) we consider two spatial patterns of selection intensities, which are determined by an environmental variable  $E_S$ : ( $i$ ) at each population fission,

the values of  $E_S$  for each descendant population are drawn from a uniform distribution centred on the value of the ancestral population (HsIMM-U) and, (ii) at each population fission, the values of  $E_S$  for the new populations are set such that they produce an environmental gradient along a linear habitat (HsIMM-C). For the isolation with migration (IMM) and stepping-stone (SS) scenarios, the values of  $E_S$  are also set to form an environmental gradient, like in case (ii) (TAB. A1.1).

The local coefficient of selection  $s_P$  is calculated as a logistic transformation of the environmental variable :

$$s_P = s \frac{1 - e^{-\beta E_S}}{1 + e^{-\beta E_S}} \quad (\text{A1.4})$$

where  $s$  is the ‘baseline’ selection coefficient and  $\beta$  is the ‘slope’ of the logistic transformation. For the scenario with a single selected locus we set  $s$  to 0.1 and  $\beta$  to 1. In the case of the polygenic scenario we use  $s = 0.004$  and  $\beta = 5$ . The difference in parameter values between the two scenarios is necessary because, for size effect  $s$  in the monogenic case and  $s/N$  in the polygenic case, local adaptation progresses much more slowly under the polygenic architecture. Therefore, it was necessary to increase both the effect size and slope of the gradient for the polygenic case so as to generate local adaptation patterns under both scenarios in a similar evolutionary time. The values were scaled so that the mean allelic frequency pattern in the polygenic case was similar to the one in the monogenic case.

We also investigate the potential for spurious selection signals due to the consideration of environmental factors unrelated to any selective pressure. For this we consider scenarios that include a selectively neutral environmental variable  $E_0$  whose values are randomly drawn from a normal distribution. Selection starts at the second fission events in the HsIM scenarios, and at the (only) first one in the two other scenarios.

TAB. A1.1: Description of the scenarios considered in this study

Scenario	Spatial Model	Demographic History	Selection Pattern
HsIMM-U	Hierarchical	Multiple Binary Fissions	Correlated with demographic history
HsIMM-C	Hierarchical	Multiple Binary Fissions	Environmental Gradient
IMM	Standard IMM	Instantaneous Fission	Environmental Gradient
SS	Stepping-Stone	Instantaneous Fission	Environmental Gradient

## Statistical analysis

### ► Error rate

For all methods, we use  $q$ -values as a significance test statistic (STOREY & TIBSHIRANI, 2003; STOREY et al., 2004). The  $q$ -value is tightly linked to the False Discovery Rate (FDR, STOREY & TIBSHIRANI, 2003). For a statistical test, the FDR is equal to the number of false positives over the total number of positives (true and false). Thus, it is the proportion of “false discoveries” among all the “discoveries” of the test. If the assumptions of the test hold, then a given threshold  $\alpha_q$  for assessing the significance of  $q$ -values should lead to a FDR of  $\alpha_q$ . For example, if one decides a cut-off threshold of 5%, then the test will yield 95% of true positives and 5% of false positives.

Note that, in this sense, a cut-off of 5% for  $q$ -values is much more stringent than the same cut-off for  $p$ -values. It is important to distinguish between false positive rate, false discovery rate and power: their relationship is explained further in SuppInf S1. Note that, for the same dataset, an increase in power would lead to a decrease of FDR, whereas an increase in false positive rate (FPR) would lead to an increase of the FDR.

### ► Power

For monogenic selection scenarios, the definition of power is straightforward: it is the proportion of truly selected loci that are significant (see also Eq. S1.6 in SuppInf S1). For polygenic selection, this definition leads to a value of power for each locus. We computed power for each locus for each simulation, and then averaged over all loci, in order to get a mean power comparable to the case of monogenic scenarios. Note that, in the case of polygenic scenarios, we have less sampling error than in monogenic scenarios, because we have 50 times more selected loci.

### ► Data specification

Some methods can be applied either to population allele frequency data or to individual genotype data. In principle, using genotypic data is more appropriate when it is difficult to clearly define population boundaries. It can also avoid potential biases introduced by differences in sample sizes across populations. We investigated the influence of data specification for the linear regression and the Latent Factor Mixed Model methods (see below).

## Genome scan methods to detect selection

There are several genome-scan methods aiming at detecting selection by identifying outlier loci. Here we focus on two genotype-environment association methods that explicitly take into account the allele frequency covariance across populations and we compare these methods to a genome-scan method based on population differentiation. We further consider a naive frequentist approach that tests for correlations between allele frequencies and environmental factors.

### ► BayEnv

A first method that takes into account the allele frequency covariance across populations generated by demographic history and spatial effects was developed by COOP et al. (2010) and is implemented in the software BayEnv. This method consists in a two-step procedure. First, a model using all loci (or a part of the data set that is known to be neutral) estimates the population structure using a variance-covariance matrix of allele frequencies between populations. Second, a model incorporating the empirical covariance matrix tests for the correlation between the allele frequencies at each locus (or only at loci of interest) and each environmental variable.

The software returns Bayes factors weighting the strength of evidence in favour of a correlation between allele frequencies and the environmental variable. We transform the Bayes Factors into posterior odds using a prior probability of the null model  $\pi_0 = 0.99$ , and use these odds to compute *q-values* (STOREY & TIBSHIRANI, 2003; STOREY et al., 2004) which are used to assess the significance of each locus. The procedure is explained in the supplementary information (SuppInf S1), where we also provide the MCMC parameters used.

### ► Latent Factor Mixed Model

Latent Factor Mixed Models (LFMMs, see FRICHOT et al., 2013) are very general and flexible models and provide an alternative approach to detect relationships between allele frequencies and environmental values, while taking into account population structure. The model can be seen as an approximate Principal Component Analysis combined with a regression. It is computationally faster than BayEnv and Bayescan (FRICHOT et al., 2013).

The K value (number of factors) needed by the software are estimated to be 15 for every scenarios, using Tracey-Widom tests. The *p-values* returned by the method are transformed into *q-values* following a standard procedure (STOREY & TIBSHIRANI, 2003). We used the version 1.2 of the software.

### ► Bayescan

Bayescan (FOLL & GAGGIOTTI, 2008) is an  $F_{ST}$ -based model (BEAUMONT & BALDING, 2004). This method is not searching for a potential correlation between allele frequencies and the environment.



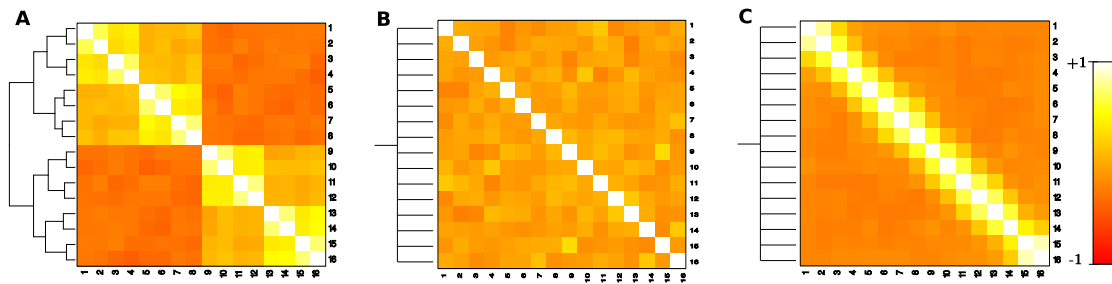


FIG. A1.1: Heatmap of allele frequencies correlation between all simulated 16 populations. Panel A: HsIMM model; Panel B: IMM model; Panel C: SS model. The red to white gradient corresponds to the  $[-1, 1]$  interval. The dendrogram illustrates proximity between populations (inferred for HsIMM, drawn for IMM and SS).

Instead, it is searching for loci exhibiting extreme  $F_{ST}$  values. Large  $F_{ST}$ s are then interpreted as signatures of local adaptation. It is testing for outliers independently of any environmental knowledge.

The statistical significance is assessed by the use of  $q$ -values (STOREY & TIBSHIRANI, 2003 ; STOREY et al., 2004) using a prior odds of 100. The MCMC parameters used are detailed in SuppInf S1. Because of computation time issues, we only used 50 replicates for this method.

### ► Corr: Allele frequencies-environment regression

This is the most naive method and only aims at detecting a correlation between population allele frequencies and an environmental variable. Typically, significance is evaluated using the  $p$ -value returned by Student's test on the slope of the regression. However, to correct for multiple tests and to easily compare with results of other methods we transform the  $p$ -values into  $q$ -values using the method presented in STOREY & TIBSHIRANI (2003).

## Results

### Genetic structure produced by the population models

As expected, our simulation models produce highly structured population genetic data. FIG. A1.1 shows the structure of the correlation in allele frequencies across populations, as estimated by the software BayEnv (COOP et al., 2010). For the binary fission model (FIG. A1.1.A), the strength of the correlation decreases with the phylogeographic distance. The isolation with migration model (FIG. A1.1.B) produces no apparent spatial pattern while the stepping stone model (FIG. A1.1.C) leads to a typical isolation-by-distance pattern.

### Monogenic selection

#### ► Error Rates

The expectation is that the False Discovery Rate (FDR) increases linearly with the threshold used to decide the significance of  $q$ -values. However, the results differ radically from this expectation. Indeed, the FDR of all methods was higher than expected under all scenarios (FIG. A1.2), except for LFMM in the IMM scenario, which is even quite conservative. Note also that BayEnv has an acceptable FDR for the SS scenario and stringent thresholds (FIG. A1.2, SS). This inflation in FDR is partly due to the fact that, when only one locus is truly selected, even a small false positive rate, when combined with high power, leads to very high FDRs. Regarding hierarchical scenarios, when the spatial selection pattern is a function of phylogeographic distance (FIG. A1.2, HsIMM-U), FDRs are highest for Bayescan and lowest for LFMM, while the FDR values for BayEnv and the linear regression methods are intermediate. When selection is a function of an environmental gradient (FIG. A1.2, HsIMM-C), the FDR is highest for the linear regression method, intermediate for Bayescan

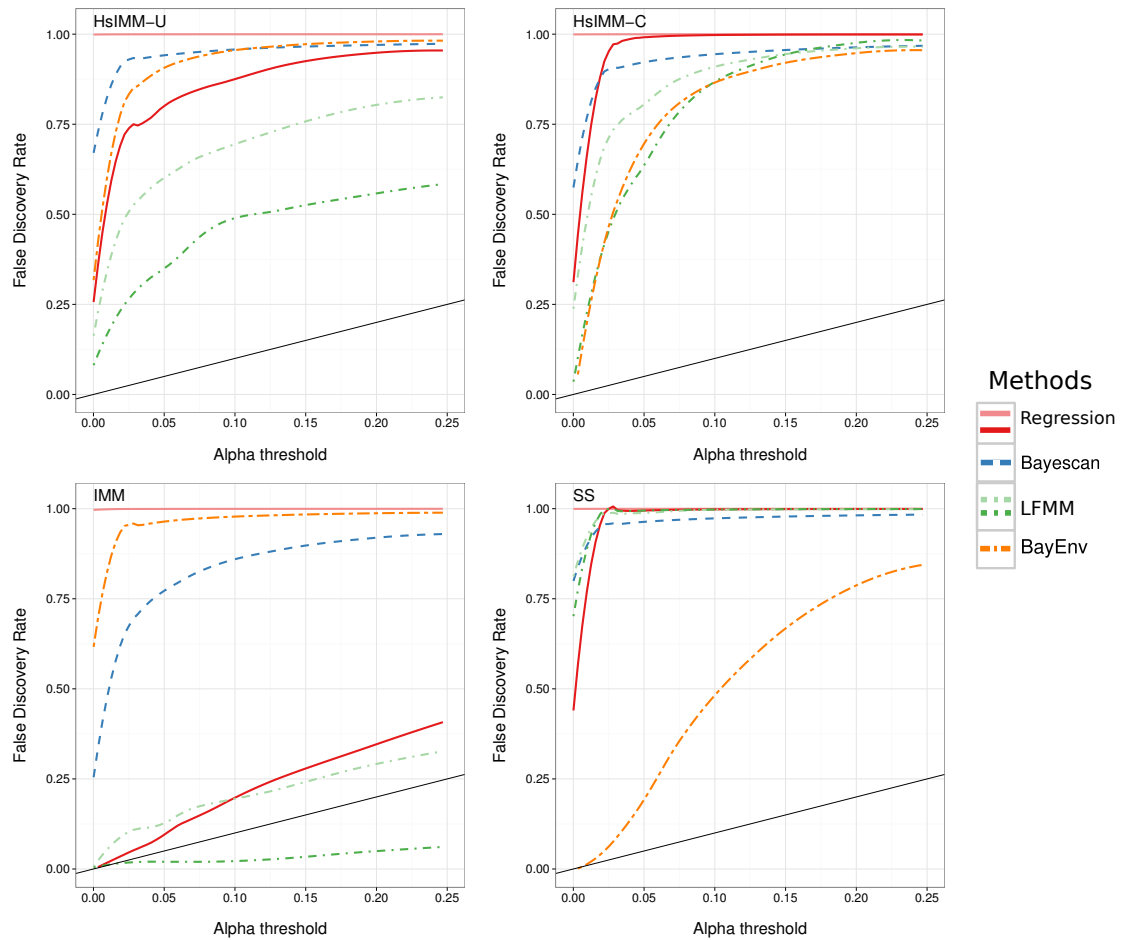


FIG. A1.2: False discovery rate against significance threshold ( $\alpha_q$ ) for monogenic selection. Black line: Expected relationship between FDR and  $\alpha_q$ . Lines are LOESS smooth for linear regression (plain red line), Latent Factor Mixed Model (LFMM, green dot-dashed line), Bayescan (blue dashed line) and BayEnv (orange two-dashed line). Light lines are for individual genotype data specification for the linear regression (light red) and LFMM (light green).

and lowest for BayEnv and LFMM. Thus, the spatial pattern in selection intensities greatly influences the relative performance of the different methods.

Note that the individual genotype data specification for the linear regression and LFMM (light lines) always lead to higher FDRs. This is especially the case for the linear regression with FDRs of almost 1. Note also that the linear regression method yields intermediate FDRs for small  $\alpha$  thresholds for both scenarios. Finally, recall that FDRs are not on the same scale as False Positive Rates (FPR). Since here we are considering a monogenic scenario, a FDR of 75% corresponds to the truly selected locus plus 3 false positives, thus to a FPR of 0.06%.

### ► Statistical power

Under the scenario HsIMM-U the power of all methods is moderate with a maximum between 75 and 80% for very permissive thresholds (FIG. A1.3., HsIMM-U, except the case of linear regression for individual genotype data specification, light red line). In this case, all recent methods had roughly similar power, although LFMM yields a lower one. The regression method has the lowest power for allele frequency data specification but the highest one when using individual genotypes. Under the other scenarios (FIG. A1.3, HsIMM-C, IMM and SS), the power of all methods is very high, being perfect for some of them regardless of the threshold value used.

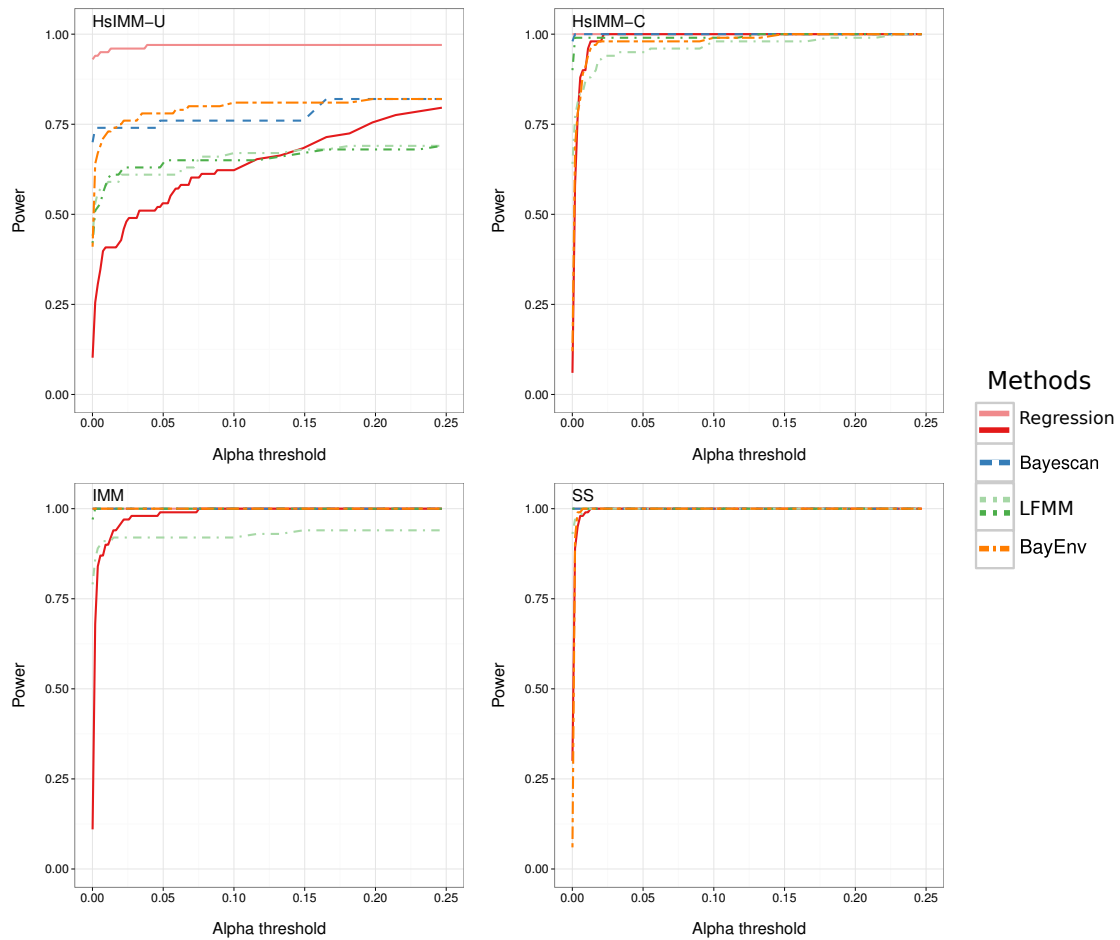


FIG. A1.3: Statistical power against significance threshold for monogenic selection. Lines are linear regression (plain red line), Latent Factor Mixed Model (LFMM, green dot-dashed line), Bayescan (blue dashed line) and BayEnv (orange two-dashed line). Light lines are for individual genotype data specification for the linear regression (light red) and LFMM (light green).

Note that, in all cases, the regression model is always among the least powerful methods. Also, whereas the “individual genotypes” specification always increase the power for the regression (light red lines in FIG. A1.3), this is not always the case for LFMM (light green lines).

## Polygenic Selection

### ► Error Rates

As it was the case for the monogenic selection scenario, the false positive rate of all methods under all scenarios was higher than expected. FIG. A1.4 shows that the expected linear increase in FDR with increasing threshold values only holds for BayEnv under the stepping-stone model (FIG. A1.4, SS). Interestingly, LFMM shows a very conservative pattern for the IMM scenario, when using the population frequency data specification (FIG. A1.4, IMM, dark green line). All other combinations of scenarios and methods are more error-prone than our theoretical expectation. Note in particular that all methods have very high FDR under the hierarchically structured IMM scenarios. While LFMM is the most conservative method in the case of environment correlated with demography (FIG. A1.4, HsIMM-U), BayEnv and Bayescan are the approaches that are the least error prone for a clinal environment (FIG. A1.4, HsIMM-C). The behaviour of LFMM and BayEnv changes radically across scenarios and they seem specially well adapted to a specific scenario (IMM and SS respectively) while the error rate of Bayescan is more intermediate across the different scenarios, although it is one of the worst under the standard IMM model.

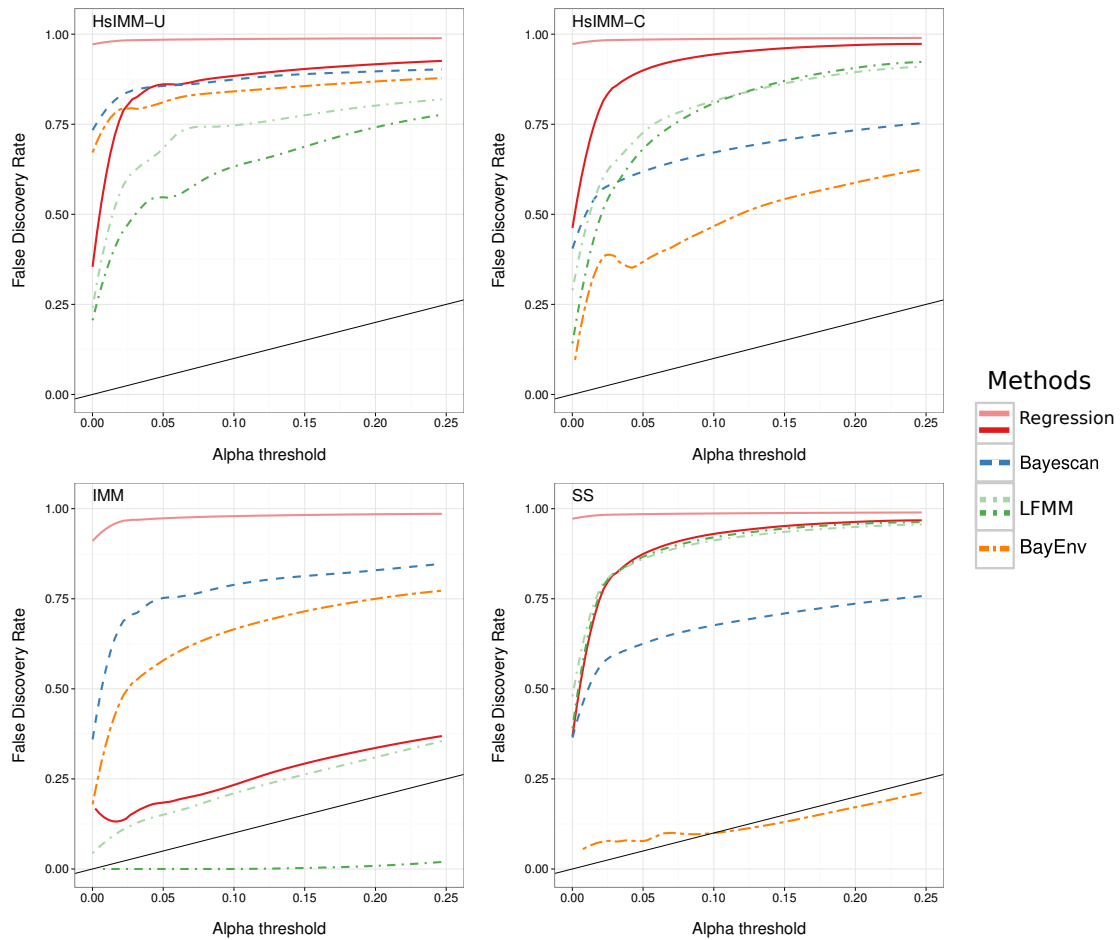


FIG. A1.4: False discovery rate against significance threshold ( $\alpha_q$ ) for polygenic scenarios. Black line: Expected relationship between FDR and threshold value  $\alpha_q$ . Lines are LOESS smooth for linear regression (plain red line), Latent Factor Mixed Model (LFMM, green dot-dashed line), Bayescan (blue dashed line) and BayEnv (orange two-dashed line). Light lines are for individual genotype data specification for the linear regression (light red) and LFMM (light green).

Regarding data specification, LFMM seems to be quite robust to its influence, although the individual specification still tends to yield more erroneous results than its allele frequency counterpart. The linear regression model, however, is much less robust: its individual genotype specification version is always the most error-prone, while its population allele frequencies specification can yield relatively conservative results (e.g. see FIG. A1.4, IMM).

### ► Statistical power

Because of the small effect size of each locus under the polygenic model, the power of all methods should be lower than under the single-gene model. Indeed, we do observe an overall decrease in power for all scenarios (FIG. A1.5 compared to FIG. A1.3). The linear regression is the method that had the highest power under scenarios HsIMM-U, HsIMM-C. This power performance is followed by LFMM (FIG. A1.5 HsIMM-U and HsIMM-C). These two methods are comparable for the SS scenario (FIG. A1.5). Regarding these three models (HsIMM-U, HsIMM-C and SS), Bayescan shows intermediate power and BayEnv is the least powerful method. Interestingly, the behaviour of the methods is very different for the IMM scenario (FIG. A1.5): here, BayEnv is one of the most powerful methods, only outperformed by the error-prone linear regression in its individual genotype specification. LFMM and Bayescan are the two worst methods.

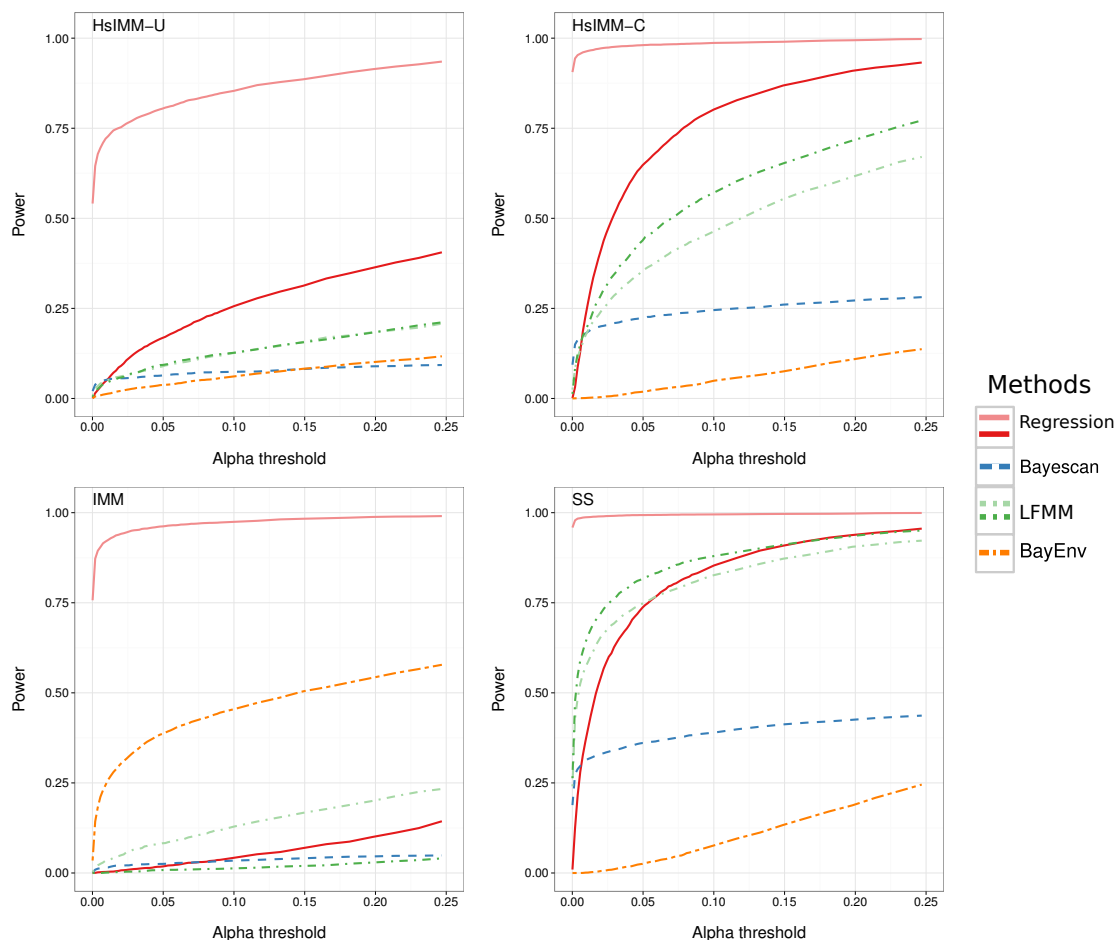


FIG. A1.5: Statistical power against significance threshold for polygenic selection. Lines are for linear regression (plain red line), Latent Factor Mixed Model (LFMM, green dot-dashed line), Bayescan (blue dashed line) and BayEnv (orange two-dashed line). Light lines are for individual genotype data specification for the linear regression (light red) and LFMM (light green).

While having a high power is an interesting feature, it needs to coincide with reasonable False Positive and False Discovery rates to be relevant. Power against False Positive (ROC curves) and False Discovery rates are provided in SuppInf S1. The ROC curves (Fig. S1.4 in SuppInf S1) illustrate the compromise between the number of true and false positives and show that all methods are comparable in this regard. The “power against FDR” graphs (Fig. S1.7 in SuppInf S1) provide information about how many true positives are detected by the methods. For a given FDR, more power means more true (and false) positives.

### ► Consistency between methods

Overall the methods tend to disagree from each other, in terms of which loci should be considered as selected (true or false positives). The percentage of overlap between loci considered as positives by two different methods is around 1% to 5%, except for the regression and LFMM (14% to 48% depending on the scenarios). Notable exceptions are the HsIMM scenarios, where Bayescan and LFMM reach an agreement on 13% of loci under selection for HsIMM-U and 18% for HsIMM-C. Still, the methods are more often in agreement regarding true positives than regarding false positives. This means that using 3 methods to assess the outlier behaviour of loci leads to a substantial decrease of the FDR. This decrease varies between 0.4 and 0.65, depending on the scenarios. For the IMM model, this strategy yields a FDR of 0% (all positives are true positives). Unfortunately, using several methods leads to a decrease of power of roughly the same magnitude as the decrease in FDR (between 0.25 and 0.55).

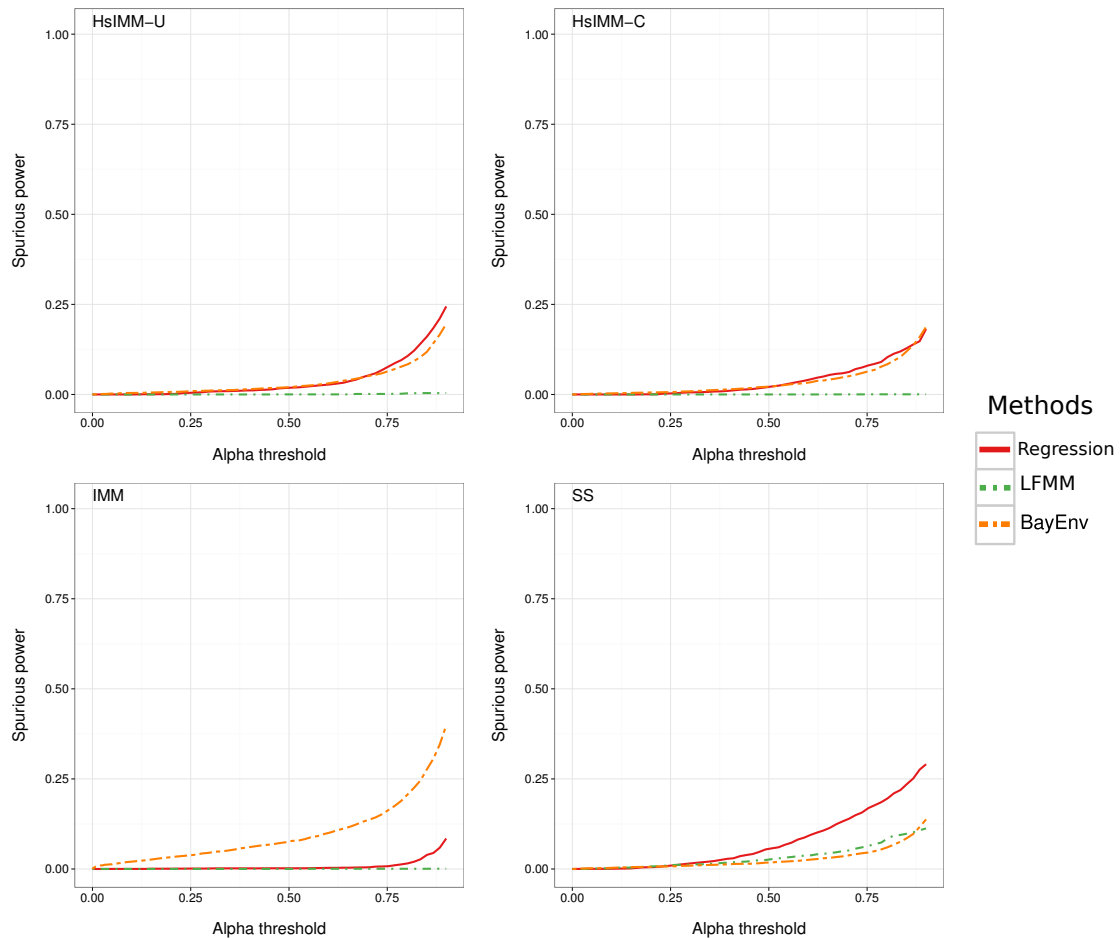


FIG. A1.6: Spurious “power” (i.e. power to detect selected loci using an independent spurious variable) against significance threshold for the null environmental variable. Line are for linear regression (plain red line), Latent Factor Mixed Model (LFMM, green dot-dashed line), and BayEnv (orange two-dashed line). Light lines are for individual genotype data specification for the linear regression (light red) and LFMM (light green).

### ► Spurious environmental variable

Methods that use environmental variables to identify outliers assume that the chosen variables exert a selective pressure or are highly correlated to the one directly involved. One possible outcome in this situation is that the statistical tests identify a truly selected locus, but assign it to the wrong environmental variable. Although detecting a locus under selection is desirable, one does not want to link it to a spurious environmental variable. We call this error rate “spurious power” and define it as the proportion of truly selected loci considered as positive using a spurious, unrelated environmental variable. FIG. A1.6 shows that for HsIMM scenarios, the linear regression and BayEnv methods do not differ much in their “spurious power” (here, we only focus on the polygenic selection case). However, LFMM has a very low spurious power. For the IMM scenario, BayEnv is the most prone to erroneous choice of selective variable. By contrast, the linear regression is the most prone to error for the SS scenario.

Note that, in principle, the spurious power should be equal to the overall false positive rate (FPR), because we expect no association between the spurious environmental variable and the selected loci. This is roughly the case for all methods, except for LFMM in the scenarios IMM and SS. It tends to detect (false) association for selected loci more often than for non selected loci (see FIG. S1.8 in SuppInf S1, note that the scale on these graphs are totally different from FIG. A1.6, since the methods differ in their False Positive Rate).

## Discussion

### ► Performances of the methods against difficult scenarios

This study aimed at assessing the performance of recent and/or popular genome scan methods, in terms of power and error rate, when applied to difficult scenarios. The relative ranking of the methods, for the polygenic case, is summarised in TAB. A1.2. Note that the relationship in ranking between the FDR and the FPR is strong. Therefore, the methods have an inflated FDR mainly because of too many false positives, not because of too few true positives.

The most important challenge to the performance of all methods is the polygenic selection process. Obviously, one would expect an overall decrease in power for all methods when using a polygenic selection model compared to a monogenic one, something that was actually observed. This decrease in power went hand in hand with an overall decrease of FDR (mostly due to the increased number of selected loci, see Eq. S1.7 in SuppInf S1). However, the impact on performance differed among methods leading to a radical change in their ranking in terms of power/FDR. While all methods performed roughly equally in the monogenic scenarios –especially regarding power, for polygenic scenarios we observed large differences. First, the regression method became one of the most powerful but also most error prone methods. Second, the relative ranking between Bayescan, BayEnv and LFMM was changed, both in terms of power and error rate.

The second most important challenge was a strongly hierarchical spatial structure. This is evident when comparing the results for scenarios HsIMM-C and SS, both of which consider selection along an environmental gradient: the HsIMM-C scenario led to lower power for all methods. Note that the FDR for BayEnv was also inflated in the HsIMM-C scenario whereas it was almost perfect in the SS scenario. Apart from this overall changes in behaviour, the ranking of the method was conserved between the two spatial scenarios (although LFMM and the linear regression tend to be alike under the SS scenario).

The last challenge under study was the correlation between the environmental variable underlying the selective pressure and demographic history. The effects of this process can be visualised by comparing scenarios HsIMM-U and HsIMM-C, which only differ in this particular aspect. Overall, we see that a correlation between environment and demography led to low power for all methods, and higher FDR for Bayescan and BayEnv, which became even more prone to error than LFMM. The ranking, in terms of power, of the methods was conserved between the two kinds of scenarios.

TAB. A1.2: Summary of the properties of each genome scan method, under the different scenarios focusing on the polygenic case. FDR: False Discovery Rate; FPR: False Positive Rate. Methods are ranked from the best (\*\*\*\*) to the worst (\*). All properties are compared against the  $\alpha$  thresholds below 0.05. When the ranking of the method was ambivalent, they were both assigned the same rank.

Scenario	Statistics	Regression	BayEnv	LFMM	Bayescan
HsIMM-U	FDR	**	**	****	*
	FPR	*	***	****	**
	Power	****	*	***	**
HsIMM-C	FDR	*	****	**	**
	FPR	*	****	**	***
	Power	****	*	***	**
IMM	FDR	***	**	****	*
	FPR	***	*	****	**
	Power	***	****	*	**
SS	FDR	*	****	*	***
	FPR	*	****	*	***
	Power	***	*	****	**

Another source of error to be considered in the case of association methods (e.g. the regression, LFMM and BayEnv) is that of associating the selected loci with a non selective (spurious) environmental variable. In this case, BayEnv and the linear regression methods yielded a stronger “spurious power” than LFMM. Also LFMM tended to associate the spurious variable with the selected loci more often than with the neutral loci.

We finally investigated the influence of the data specification (population allele frequencies or individual genotypes) for the linear regression and LFMM methods. The population allele frequencies data specification allowed for better performance in terms of error rate and most of the time in terms of power, at least under our simulated scenarios. This can be due to the fact that using genotypic data involved a larger sample size, which led to a higher rate of null model rejection due to slight violations of its underlying neutral hypotheses (higher power, but higher error rate). Note that, for polygenic selection, LFMM was less sensitive to the data specification. More puzzling, the genotypes specification sometimes led to a lower power.

### ► Characteristics of the methods and comparison to previous studies

Overall, we see that methods using an environmental variable have generally more power than genome-only based methods. Notably, Bayescan was always less powerful than at least one of the other methods. This is expected, since the method is not taking advantage of as much information as the others. One has to note, however, that sometimes it may not be possible to identify the environmental variable that should be considered, in which case a “blind” genome scan method must be used. Although Bayescan has been shown to perform quite well under Island and Stepping Stone scenarios (FOLL & GAGGIOTTI, 2008 ; NARUM & HESS, 2011 ; VILAS et al., 2012 ; DE MITA et al., 2013), it did not perform very well under our Isolation with Migration (IMM) model and polygenic selection. This is a potentially common scenario so the results of our study differ from those of previous ones in that they suggest caution when using  $F_{ST}$ -based genome scans. Note, however, that the low power under the IMM scenario was only severe for the polygenic case.

Regarding LFMM and BayEnv, the two methods have much in common: both approaches employ mixed models in which environmental variables are introduced as fixed effects whereas population structure is introduced using unobserved variables or hidden factors. Yet, there are two main differences between the two methods. First, whereas BayEnv is a two-step procedure, estimating first the covariance structure of the population allele frequencies, and only then testing for association with an environmental variable, LFMM uses hidden factors to capture the part of genetic variation that cannot be explained by the set of measured environmental variables, all at once. This variation could include unknown demographic history, IBD patterns or environmental gradients not accounted for in the study. Second, the PCA-related nature of LFMM would *a priori* allow the method to take into account more complex scenarios. In particular, BayEnv has already been shown to perform poorly when confronted to hierarchical structure, and perform quite well in an island model (DE MITA et al., 2013). In our study, on the other hand, we also included LFMM and observed that this method over-corrected under the low-structure IMM scenario, leading to a very low FDR, but also a lower power.

Since the regression is not correcting for any population structure, we would expect it to yield more false positives, most likely accompanied by higher power. The regression is indeed the most error-prone method for all scenarios, except for the IMM one (which is the least structured scenario). Note that, when increasing the number of selected loci (i.e. from the monogenic to the polygenic case), the compromise between false positives and power gets better for the regression model (see FIG. S1.3 and S1.4). This could be caused by the fact that the regression is a more sensitive method, which is less ‘reluctant’ to identify loci as selected. Thus, when many loci are selected, each with a small effect, we can expect this method to yield better power.

The results of this study differ from the previous ones (e.g. PÉREZ-FIGUEROA et al., 2010 ; NARUM & HESS, 2011 ; VILAS et al., 2012 ; DE MITA et al., 2013) in several aspects. First, we used the same metric (the *q-value*) for all methods, which allows for a fair comparison. Second, while other studies



investigated polygenic selection (NARUM & HESS, 2011 ; VILAS et al., 2012), they only considered up to 10 loci, and only investigated  $F_{ST}$ -based methods. Third, we used more complex models with strong hierarchical structure.

### ► General issues and properties of genome-scan methods

The results about polygenic selection tell us that assessing methods for monogenic scenarios only is not sufficient, especially because we expect the polygenic case to be the norm rather than the exception *in natura* (PRITCHARD & DI RIENZO, 2010). Of course, we have assumed a model of small locus effects, which could be one of the most difficult for genome scan methods. All methods may perform better under an L-shaped distribution of locus effects (see an example in KULWAL et al., 2003), where a few loci have strong effects among numerous small effect loci. Yet, although there is evidence for the L-shaped architecture in the context of local adaptation (YEAMAN & WHITLOCK, 2011), there is also evidence that some phenotypic traits are under the control of many small-effect loci (reviewed in STRANGER et al., 2011 ; ROCKMAN, 2012).

Another important issue concerns methods that can consider both population- and individual-level data. In principle one expects that individual based data (genotypes) should lead to better performance, however, this is not necessarily the case. The type of data used has a large effect on the rate of false positives and consequently the FDR. We here illustrated this fact using LFMM and the linear regression models. Although we did not test it for BayEnv because the current implementation does not allow it, the results should be similar.

This result is due to the simple fact that using the individual genotypes instead of allele frequencies (by frequencies here, we mean allele count data) increases the number of observations. This has the desirable property of increasing the power, but also leads to the undesirable increase in number of false positives, because the null models are essentially false. Indeed no model is a perfect description of the data; there will always be a discrepancy with the underlying processes that lead to the data (because of non linearity of effects, small differences between the potentially assumed and real demographic history, non-uniform mutation rates, etc.) and increasing the number of observations lead to the rejection of the null model for most loci instead of only the outlier ones (c.f. RAFTERY, 1995).

Using population frequencies instead of genotypes is then a more conservative method. Yet it is not always possible to use frequencies, because of non homogeneous sample sizes, pooled sampling or the use of dominant data (e.g. AFLP). In those cases, one has to be aware that the statistical methods are not that robust to departures from the underlying model, and the more observation points there are, the higher is the overall false positive rate (RAFTERY, 1995). Note that this is true for the total number of sampled individuals and the way they are implemented in the models, but not for the number of loci, which does not *a priori* increase the false positive rate.

Finally high FDR, especially in the case of monogenic selection, corresponds to an acceptable (though still inflated) false positive rate (FPR). For example, a FDR of 75% for the monogenic case corresponds to a FPR of  $6.10^{-4}$  (see Eq. S1.4 in SuppInf S1). For the polygenic case though, and assuming a power of 20%, it will correspond to a FPR of  $6.10^{-3}$ . The fact that the methods tend to disagree might seem like a drawback, but it is in fact advantageous, because they tend to agree more on true positives than on false positives. Thus, by using all 4 methods together, we obtained FDRs between 0% and 40%, which are by far more acceptable.

### ► Perspectives & Conclusion

The results of our study pointed out two main directions in which statistical genomic studies should direct attention. First, we need more general and robust likelihood models that would be flexible enough to accommodate for strong departures from classical models. LFMM is an attempt in this direction, because its likelihood does not depend on a particular population model (FRICHOT et al., 2013).

Second, we need methods better adapted to polygenic selection scenarios. The *q-value* framework

allows to control for false discovery rate (STOREY & TIBSHIRANI, 2003 ; STOREY et al., 2004), which allow for test statistics that balance power and false positive rate. Another direction would be to develop a test that is suitable for polygenic selection. The difficulty in this case is that it would require to infer the genetic architecture of the trait(s) under selection, a very difficult task especially in absence of any phenotypic data.

Since polygenic selection and complex spatial population structures are likely to be quite common in the wild, it is important to tackle these two issues in order to develop reliable genome scan methods that can be applied to new NGS data from non-model species.

## Acknowledgement

We thank K. Csilléry and the anonymous reviewers for their useful insights. PdV was supported by a doctoral studentship from the French *Ministère de la Recherche et de l'Enseignement Supérieur*. OEG was supported by French ANR grant No 09 GENM 017 001 and by the Marine Alliance for Science and Technology for Scotland (MASTS). EF and OF were supported by a grant from *la Région Rhône-Alpes*. OF was further supported by Grenoble INP.

## 5 Conclusions principales de l'article

### 5.1 Taux d'erreur global des méthodes

La principale conclusion de cette analyse est que les méthodes de scan génomique ont clairement tendance à produire un fort nombre de faux positifs lorsque confrontés à des situations complexes mais réalistes. Pour les deux scénarios hiérarchiques ainsi que pour le modèle en "pas japonais" (*stepping-stone*), les taux de fausse découverte étaient ainsi généralement supérieurs à 50% pour un seuil de  $q$ -valeur à 5%. Ils pouvaient même dépasser les 75% soit les trois-quarts des résultats significatifs qui sont en réalité des faux positifs.

Sans être alarmiste ou défaitiste sur le comportement de ces méthodes face à des données réelles, il convient cependant d'être extrêmement circonspect en analysant les locus significatifs qu'elles produisent. Rappelons que, par définition, rien ne ressemble plus à un vrai positif qu'un faux positif et il est donc quasiment impossible d'éliminer ces faux positifs "à l'œil". Un fait intéressant, souligné dans cet article est qu'en combinant les méthodes, il est possible de drastiquement diminuer le taux de fausse découverte, au coût de perdre en puissance.

D'autre part, si les simulations peuvent être considérées comme réalistes, il faut souligner qu'elles représentent tout de même un défi pour les méthodes, puisque le gradient environnemental est totalement orienté dans le sens de la différenciation entre populations (du moins pour les scénarios HsIMM-U et SS). Cette orientation presque parfaite entre les deux gradients constitue la situation la plus complexe à gérer pour toute méthode d'association environnementale (FRICHOT et al., 2015).

### 5.2 L'influence du type de sélection sur les résultats

La seconde conclusion majeure de cet article est que simuler des scénarios avec une sélection monogénique (comme fait couramment dans la littérature) ne suffit pas. Bien sûr, comme attendu, le fait que la sélection soit polygénique "dilue" le signal adaptatif et donc la puissance des méthodes de scan génomique. Mais pas seulement : dans notre étude, le classement des différentes méthodes était très largement influencé par le fait que la sélection soit monogénique ou polygénique. Il est fort probable que ces changements soient principalement liés à la puissance des méthodes, étant donné que le taux de faux positif n'est que marginalement impacté par le type de sélection.

Étant donné que la majeure partie des traits sélectionnés sont des traits complexes (donc polygéniques), diluant ainsi le signal entre les locus sous-jacent à l'architecture génétique de ces traits, il est nécessaire que les méthodes de scan génomique soient validées par des scénarios polygéniques, en plus des scénarios monogéniques.

## 6 Un mot sur la simulation de sélection polygénique

Le passage de la simulation d'une sélection monogénique à la simulation d'une sélection polygénique a été étonnamment complexe. Dans un premier temps, le coefficient de sélection par locus avait été simplement divisé par le nombre de locus afin de retrouver la même force de sélection globale que dans le cas monogénique. Il s'est avéré qu'afin d'obtenir une réponse moyenne (sur l'ensemble des 50 locus) détectable, il a fallu augmenter substantiellement le coefficient de sélection par locus et la pente du gradient ( $s = 0.1$  et  $\beta = 1$  pour le cas monogénique,  $s = 0.004$  et  $\beta = 5$  pour le cas polygénique). En multipliant le coefficient de sélection par le nombre de locus, nous obtenons

une valeur totale de 0.2 pour le cas polygénique (soit 2 fois plus que pour le cas monogénique, sans prendre en compte la pente).

La raison de cette non linéarité de l'effet de la sélection vient du changement considérable de contrainte sur les fréquences alléliques des locus sélectionnés entre le cas monogénique et polygénique. Dans le premier cas, la sélection contraint directement les fréquences alléliques du seul locus sélectionné. Dans le second cas, il existe une infinité de possibilités de changements des fréquences alléliques qui vont augmenter la fitness. Du fait des "degrés de liberté" gagnés par le système, le bruit sur les fréquences alléliques est considérable et l'effet de la sélection, même moyenné sur l'ensemble des locus, diminué. Ainsi, pour obtenir une réponse détectable (avec, pourtant, une puissance faible, cf. FIG. A1.5), il nous a fallu augmenter le coefficient de sélection ainsi que la pente du gradient.

# CHAPITRE 2

## Une nouvelle méthode de scan génomique pour détecter l'adaptation locale

---

### 1 Une méthode à $F_{ST}$ plus robuste ?

#### 1.1 Différentes sources d'erreur

##### ► La structure de population

L'étude précédente a montré que le principal problème des méthodes de scan génomique vient de forts taux d'erreur, notamment des taux de fausse découverte pouvant dépasser 75% pour des seuils classiques (ici 5%). Dans les scénarios simulés, ces faux positifs venaient principalement de la structure de population simulée, et de la confusion entre cette structure et le gradient environnemental.

##### ► Le surf d'allèle

Dans les cas réels, il existe encore d'autres sources de faux positifs. Certains phénomènes, comme celui du "surf d'allèle" (*allele surfing* en anglais, EDMONDS et al., 2004), sont des conséquences subtiles de l'histoire démographique de la population. Il apparaît lors d'épisodes de (re)colonisations successives à partir d'une population ancestrale : si les événements de colonisations successifs s'accompagnent d'une forte dérive (effet de fondation), alors il est possible, avec une probabilité faible, qu'un allèle rare dans la population d'origine augmente en fréquence jusqu'à atteindre de fortes fréquences alléliques dans les populations du front de colonisation.

Cette forte augmentation en fréquence est virtuellement impossible à distinguer de l'effet d'une sélection graduelle lors de la (re)colonisation. Le fait que le surf d'allèle soit un événement rare pourrait être rassurant, malheureusement, ce n'est pas tout à fait le cas, puisque la rareté de ce type de locus les rend d'autant plus ressemblants à des locus sous sélection et rend le phénomène d'autant plus difficile à détecter.

##### ► Hétérogénéité du génome

La valeur de la  $F_{ST}$  peut être influencée par de nombreux facteurs génomiques comme le taux de mutation ou de recombinaison (EDELAAAR et al., 2011), certains régimes de sélection (CHARLESWORTH, 2013) ou des événements très particuliers d'incompatibilités entre allèles (KRUUK et al., 1999).

Une très forte hétérogénéité sur les taux de mutation peut entraîner une augmentation locale de la  $F_{ST}$  par le mécanisme suivant : si une région du génome mute beaucoup moins souvent que le reste du génome, le hasard des mutations n'agira pas dans le sens d'une moindre différenciation entre

populations à cet endroit du génome. Par conséquent, la différenciation génétique entre populations sera localement plus importante que le reste du génome, sans que la sélection naturelle n'en soit la responsable.

Un phénomène intéressant, qui peut être aussi responsable d'une augmentation de la  $F_{ST}$  est la sélection d'arrière-plan (CHARLESWORTH, 2013). Ce phénomène apparaît lorsqu'un locus est sous forte sélection purifiante : par un phénomène semblable à celui du faible taux de mutation ci-dessus (l'apparition de nouveaux variants étant limitée par la sélection purifiante), la  $F_{ST}$  des locus adjacents et liés à celui sous sélection purifiante peut augmenter considérablement.

## 1.2 Améliorer Bayescan ?

### ► Le comportement de Bayescan

Dans l'ARTICLE 1, le comportement de la méthode à  $F_{ST}$  est intéressant en cela que son taux de fausse découverte est à peu près stable par rapport aux différents scénarios considérés (FIG. A1.4). Pourtant, il reste relativement élevé, ce qui montre que la robustesse de cette méthode pourrait être améliorée, non seulement en ce qui concerne les problèmes de structure de population mis en évidence dans cet article, mais aussi potentiellement les autres phénomènes décrits ci-dessus.

### ► La stratégie adoptée

Afin d'arriver à une méthode plus robuste, nous avons décidé de rajouter une information que les méthodes à  $F_{ST}$  n'utilisaient pas encore : l'environnement. Utiliser cette information permet d'augmenter la certitude d'un scénario d'adaptation locale, puisque l'on distingue maintenant les locus qui sont fortement différenciés entre population au hasard et ceux qui sont différenciés suite à l'action de l'environnement.

## 2 Le modèle derrière Bayescan : le modèle $F$

### 2.1 Le modèle $F$

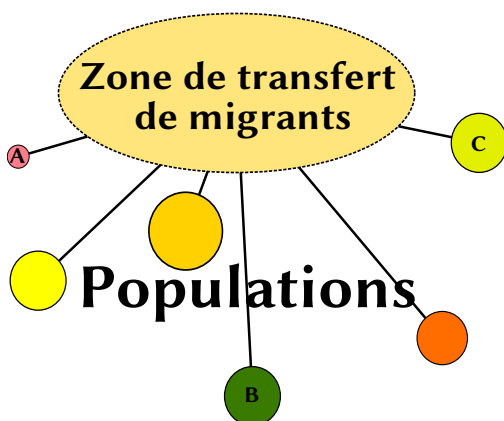


FIG. 7 : Représentation schématique du scénario démographique derrière le modèle  $F$ . Les populations échangent des migrants à travers une "zone de transfert" commune. La distance à cette zone symbolise la force de ces échanges et les couleurs symbolisent la différenciation génétique locale

Le modèle  $F$  est un modèle de génétique des populations développé par BALDING & NICHOLS (1995) et conceptualisé sous ce nom par GAGGIOTTI & FOLL (2010). Le scénario derrière ce modèle est schématisé par la figure 7 : les populations échangent des migrants uniquement à travers une "zone de transfert" commune à toutes les populations (océan, dispersion de pollen à large échelle...). Les petites populations (A) ou les populations avec des taux de migration plus faibles (B) échangent moins de migrants que les autres (C) et deviennent donc plus différenciées.

Ce modèle permet donc de prendre en compte les différences d'histoire démographique entre les populations, qui conduisent à des taux de dérive par population différents. Il constitue donc un modèle de génétique des populations relativement général.

Enfin, ce modèle possède une caractéristique particulière : il permet de calculer une valeur de  $F_{ST}$  pour chacune des populations. Dans ce cas, les valeurs de  $F_{ST}$  représentent, en quelque sorte, la différenciation génétique entre la population focale et les migrants de la “zone de transfert”. Notons que cette zone constitue, conceptuellement, une moyenne pondérée des différentes populations.

## 2.2 Vraisemblance statistique

L’intérêt de ce modèle, outre sa généralité, réside dans le fait qu’il est possible, en le supposant vrai, d’obtenir la distribution attendue des fréquences alléliques de chaque population en fonction de des différents paramètres de ce modèle.

Cette distribution prend la forme d’une distribution de Dirichlet (BEAUMONT & BALDING, 2004). Elle possède deux types de paramètres : les fréquences alléliques “ancestrales” du locus étudié  $i$  ( $p_{i1}, \dots, p_{iK}$  où  $K$  est le nombre d’allèles) et le paramètre  $\theta_j$  qui mesure la dérive que subit la population focale  $j$ . La  $F_{ST}$  au niveau de la population  $j$  (notée  $F_{ST}^j$ ) dépend de ce paramètre  $\theta_j$  par la relation suivante :

$$\theta_j = \frac{1 - F_{ST}^j}{F_{ST}^j} \quad (2.1)$$

Ce modèle permet donc une représentation statistiquement manipulable des fréquences alléliques des populations, ainsi que de la différenciation génétique entre populations (ou plus précisément entre chaque population et l’ensemble des populations).

## 2.3 Pourquoi utiliser ce modèle ?

Puisque le modèle  $F$  est relativement général, qu’il permet de prendre en compte des propriétés de chaque population qui peuvent influencer la dérive (taille, isolement) et qu’il possède une forme mathématique et statistique manipulable, il constitue un modèle très intéressant pour servir de “modèle nul” à un test statistique.

C’est la raison pour laquelle le modèle  $F$  constitue le modèle neutre dans le test de  $F_{ST}$  du logiciel Bayescan (FOLL & GAGGIOTTI, 2008) et d’autres (FOLL & GAGGIOTTI, 2006 ; GAGGIOTTI & FOLL, 2010). La méthode présentée dans ce chapitre étant elle-même basée sur ce logiciel, elle utilise ce même modèle en tant que modèle neutre. Contre ce modèle de neutralité seront testés deux autres modèles : l’un modélise un scénario d’adaptation locale vis-à-vis d’une variable environnementale particulière, l’autre un scénario de différenciation génétique forte pour un locus particulier, mais sans lien avec cette adaptation locale.

## 3 Présentation de l’ARTICLE 2

Basé sur les considérations ci-dessus, nous avons développé une méthode à  $F_{ST}$  qui utilise l’information environnementale. Contrairement aux méthodes d’association classique, notre méthode ne teste pas s’il existe une relation linéaire entre les fréquences alléliques et l’environnement, mais s’il existe une relation linéaire entre la différenciation génétique (mesurée par la  $F_{ST}^j$  locale de la population  $j$ ) et ce que nous appelons la “différenciation environnementale”, c’est-à-dire la distance à un repère environnemental supposé refléter l’environnement non sélectif.

Outre l'originalité de ce test, l'intérêt de cette nouvelle méthode vient de ce que nous avons mis l'accent sur la limitation du taux de faux positifs. En plus du modèle testant une hypothèse d'adaptation locale pour un locus particulier, nous avons introduit un modèle concurrent expliquant la forte différenciation du locus sans utiliser la variable environnementale. Cette astuce permet au modèle de mieux trier entre les locus pour lesquels l'hypothèse d'adaptation locale est très vraisemblable de ceux pour lesquels n'importe quelle hypothèse de forte différenciation entre populations possède une probabilité non négligeable.

### 3.1 Contributions

J'ai conçu la méthode conjointement avec Oscar Gaggiotti. J'ai codé, fait les simulations et distribué la méthode. J'ai analysé les résultats et écrit l'article en collaboration avec Oscar Gaggiotti.

### 3.2 Résumé de l'article

Les méthodes de scan génomique sont utilisées pour passer au crible le polymorphisme génétique de l'intégralité des génomes pour détecter des signatures de sélection. Il y a deux types de méthodes : (i) les méthodes de détection de locus "hors-normes" basées sur la  $F_{ST}$ , qui cherchent à détecter des locus avec une forte différenciation génétique comparé au reste du génome, et (ii) les méthodes d'association environnementale qui cherchent à associer statistiquement les fréquences alléliques et des variables environnementales.

Dans cet article, nous présentons une nouvelle méthode de scan génomique à  $F_{ST}$ , BayeScEnv, qui utilise l'information environnementale sous la forme d'une "différenciation environnementale". Elle est basée sur le modèle  $F$ , mais contrairement aux méthodes existantes, elle considère deux effets possibles pour un locus donné : l'un est causé par la sélection divergente, et l'autre par plusieurs processus différents de l'adaptation locale (comme l'extension d'aire de répartition, les taux de mutations variables entre locus ou la sélection d'arrière-plan). La méthode a été développée en C++ et est disponible à l'adresse <http://github.com/devillemereuil/bayescenv>.

Une étude par simulation montre que notre méthode possède un taux de faux positifs bien inférieur à une autre méthode à  $F_{ST}$ , BayeScan, pour un grand nombre de scénarios démographiques. Bien qu'elle possède une puissance inférieure, elle possède un meilleur compromis entre puissance et taux de faux positifs. Une application à un jeu de données humain montre que cette méthode peut être utilisée avec succès pour détecter de l'adaptation locale. La portée de la méthode est discutée et elle est comparée à d'autres méthodes existantes.



**Journal:** *Methods in Ecology & Evolution*, Early View (2015)  
Les *Supplementary Information* de cet article sont disponibles dans l'Annexe S2

## ARTICLE 2

### A new $F_{ST}$ -based method to uncover local adaptation using environmental variables.

Pierre de Villemereuil & Oscar E. Gaggiotti

#### Abstract

Genome-scan methods are used for screening genome-wide patterns of DNA polymorphism to detect signatures of positive selection. There are two main types of methods: (i) “outlier” detection methods based on  $F_{ST}$  that detect loci with high differentiation compared to the rest of the genome, and (ii) environmental association methods that test the association between allele frequencies and environmental variables. We present a new  $F_{ST}$ -based genome-scan method, BayeScEnv, which incorporates environmental information in the form of “environmental differentiation”. It is based on the  $F$  model, but, as opposed to existing approaches, it considers two locus-specific effects; one due to divergent selection, and another due to various other processes different from local adaptation (e.g. range expansions, differences in mutation rates across loci or background selection). The method was developed in C++ and is available at <http://github.com/devillemereuil/bayescenv>. A simulation study shows that our method has a much lower false positive rate than an existing  $F_{ST}$ -based method, BayeScan, under a wide range of demographic scenarios. Although it has lower power, it leads to a better compromise between power and false positive rate. We apply our method to a human dataset and show that it can be used successfully to study local adaptation. We discuss its scope and compare it to other existing methods.

#### Introduction

One of the most important aims of population genomics (LUIKART et al., 2003) is to uncover signatures of selection in genomes of non model species. Of special interest is the process of local adaptation, whereby populations experiencing different environmental conditions undergo adaptive, selective pressures specific to their local habitat. As a result, populations evolve traits that provide an advantage in their local environment. Many experimental approaches focused on potentially adaptive traits have been developed to test for local adaptation (reviewed in BLANQUART et al., 2013), but only recently it has become possible to make inferences about the genomic regions involved in local adaptation processes. Indeed, the advent of next generation sequencing (NGS, SHENDURE & JI, 2008) has fostered the development of so-called genome-scan methods aimed at identifying regions of the genome subject to selection. These methods are now widely used in studies of local adaptation (FARIA et al., 2014).

There are two main types of genome-scan methods. The first type detects ‘outlier’ loci using locus-specific  $F_{ST}$  estimates, which are compared to either an empirical distribution (AKEY et al., 2002), or to a distribution expected under a neutral model of evolution (BEAUMONT & BALDING, 2004; FOLL & GAGGIOTTI, 2008). The rationale behind these methods is that local adaptation leads to strong genetic differentiation between populations, but only at the selected loci (or marker loci linked to them). Thus, loci with very high  $F_{ST}$  compared to the rest of the genome are suspected to be under strong local adaptation and are referred to as outliers. The outlier approach was further extended to statistics akin to  $F_{ST}$  (BONHOMME et al., 2010; GÜNTHER & COOP, 2013), and also to other unrelated statistics (DUFORET-FREBOURG et al., 2014). One limitation of these methods is that they are not designed to test hypotheses about the environmental factors underlying the selective pressure.

A second type of methods focuses on environmental variables and aims at associating patterns of allele frequency to environmental gradients. The rationale is that selective pressures should create associations between allele frequencies at the selected loci and the causal environmental variables (COOP et al., 2010). In the presence of population structure, performing a simple linear regression would be an error-prone approach (DE MITA et al., 2013; de VILLEMEREUIL et al., 2014). Instead, existing methods account for population structure by modelling the allele frequency covariation across populations (COOP et al., 2010; FRICHOT et al., 2013; GUILLOT et al., 2014).

One disadvantage of most of these approaches is that the parameters that capture the effect of demographic history on genetic differentiation do not have a clear biological interpretation, which in turn makes the rejection of the null model hard to interpret in terms of detection of local adaptation. Especially, note that although the elements of the covariance matrix estimated by COOP et al. (2010) could in principle be interpreted as parametric estimates of the pairwise and population-specific  $F_{ST}$ , this is only true when levels of genetic drift are low (NICHOLSON et al., 2002).

It is important to note that, regardless of the type of genome-scan method under consideration, processes other than local adaptation might be responsible for the observed spatial patterns in allele frequency or  $F_{ST}$ . These include demographic processes such as allele surfing (EDMONDS et al., 2004) or hierarchical population structure (EXCOFFIER et al., 2009), large differences in mutation rate across loci (EDELAAR et al., 2011), hybrid incompatibility following secondary contact (KRUUK et al., 1999) and background selection (CHARLESWORTH, 1998). It is therefore possible that some of the loci identified as outliers are in fact false positives. Accounting for processes other than selection would require introducing parameters that could appropriately capture the effect of these other processes.

Here, we present a method that incorporates features of the two types of genome-scans described above. The objective is to allow inferences about the environmental factors underlying selective pressures, and simultaneously better discriminate between true and false genetic signatures of local adaptation. Note that our new method focuses only on local adaptation driven by a focal environmental variable and, therefore, differs from other  $F_{ST}$ -based methods that carry out "blind" genome scans. Thus, this new approach is aimed at testing hypothesis about specific drivers of local adaptation such as altitude (BIGHAM et al., 2010; FOLL et al., 2014), salinity (LARSSON et al., 2007; DAUB et al., 2013), pathogens (FUMAGALLI et al., 2011; DAUB et al., 2013), etc.

Our method is based on the Bayesian approach first proposed by BEAUMONT & BALDING (2004) and later extended by FOLL & GAGGIOTTI (2008). The original formulation considers population- and locus-specific  $F_{ST}$ 's, which are described by a logistic model with three parameters: a locus-specific term,  $\alpha_i$ , that captures the effect of mutation and some forms of selection, a population-specific term,  $\beta_j$ , that captures demographic effects (e.g.  $N_e$  and migration) and a locus-by-population interaction term,  $\gamma_{ij}$ , that reflects the effect of local adaptation. The estimation of the first two terms benefits from sharing information across loci or populations, but this is not the case for the interaction term, which is therefore poorly estimated (BEAUMONT & BALDING, 2004, but see RIEBLER et al., 2008).

In practice signatures of local adaptation are therefore inferred from the locus-specific effects ( $\alpha_i$ ) under the assumption that large positive values reflect adaptive selection. The implicit assumption is that background selection and mutation should not have much of an effect on this term. In order to relax this assumption and to better estimate the interaction term we introduce environmental data so that  $\gamma_{ij} = g_i E_j$ , where  $E_j$  is the "environmental differentiation" observed in population  $j$  and  $g_i$  is a locus-specific coefficient.

In what follows, we first describe in detail the probabilistic model underlying our Bayesian approach. We then evaluate its performance using simulated data and then present an application using a human dataset. Finally, we discuss the scope of our method and compare it with other existing genome-scan approaches.

## Statistical model

### Modelling allele frequencies using the $F$ model

Our new genome-scan approach is based on the  $F$  model (BEAUMONT & BALDING, 2004; FOLL & GAGGIOTTI, 2008) and extends the software BayeScan (FOLL & GAGGIOTTI, 2008) by incorporating environmental data so as to explicitly consider local adaptation scenarios. Full details of the  $F$  model are given by GAGGIOTTI & FOLL (2010), so here we only provide a brief description. The core assumptions of the  $F$  model is that all populations share a common pool of migrants, but that their effective sizes and immigration rates are population-specific. Thus, population structure at each locus is described by local  $F_{ST}$ 's that measure genetic differentiation between each local population and the migrant pool.

The  $F$  model uses the multinomial-Dirichlet likelihood for the allele counts  $\mathbf{a}_{ij} = (a_{ij1}, \dots, a_{ijK_i})$  at locus  $i$  within population  $j$  (where  $K_i$  is the number of distinct alleles at locus  $i$ ) with parameters given by the migrant pool allele frequencies,  $\mathbf{f}_i = (f_{i1}, \dots, f_{iK_i})$ , and a population- and locus-specific parameter of similarity,  $\theta_{ij} = \frac{1 - F_{ST}^{ij}}{F_{ST}^{ij}}$ :

$$\mathbf{a}_{ij} \sim \text{multDir}(\theta_{ij}f_{i1}, \dots, \theta_{ij}f_{iK_i}), \quad (\text{A2.2})$$

where *multDir* stands for the multinomial-Dirichlet distribution.

Although, for the sake of simplicity, we only present here the formulation for co-dominant data, the software implementing our approach also allows for dominant data (e.g. AFLP markers) using the same probabilistic model as FOLL & GAGGIOTTI (2008). Note finally that, for bi-allelic co-dominant markers (e.g. SNP markers), the likelihood reduces to a beta-binomial model.

### Alternative models to explain population structure

Our purpose is to better discriminate between true signals of local adaptation and spurious signals left by other processes. Therefore, we assume that genetic differentiation at individual loci is influenced by three type of effects: (i) genome-wide effects due to demography, (ii) a locus-specific effect due to local adaptation caused by the focal environmental variable, and (iii) locus-specific effects unrelated to the focal environmental variable. Although in principle one could consider all seven alternative models that can be constructed with different combinations of these three effects, most of them would not have any biological meaning. For example, all models should include genome-wide effects associated with genetic drift. Additionally, we do not consider the two types of locus-specific effects simultaneously in a full model. The reason for this is that the inclusion of  $\alpha_i$  along with  $g_i$  is not justified biologically. This is because the joint effect of local adaptation and another locus-specific effect such as allele surfing or background selection on the same locus is extremely unlikely either because of the strong effect of genetic drift in the first instance or the implausibility of a favourable variant arising and increasing in frequency in a highly conserved region subject to strong purifying selection. Thus, we focus on three different models to explain the genetic structuring at individual loci.

#### ► Null model of population structure

Under the null hypothesis that all loci are neutral, the local differentiation parameter  $F_{ST}^{ij}$  will be driven only by local population demography and, hence, should be common to all loci:

$$\log \left( \frac{F_{ST}^{ij}}{1 - F_{ST}^{ij}} \right) = \log \left( \frac{1}{\theta_{ij}} \right) = \beta_j. \quad (\text{A2.3})$$

A high  $\beta_j$  value means that the population  $j$  is strongly differentiated from the pool of migrants. This could be due to a lack of immigration from the other populations, a reduced effective size, or a particular spatial structure.

### ► Alternative model of local adaptation

In this model, we focus on a particular signature left by a process of local adaptation. If selection is driven by a putative environmental factor, we expect that genetic differentiation for the locus or loci under selection will be stronger than expected under neutrality for populations with strong environmental differentiation. Any measure of distance between the environmental value of population  $j$  and the average environment could serve as a measure of differentiation. For the sake of simplicity, we here only consider the absolute value. Furthermore, in order to facilitate the calibration of prior distributions, we consider standardised environmental values with unit variance.

To model the effect of local adaptation on locus  $i$ , we consider the impact of environmental differentiation  $E_j$  of population  $j$  on the locus, we thus modify EQ. A2.3 as follows:

$$\log\left(\frac{F_{ST}^{ij}}{1 - F_{ST}^{ij}}\right) = \beta_j + g_i E_j, \quad (\text{A2.4})$$

where  $g_i$  quantifies the sensitivity of locus  $i$  to the environmental differentiation.

### ► Alternative model of locus-specific effect

Local adaptation with respect to the focal environmental variable is not the only evolutionary phenomenon that could lead to departures from the neutral model. Other phenomena that could produce such locus-specific effects include local adaptation due to other unknown factors, large differences in mutation rate across loci, the so-called allele surfing phenomenon (EDMONDS et al., 2004) and background selection (CHARLESWORTH, 2013).

This is accounted for by using the following parametrisation for local differentiation:

$$\log\left(\frac{F_{ST}^{ij}}{1 - F_{ST}^{ij}}\right) = \alpha_i + \beta_j. \quad (\text{A2.5})$$

The main advantage of implementing both of the above alternative models is that we can distinguish between departures from the neutral model of unknown origin (using EQ. A2.5) and departures due to local adaptation caused by a particular environmental factor (using EQ. A2.4).

## Material & Methods

### Implementation of the statistical model

Our method, summarised in FIG. A2.1, uses two types of data: (i) the allele counts  $\mathbf{a}$  for each locus in each population sample, and (ii) observed values  $E$  of an environmental variable (one value per population), which are transformed into environmental differentiation using an appropriate function. Indeed, our model aims at associating genetic distance (i.e. the  $F_{ST}^{ij}$ ) with an environmental distance. Note that measuring an environmental distance requires to define a reference. The most natural reference would be the average of the environmental values, but this would not be always the case (see the example of adaptation to altitude in humans presented below). Also, it is strongly advised to standardise the environmental values by dividing by the standard deviation, in order to avoid effect size issues regarding the inference of the parameter  $g$ .

As stated in the previous section, there are three different models:

**M1** Neutral model:  $\beta_j$ ,

**M2** Local adaptation model with environmental differentiation  $E_j$ :  $\beta_j + g_i E_j$ ,

**M3** Locus-specific model:  $\alpha_i + \beta_j$ .

Note that in our framework, the focal model being tested against the two others is **M2**. Thus, power and error rates (FPR and FDR) are computed for model **M2**. Model **M3** can be considered as a “nuisance model” whose role is to reduce the overall false positive rate by explaining the inflation of the variance in  $F_{ST}$  due to locus-specific effects other than selection driven by the focal environmental factor.

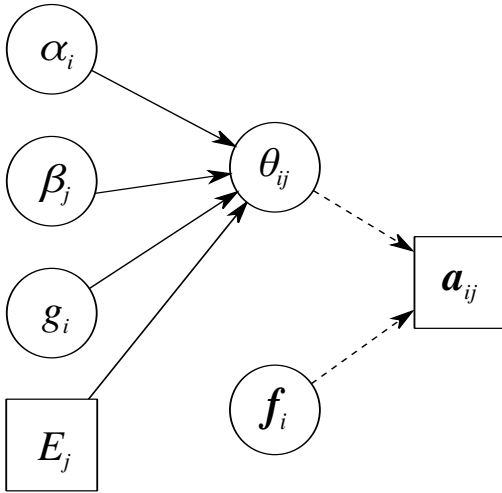


FIG. A2.1: Directed Acyclic Graph (DAG) of the model. Squared nodes denote known quantities ( $E$  for environmental data, and  $A$  for genetic marker data). Circled nodes denote unknown parameters. Plain arrows stand for deterministic relationships, and dashed arrows stand for stochastic relationships.

The mathematical details of the transition between models can be found in the Supplementary Information S2 (sections S2.1, S2.2 and S2.3). Pilot studies showed that using values of  $p$  above 0.5 yielded extremely conservative results (note that setting  $p = 1$  would mean that model **M3** is always favoured over **M2**, in which case the power of the method is zero, yielding no positives whatsoever). We used a uniform Dirichlet prior for the allele frequencies  $f_i \sim Dir(1, \dots, 1)$ . The priors for the hyperparameters  $\alpha$  and  $\beta$ , were Normal with mean -1 and variance 1 (note that the results of our method will be especially sensitive to the prior mean of  $\alpha$ , but our pilot studies showed that -1 was a good default). Since under a local adaptation scenario the parameter  $g$  is only expected to be positive, it was assigned a uniform prior between 0 and 10.

Our method outputs posterior error probabilities and  $q$ -values, which are test statistics related to the False Discovery Rate (FDR) (STOREY, 2002; KÄLL et al., 2008). Contrary to the commonly used False Positive Rate (FPR), which is the probability of declaring a locus as positive given that it is actually neutral, the FDR is the proportion of the positive results that are in fact false positives, and is more appropriate for multiple testing (KÄLL et al., 2008). See the Supplementary Information S2.3 for more details.

## Simulation analysis

We performed a simulation study to evaluate the performance of our method and compare it with that of BayeScan (FOLL & GAGGIOTTI, 2008). We modelled 16 populations each with 500 individuals genotyped at 5,000 loci, among which one (monogenic scenario) or 50 (polygenic scenario) were under selection. We modelled three kinds of population structure: (i) a classical island model (IM), (ii) a one-dimension stepping-stone (SS) model and (iii) a hierarchically structured (HS) model.

Hence, the statistical significance of the parameter  $\alpha_i$  is not of interest for BayeScEnv: only the significant values of  $g_i$  are considered. All three models were implemented using an RJMCMC algorithm (GREEN, 1995). In order to propose relevant values for new parameters during the jumps, the RJMCMC is preceded by pilot runs. These are aimed at both calibrating the MCMC proposals to reach efficient acceptance rates, and approximating the posterior distribution of parameters, as proposed by S. BROOKS (1998) and already implemented in BayeScan (FOLL & GAGGIOTTI, 2008). Our code is based on the source code of BayeScan 2.1 and is written in C++. The source and binaries are available at <https://github.com/devillemereuil/bayescenv>.

Our prior belief in the three models is described by two parameters: the probability  $\pi$  of moving away from the neutral model and the preference  $p$  for **M3** against **M2** as alternative models. We can calculate the prior probability for each model as:

$$\begin{aligned} P(\mathbf{M1}) &= 1 - \pi, \\ P(\mathbf{M2}) &= \pi(1 - p), \\ P(\mathbf{M3}) &= \pi p. \end{aligned} \quad (\text{A2.6})$$

The genome was composed of 5,000 bi-allelic SNPs spread along 10 chromosomes. The loci under selection due to an environmental variable  $E$  (see FIG. S2.2 and EQ. S2.7 and S2.8), one for the monogenic case and 50 for the polygenic case, were randomly distributed across the genome. Since all markers were independently initialised, our simulations yielded negligible linkage disequilibrium. Consequently, we considered as true positives only the loci subject to selection. For the IM and SS scenarios, we directly initialised all 16 populations. For the HS scenario, we initialised the ancestral population, which, following successive and temporally spaced-out fission events, gave rise to 2, 4, ..., 16 populations. This hierarchical structure is reinforced by preferential migration between related populations. More details regarding migration and population history are available in the SI. This model is very close to that used by de VILLEMEREUIL et al. (2014). It should be particularly difficult for our method, because all populations are equally differentiated (i.e. the  $\beta_j$  parameters are expected to be roughly the same across populations), but a phylo-geographic covariance exists between related populations, which is not explicitly accounted for by our probabilistic model. More information regarding the environmental gradient and the fitness function are available in the SI, but, briefly, a polygenic multiplicative model was used with a selection strength of 0.02 (0.1 for the monogenic case).

The simulations were performed using the SimuPOP Python library (PENG & KIMMEL, 2005) and the scripts are available online in the data section. Our simulated datasets were analysed using our C++ code and version 2.1 of BayeScan (FOLL & GAGGIOTTI, 2008).

We generated 100 datasets for each scenario and computed the realised FDR, FPR and power yielded by BayeScan and our new environmental method (BayeScEnv). For the latter, we also compared several parametrisations using a prior probability  $\pi$  of jumping away from the neutral model of 0.1 (equivalent to the default prior odds used by BayeScan, which is 10) or 0.5, as well as a preference for the locus-specific model  $p$  of 0.5 (environmental and locus-specific models are equiprobable) or 0 (the locus-specific model is forbidden and only the environmental model is tested against the neutral one).

We supplemented these scenarios with a heterogeneous mutation rate case, based on the IM scenario above, where most of the genome had a high mutation rate of 0.05, whereas 50 loci had a low mutation rate of  $10^{-7}$ . The result was an overall low  $F_{ST}$  of 0.05 for the whole genome, and of 0.10 for the low mutating loci.

## HGDP SNP data analysis

In order to test our new method against a real dataset, we focused on 26 Asian populations from the Human Genome Diversity Panel (HGDP) SNP Genotyping data. This data set consists of 660,918 SNP markers genotyped using Illumina 650Y arrays. After cleaning the dataset from mitochondrial and sex-linked markers, we removed all markers with minor allele frequency below 5%. This left us with a total of 446,117 SNPs. For all populations, we obtained the following environmental variables from the BIOCLIM database: mean annual temperature, precipitation, and altitudinal data. We ran separate BayeScEnv analysis for each variable and compared the results with BayeScan (which doesn't use environmental variables). After standardisation of the environmental variables, we computed environmental differentiation from the mean for temperature and precipitation, and from the sea level for elevation. Gene ontology enrichment tests for the detected genes were performed using the "SNP mode" of the Gowinda software (KOFLER & SCHLÖTTERER, 2012). The prior odds for BayeScan was 10 for this analysis. BayeScEnv prior parameters for this analysis were  $\pi = 0.1$  and  $p = 0.5$ .

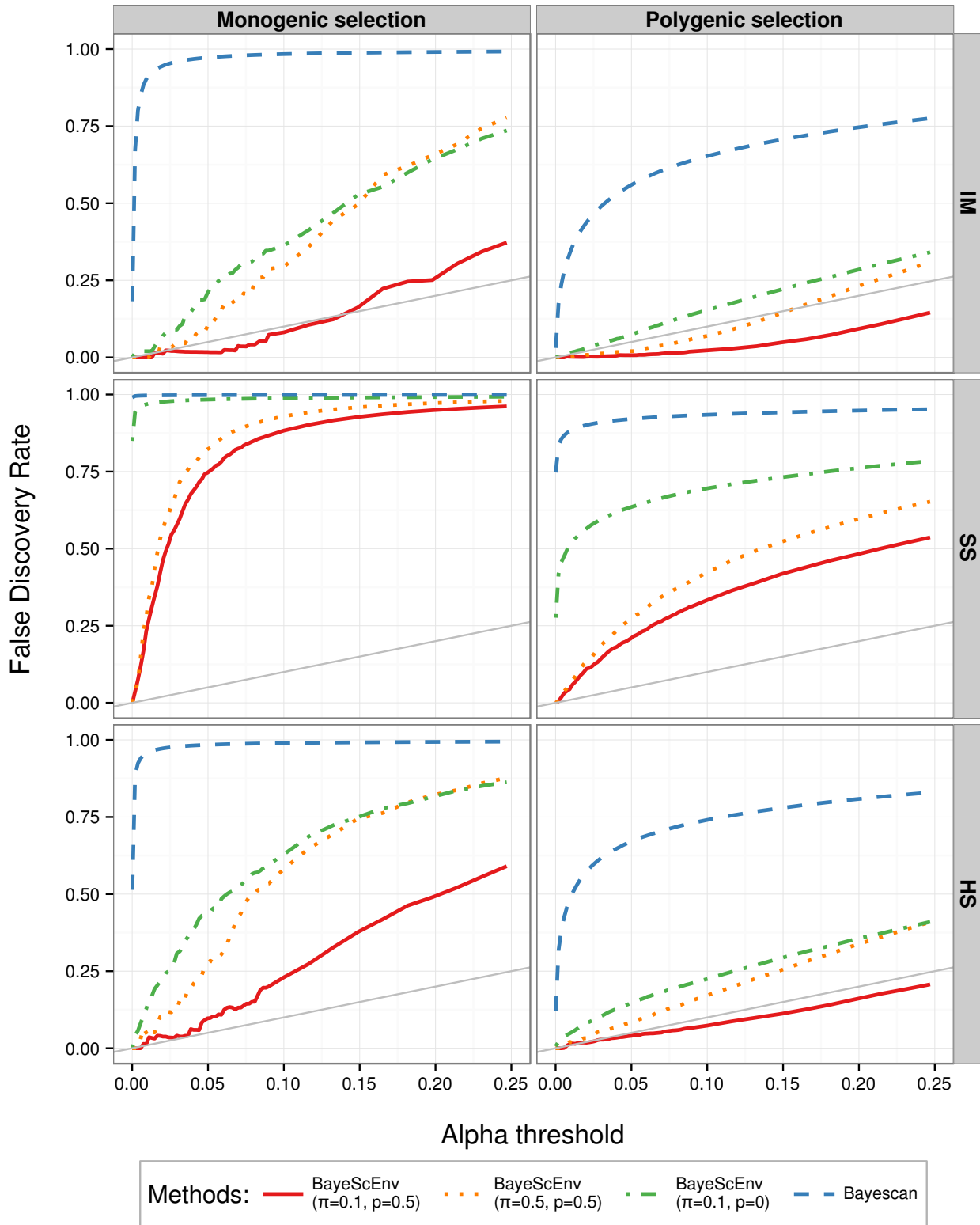


FIG. A2.2: False Discovery Rate (FDR) against significance threshold  $\alpha$  for three scenarios (IM: Island model, SS: Stepping-Stone model and HS: Hierarchically Structured model) and monogenic/polygenic selection. The grey line is the expected identity relationship between the FDR and  $\alpha$ . The models tested are BayeScan (blue dashed), and BayeScEnv (orange dotted, green dot-dashed and solid red) with different probabilities  $\pi$  of jumping away from the neutral model ( $M_1$ ) and different preferences  $p$  for the locus-specific model ( $M_3$ ). Note that  $p = 0$  means the environmental model ( $M_2$ ) is tested against the neutral one only.

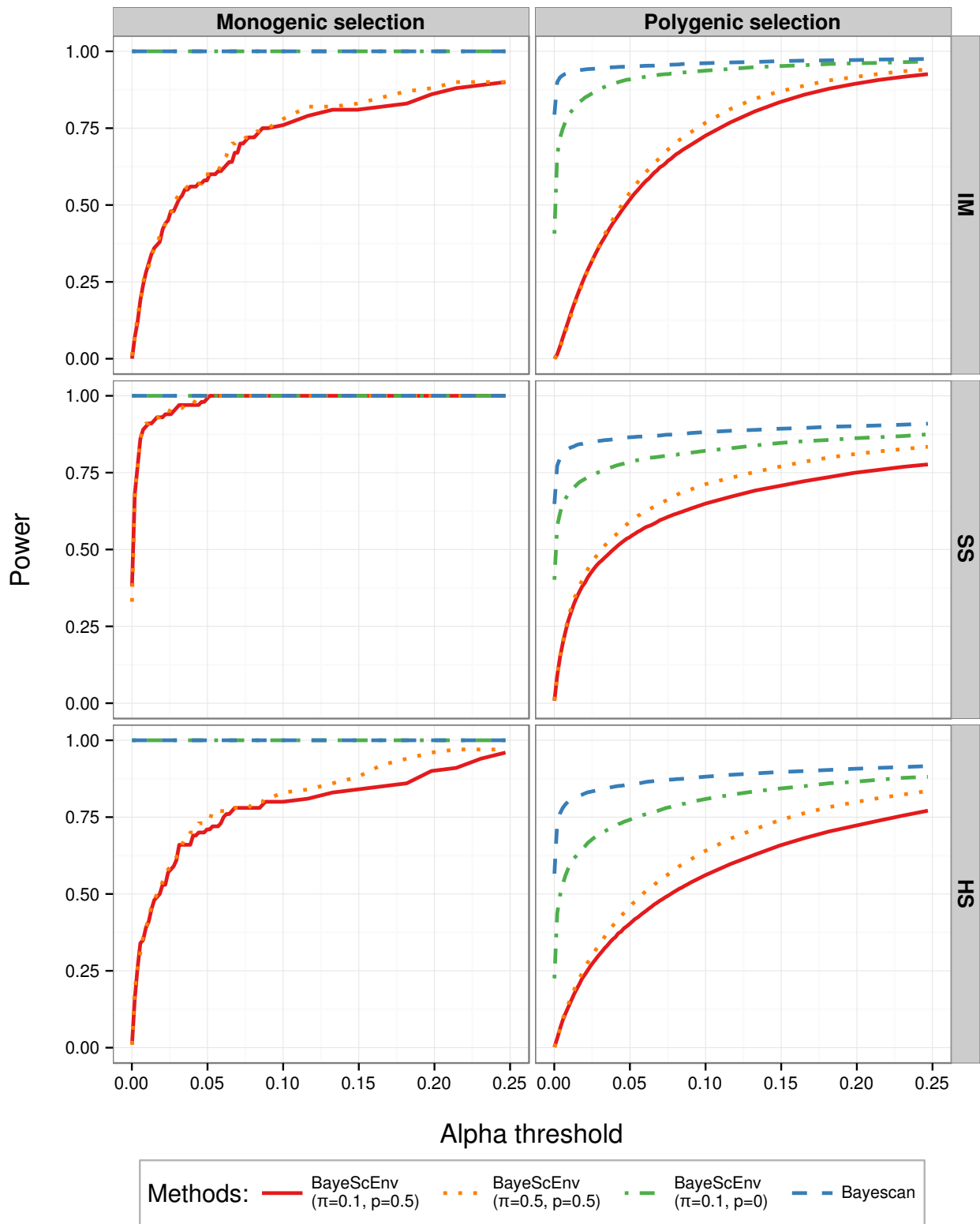


FIG. A2.3: Power against significance threshold  $\alpha$  for three scenarios (IM: Island model, SS: Stepping-Stone model and HS: Hierarchically Structured model) and monogenic/polygenic selection. The models tested are Bayescan (blue dashed), and BayeScEnv (orange dotted, green dot-dashed and solid red) with different probabilities  $\pi$  of jumping away from the neutral model (M1) and different preferences  $p$  for the locus-specific model (M3). Note that  $p = 0$  means the environmental model (M2) is tested against the neutral one only.



## Results

### Simulation results

By definition, a threshold value of  $\alpha$  used to decide whether  $q$ -values are significant or not is expected to yield an FDR of  $\alpha$  on the long run, when the model is robust and priors are calibrated. Recall that in BayeScEnv all  $q$ -value tests below were performed on the parameter  $g$  to test for local adaptation. In the case of BayeScan, on the other hand, the  $q$ -values correspond to parameter  $\alpha$ .

As shown in FIG. A2.2, BayeScan was less well calibrated, yielding higher FDRs than BayeScEnv under all scenarios and for both monogenic and polygenic selection. Additionally, for BayeScEnv, the implementation using  $\pi = 0.1$  was fairly well calibrated (i.e. the curve is close the grey line in FIG. A2.2) under the IM scenario (for both monogenic and polygenic versions) and under the polygenic version of the HS scenario. This implementation was much more conservative than the one using  $\pi = 0.5$ . For  $\pi = 0.1$  and  $p = 0$ , the FDRs were closer to those yielded by BayeScan, but still lower.

The higher FDR for BayeScan and BayeScEnv with  $\pi = 0.5$  or  $p = 0$  was mainly driven by a higher FPR rather than a lack of power (FIG. A2.3, see also FIG. S2.3 in the Supplementary Information). Notably though, BayeScan had a quite high power, higher than that of BayeScEnv. Note, however, that BayeScEnv with  $p = 0$  had, as BayeScan, a maximal power in the monogenic scenarios, and was almost as powerful as BayeScan in the polygenic scenarios. Yet its FDR was lower (sometimes much lower) than that of BayeScan. This indicates that the incorporation of environmental data helps to reduce the error rate both with or without the inclusion of spurious locus-specific effects ( $\alpha_i$ ). More details regarding the FPR results are available in the Supplementary Information (FIG. S2.3).

Another traditional way to apprehend the compromise between power and false positives is the so-called Receiver Operating Characteristics (ROC) curve, plotting power against FPR (FIG. A2.4). In these plots, the curve that is “more to the left” is preferred because this means it offers higher power for a lower FPR. FIG. A2.4 shows that BayeScEnv with  $\pi = 0.1$  and  $p = 0$  performed best under the IM and HS scenarios, whereas BayeScEnv with  $\pi = 0.1$  and  $p = 0.5$  performed better under the “harder” SS scenario. Overall, although BayeScan has higher power to detect local adaptation, it is still too liberal when deciding that a locus is under selection for the scenarios we investigated.

The heterogeneous mutation scenario lead to a dramatically high false positive rate for the low mutating loci in the case of Bayescan (62%). BayeScEnv, on the other hand, yielded a much lower false positive rate for these loci (4.9%). Of course, because the higher differentiation due to low mutation rate can be seemingly distributed according to the environmental variable, higher false positive rates will always be expected in such a scenario. Nevertheless, BayeScEnv is an improvement over Bayescan in that regard.

### Analysis of human data from Asia

The results of the human dataset analysis (TAB. A2.1) show a dramatic discrepancy between the two methods. Whereas BayeScan yields a very large number (66,316) of markers considered as significant at the 5% threshold, many fewer markers (154 to 2728) are considered significant by BayeScEnv. Gene Ontology (GO) enrichment tests identified many significant terms (TAB. A2.1). Note, however, that in the altitude and temperature analyses they correspond to a small number of genes (11 and 20 respectively, see TAB. A2.1). The number of genes is larger for the precipitation analysis (359) and even larger for the analysis using BayeScan (5628).

Regarding the altitude, significant biological processes included the fatty acid metabolism (e.g. SCARB1), skin pigmentation (e.g. MLANA, SLC24A5), kidney activity (e.g. SLC12A1) and oxidoreductase activity (e.g. NOS1AP). Regarding the temperature, significant biological process included cardiac muscle activity (e.g. SLC8A1) and development (e.g. NRG1, FOXP1), fatty acid metabolism (e.g. FADS1, FADS2) and response to hypoxia (e.g. SLC8A1, SERPINA1). For the precipitation analysis

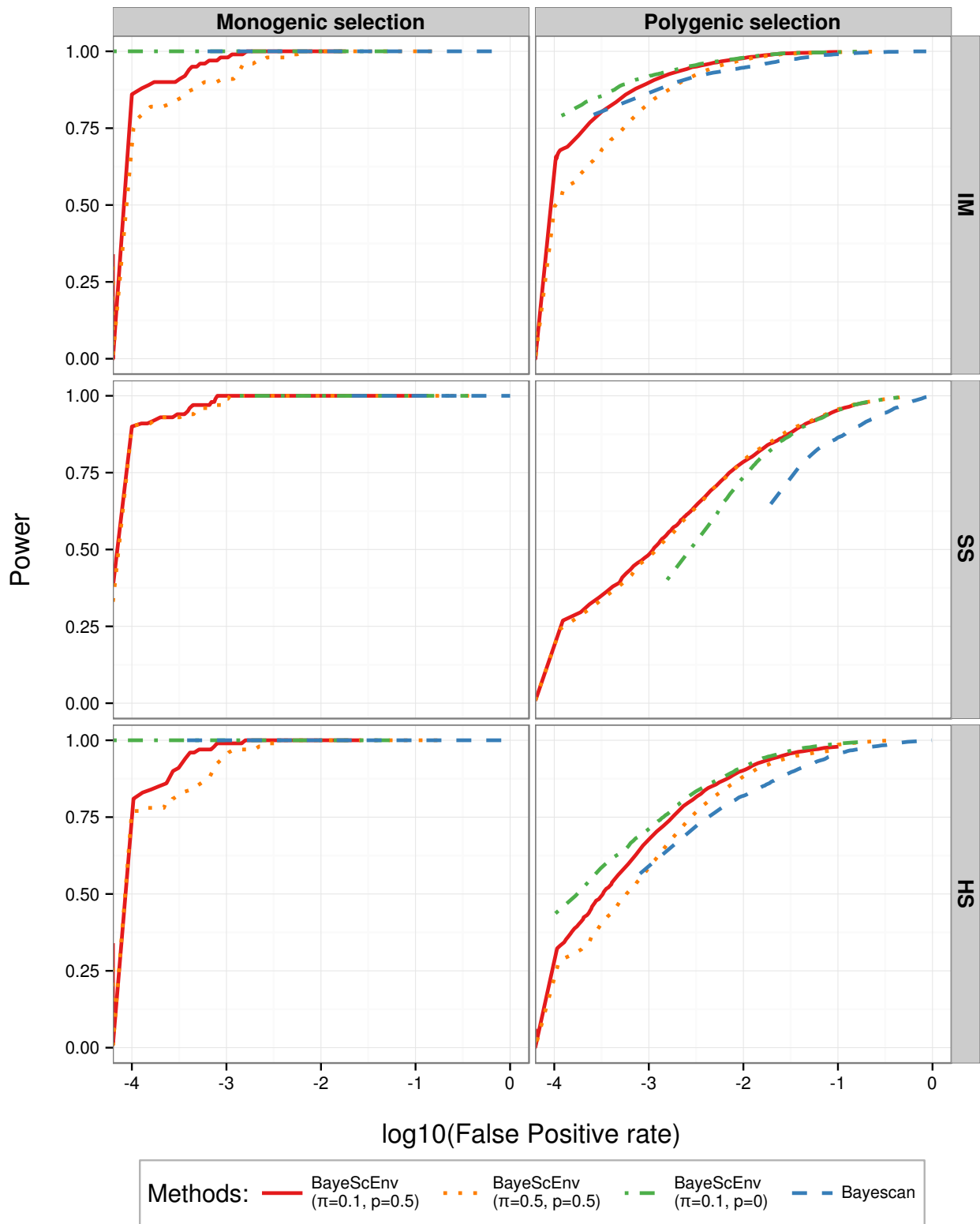


FIG. A2.4: Power against False Positive Rate (FPR), a.k.a. ROC curve, for three scenarios (IM: Island model, SS: Stepping-Stone model and HS: Hierarchically Structured model) and monogenic/polygenic selection. The models tested are BayeScan (blue dashed), and BayeScEnv (orange dotted, green dot-dashed and solid red) with different probabilities  $\pi$  of jumping away from the neutral model (M1) and different preferences  $p$  for the locus-specific model (M3). Note that  $p = 0$  means the environmental model (M2) is tested against the neutral one only.

**Tab. A2.1:** Results from BayeScan and BayeScEnv on the human dataset. FDR significance threshold was set to 5%. The total number of tested markers was 446,117.

Method	Variable	Nr of significant SNPs	Nr of significant GO terms	Nr of genes associated with a significant GO term
BayeScEnv	altitude	154	32	11
	temperature	170	103	20
	precipitation	2728	439	359
BayeScan	—	66,316	469	5,628

with BayeScEnv, as well as the BayeScan analysis, the number of significant terms was too large for hand-picked examples to be feasible.

The significance results ( $q$ -values) are displayed as a Manhattan plot in FIG. A2.5, along with the above mentioned genes for the altitude and temperature analyses (FIG. A2.5, A and B). Other regions of the genome also include outlier loci but they correspond to non-coding regions, or are close to genes associated to GO terms that were not significant, or to proteins without a known function (e.g. C9orf91, which was the most significant gene in the temperature analysis). Pattern of linkage disequilibrium was visible, which sometimes strongly supported some candidate genes (FIG. A2.5, A, SLC12A1 and SLC24A5). Finally, comparing BayeScEnv (FIG. A2.5, A, B and C) and BayeScan analyses (FIG. A2.5,D), we see that BayeScan yielded too many significant markers for a Manhattan plot to be a useful display of the results. An interesting pattern is that BayeScan yielded far more outlier markers with maximal certainty (e.g. posterior probability of one) than BayeScEnv. For the present dataset, 22,516 markers had a posterior probability of one, whereas the maximal posterior probability yielded by BayeScEnv was 0.9998. Finally, almost all loci detected using BayeScEnv were also found when using BayeScan (between 98% for altitude to 100% for the two other variables).

## Discussion

### Features and performance of the method

The method we introduce in this paper, BayeScEnv, has several desirable features. First, just as BayeScan, it is a model-based method. This means that the null model can be understood in terms of a process of neutral evolution. One can thus predict what the method is able to fit or not. Second, we explicitly model a process of local adaptation caused by an environmental variable. Third, in order to render the model more robust, we account for locus-specific effects unrelated to the environmental variable under consideration. These departures can be due to another process of local adaptation (i.e. caused by unknown environmental variables), to large differences in mutation rates across loci, to background selection (CHARLESWORTH, 2013) or complex spatial effects, such as allele surfing (EDMONDS et al., 2004) and hierarchical population structure (EXCOFFIER et al., 2009). Our simulation results show that when compared to BayeScan, BayeScEnv has a better control of its false discovery rate under various scenarios (FIG. A2.2), yielding fewer, but more reliable candidate markers. Obviously, this has a cost in terms of absolute power (FIG. A2.3), but BayeScEnv still performs better than BayeScan in terms of the investigated compromises between true and false positives (i.e. FDR and ROC, FIG. A2.2 & A2.4).

Besides, the parametrisation of BayeScEnv allows for a fine and intuitive control of the false positive rate and power. For example, setting  $p$  to 0 increases both power and false positive rate, whereas setting  $p = 0.5$  will allow for a more conservative test. This is because with  $p = 0$ , that is when the locus-specific effect model (M3) is excluded, the local adaptation model (M2) will absorb much of the signal in the data, yielding a higher probability of detecting true positives, but also a higher sensitivity to false positives. Our simulation results show that, if the species under study has moderate to large dispersal abilities (c.f. hierarchical structure or island model), the former parametrisation will

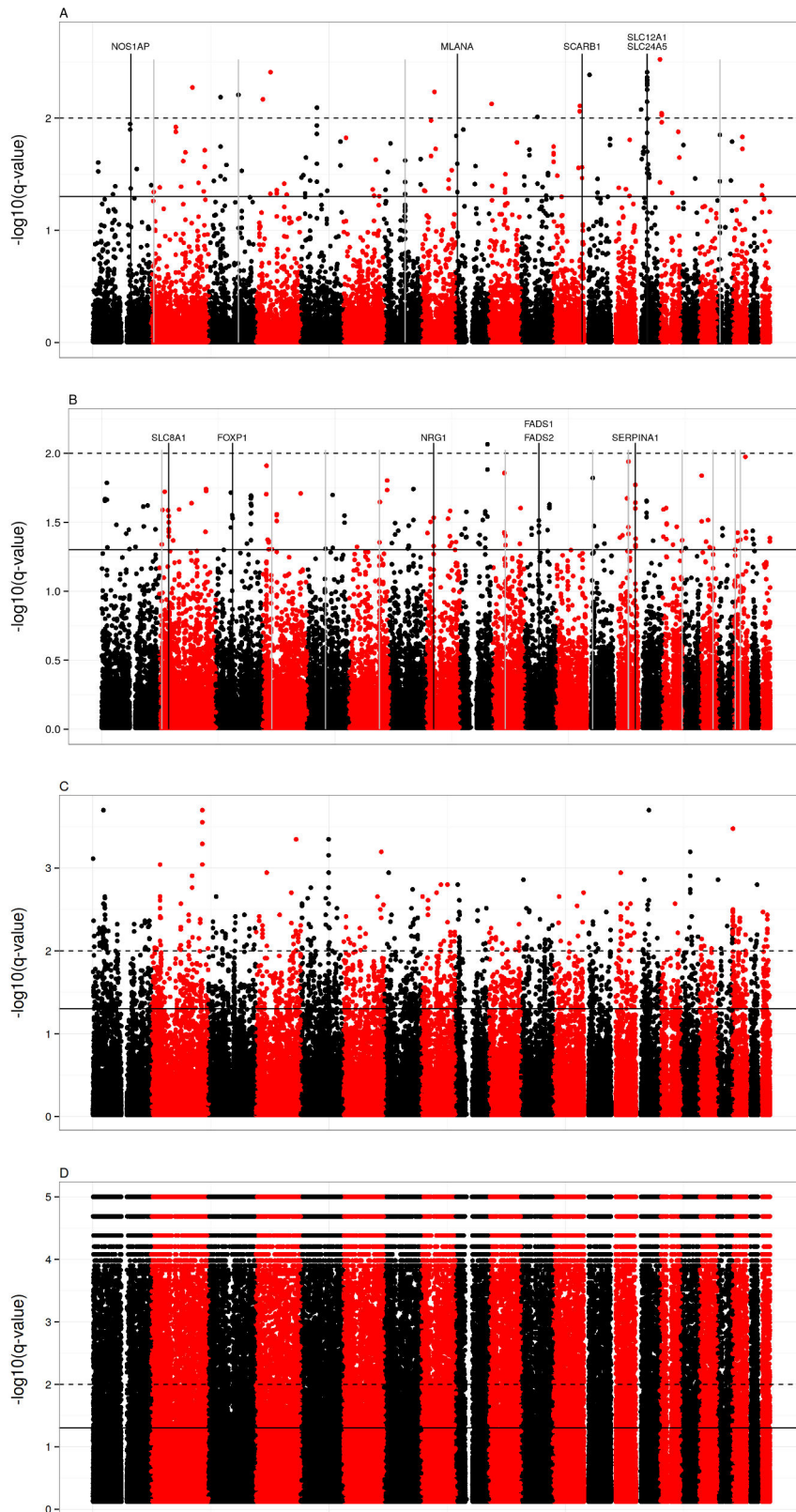


FIG. A2.5: Manhattan plot of the  $q$ -values for the human dataset when using BayeScEnv with altitude (A), temperature (B), precipitations (C) or when using BayeScan (D). For altitude and temperature (A and B), genes mentioned in the text are displayed using black lines and genes associated with a significant GO term using grey lines. Top “stripes” for BayeScan (D) are artefacts due to finite number of iterations in RJMCMC (e.g. 0, 1, 2, 3... iterations outside of the non-neutral model), corresponding to determined posterior probabilities when divided by the total number of iterations.

be more appropriate, whereas for species with low dispersal abilities (c.f. stepping-stone model) the latter should be preferred. Thus, being able to choose the right parametrisation only requires limited knowledge about the dispersal abilities of the species.

We note that BayeScan was recently extended to consider species with hierarchical population structure (BayeScan3, FOLL et al., 2014). With BayeScan3 it is now possible to study widely distributed species covering several continents or geographic regions. It is also possible to better focus on local adaptation by considering groups that include pairs of populations inhabiting different environments such as low and high altitude habitats. Thus, BayeScan3, allows for the consideration of categorical environmental variables. Our new approach on the other hand, allows the study of local adaptation related to continuous environmental variables in species with a more restricted range.

## How to quantify ‘environmental differentiation’?

To model local adaptation, we compute an “environmental differentiation” in terms of the distance (absolute value) to a reference value. Although this reference can conveniently be chosen as the average of the environmental values across the sampled populations, other kinds of reference may be biologically more relevant. For example, in our analysis of the effect of elevation in humans, it seems appropriate to use sea level as the reference. Indeed, given the kind of environmental variables elevation is a proxy for (e.g. partial pressure of oxygen, temperature, solar radiation, etc.), for most systems we would consider the sea level as a neutral environment rather than the differentiated one.

Another way to account for environmental differentiation is to use Principal Component Analysis (PCA), providing one of the axes to BayeScEnv as a description of the distance between environments. Despite this practice being an elegant way to summarise environmental distance between populations, it also has the drawback of making it more difficult to identify the “causal” variable.

Note that the environmental variables must be standardised so as to avoid scale inconsistencies between  $g$  and  $\alpha$  and  $\beta$ . If we choose the average environmental value as reference, then standardisation involves mean-centring and rescaling to have unit variance. However, if we choose another reference, then standardisation only involves rescaling to have unit variance.

Finally, the software implementation of our method only accepts one environmental variable at a time as including more than one variable would considerably slow the algorithm down, and render the biological interpretation of  $g$  quite tedious. Also, when using several correlated variables, it is important to realise that statistically distinguishing between the relative selective roles of each one would require many populations.

## Comparison with other environmental association methods

There are several genome-scan approaches that incorporate environmental information, such as Bayenv (COOP et al., 2010), LFMM (FRICHOT et al., 2013) and gINLAnd (GUILLOT et al., 2014). These methods perform a regression between allele frequencies and environmental values. Yet non-equilibrium situations combined with complex spatial structuring can lead to spatial correlations in allele frequencies, which in turn can lead to high false positive rates. To minimise this problem, the above methods take into account allele frequency correlations across populations while performing the regression.

BayeScEnv, on the other hand, assumes that all populations are independent, exchanging genes only through the migrant pool. However, it includes a locus-specific effect unrelated to the environmental variable that helps to take into account locus-specific spatial effects due to deviations from the underlying demographic model. The fact that this approach works is illustrated by our simulation study, which showed that BayeScEnv was fairly robust to isolation-by-distance and a hierarchically structured scenario. Moreover, the analyses of simulated datasets from de VILLEMEREUIL et al. (2014), available in the SI, show that even under very complex scenarios, BayeScEnv can compete

with other environmental association methods. In particular, most of these scenarios assume an environmental selective gradient confounded with population structure, which is particularly hard for genome scan methods (FRICHOT et al., 2015) : the results show that, in that case, BayeScEnv suffer from low power, but not from an excess of false positives. When compared with the other methods (including Bayenv and LFMM), BayeScEnv typically yields a medium FDR for most scenarios, and is less scenario-sensitive than Bayenv and LFMM. Nevertheless, we note that BayeScEnv is best suited for species with medium to high dispersal abilities such as marine species and anemophilous plants.

Another point that distinguishes BayeScEnv from these methods is that it does not assume any particular functional form for the relationship between environmental values and allele frequencies. While existing association methods all assume a clinal pattern, BayeScEnv only assumes that genetic differentiation increase exponentially with environmental differentiation. This allows for a more diverse family of relationships between allele frequencies and the environment.

Finally, BayeScEnv is one of the very few methods to study gene-environment associations that can be used with dominant data (but see also GUILLOT et al., 2014).

## Data analysis

When confronted with real datasets, BayeScEnv typically returned fewer significant markers than BayeScan. This is explained both by the focus on searching for outliers linked to a specific environmental factor and by the lower false positive rate of our approach. When applied to the human dataset, BayeScEnv identified several genomic regions that are enriched for gene ontology terms relevant to potential local adaptation to altitude or temperature. We emphasise that this study was not meant to exhaustively and rigorously investigate local adaptation in Asian human populations. However, our results tend to demonstrate that the candidates yielded by BayeScEnv have a biological interpretation. For example, skin pigmentation and cardiac activity could clearly be involved in responses to increased solar radiation and depleted oxygen availability at high elevation.

Much of the ontologies linked to temperature were potentially confounded with adaptation to altitude, such as the response to hypoxia and cardiac muscle activity. Also, fatty acid metabolism was associated to both altitude and temperature. Of course, the biological functions described here do not account for all the signals yielded by BayeScEnv (see FIG. A2.5, A and B). Other genomic significant regions include genes with less obvious biological function regarding local adaptation, non-coding regions and proteins without a known function. Finally, the analysis using the precipitation variable yielded too many significant markers for a detailed analysis of the biological functions involved. This may not necessarily be due to a confounding effect of the spatial structure (the human Asian populations being structured mainly from West to East, while the Eastern climate is characterised by strong precipitations during the monsoon), since precipitation may behave as a surrogate for several environmental variables.

## Conclusion

The main improvement introduced by our new method, BayeScEnv, over existing  $F_{ST}$ -based genome-scan approaches is the possibility of focusing on the detection of outlier loci linked to genomic regions involved in local adaptation and better distinguishing between the signal of positive selection and that of other locus-specific processes such as mutation (see the heterogeneous mutation rate scenario in the Results) and background selection. Although it does not explicitly model complex spatial effects, the consideration of two different locus-specific effects make it more robust to potential deviations from the migrant pool model. This is reflected in its much lower false discovery rate when compared to BayeScan.

Our new formulation also allows for an improved control of the true/false positives compromise through the parameter  $p$ , which describes our preference for the model that includes a locus-specific effect unrelated to the environmental factor over the model that includes environmental effects.

Although we recommend using  $p = 0.5$ , lower values (including 0) could be used if population structure is weak or maximising power is more important than reducing the false positive rate.

With this new method, there are now three alternative formulations of genome-scan methods based on the  $F$  model. BayeScan detects a wide range of locus-specific effects (including background selection). Although its false discovery rate is higher than that of the two extensions, it is able to detect regions of the genome subject to purifying selection. The hierarchical version of this original formulation, BayeScan3, allows the study of local adaptation due to categorical environmental factors. Finally, our new method, BayeScEnv, is more appropriate to detect genomic regions under the influence of selective pressures exerted by continuous environmental variables. Thus, all three methods are complementary and jointly cover scenarios applicable to a wide range of species

## Acknowledgement

We thank M. Foll for providing the source code of BayeScan and for clarifying several issues related to the code, J. Renaud for his help on getting the average altitude out of the HGDP latitude/longitude data, S. Schoville for the BIOCLIM data, E. Bazin for his help on the HGDP data analysis. PdV was supported by a doctoral studentship from the French *Ministère de la Recherche et de l'Enseignement Supérieur*. OEG was supported by the Marine Alliance for Science and Technology for Scotland (MASTS).

## Data Accessibility

The Python code used to simulate data is available online in the Supplementary Information. The software and its source code are available online at GitHub: <http://github.com/devillemereuil/bayescenv>. The HGDP dataset is available at <http://www.hagsc.org/hgdp/files.html>. The BIOCLIM database is available at <http://worldclim.org/bioclim>.

## Author contributions

PdV and OEG designed the statistical model. PdV modified the C++ code and performed the simulation and data analysis. PdV and OEG wrote the article.

## 4 Caractéristiques de la nouvelle méthode

### 4.1 Puissance et taux d'erreur

Les analyses présentées dans cet article montrent plusieurs choses. Tout d'abord, la méthode fonctionne et est capable de retrouver les locus sous sélection même dans des conditions relativement difficiles (sélection polygénique par exemple). Ensuite, comme le montre l'analyse par simulation, mais surtout sur l'Humain, BayeScEnv possède un taux de faux positifs très inférieur à celui de Bayescan. Cette amélioration possède un coût en terme de puissance, mais les analyses montrent que ce coût est assez réduit pour que le compromis entre puissance et faux positifs soit en faveur de la nouvelle méthode.

D'autres part, la comparaison aux autres méthodes en Annexe (S2, section S2.8) montre que les performances de BayeScEnv sont comparables aux autres, bien que, là, on retrouve un manque de puissance dans les scénarios où celle-ci est généralement faible.

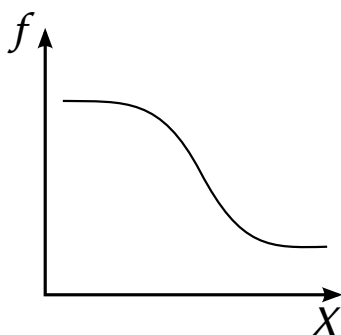
D'un point de vue statistique, les taux d'erreur sont généralement tellement forts pour ces méthodes qu'il paraît logique de limiter les risques de faux positifs le plus possible, quitte à sacrifier un peu de puissance. En d'autres termes, il vaut mieux qu'une méthode renvoie peu de locus dont une bonne proportion seront positifs plutôt qu'une longue liste de locus significatifs dont la majeure partie seront des faux positifs.

### 4.2 Patrons de sélection détectés

Puisque BayeScEnv ne cherche pas spécifiquement une relation linéaire entre les fréquences alléliques et l'environnement, mais une relation entre la différenciation génétique et environnementale, cette méthode est capable de détecter une grande variété de signaux.

Considérons la variable environnementale  $X$  prenant la valeur  $X_j$  dans la population  $j$ . À partir de cette variable, il est possible de calculer une différenciation environnementale  $E_j = |X_j - R|$  où  $R$  est la référence environnementale (mettons l'environnement moyen sur la méta-population). Notons donc que la différenciation environnementale  $E_j$  est forte quand  $X_j$  est fort *et/ou* faible dans les graphiques ci-dessous.

On peut alors distinguer différents patrons de fréquences alléliques  $f_j$  en fonction de  $X_j$ , dont les trois ci-dessous :



#### **Le cline**

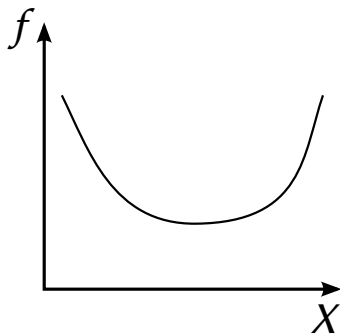
C'est le scénario classique, qui a été simulé dans les ARTICLES 1 et 2 et qui est le plus étudié. C'est aussi le scénario testé par toutes les méthodes d'association environnementale.



### Le scénario “plastique”

Dans ce scénario, les populations se trouvant dans des environnements extrêmes ont des fréquences alléliques similaires. Ce scénario peut survenir lorsque la variable environnementale utilisée est un *Ersatz* de la vraie variable sélective et possède une relation non monotone avec elle. Par exemple, les milieux très chauds ou très froids sont aussi souvent très arides.

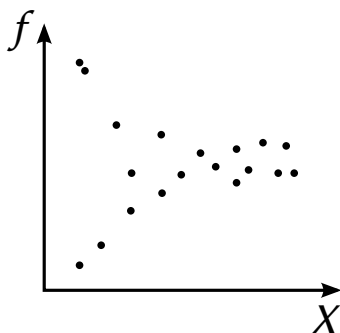
Plus intéressant, ce type de patron est attendu lorsque les gènes sous sélection sont des gènes régulateurs, de réponse au stress, ou plus généralement, impliqués dans la plasticité phénotypique (M. R. J. MORRIS et al., 2014).



### Les fréquences extrêmes

Dans ce scénario, les populations se trouvant dans un environnement différencié possèdent des fréquences alléliques particulièrement différenciées.

Il est très probable que ce type de résultats soit des faux positifs : pourtant, il faut noter que ce type de patron peut émerger dans des cas très particuliers. C'est le cas par exemple d'une sélection fréquence-dépendante déclenchée par une variable environnementale.



Bien sûr, il est possible de trier ces différents scénarios en utilisant des analyses *post-hoc* sur les fréquences alléliques. Pourtant, dans l'absolu, si l'un de ces patrons ressort comme étant significatif, c'est qu'il se passe vraiment quelque chose d'étrange sur ce locus, et que cette étrangeté peut être reliée d'une manière ou d'une autre à l'environnement. Il faut donc être prudent de ne pas ignorer des signaux “bizarres” mais intéressants sous prétexte qu'ils seraient des faux positifs. Par définition, rien ne ressemble plus à un vrai positif qu'un faux positif.

## 5 Pourquoi cette profusion de méthodes ?

Étant donné le grand nombre de méthodes de scan génomique (passées en revue par RELLSTAB et al., 2015), on peut se demander s'il est bien raisonnable d'en rajouter une nouvelle et si l'énergie dépensée dans une course à la méthode idéale ne serait pas mieux allouée au développement d'une méthode qui serait utilisée par tout le monde.

Il existe cependant un problème à cette idée qu'une seule méthode pourrait être utilisée par tous les biologistes évolutifs pour la détection de la sélection : la diversité, non seulement celle des signaux de sélection à détecter, mais aussi celle des structures de population. En ce qui concerne les signaux de sélection, les méthodes d'association environnementale sont conçues pour détecter les signaux de type “cline de sélection”, d'autres méthodes, comme Bayescan (FOLL & GAGGIOTTI, 2008) ou PCA-dapt (DUFORET-FREBOURG et al., 2014) cherchent des signaux de différenciation entre populations

différents du reste du génome (sans environnement), d'autres méthodes enfin (revues et comparées par VATSIUO et al., 2015) sont dédiées à la recherche de balayage sélectif récent.

BayeScEnv quant à elle, tente de détecter tout signe d'adaptation locale qu'elle peut ramener à la variable environnementale testée, quelle que soit la forme que peut prendre la sélection (ce qui ne veut pas dire qu'elle soit absolument générale).

La question de la diversité des structures de population est tout aussi essentielle. Les simulations de l'ARTICLE 1 ont bien montré que les méthodes étaient différemment sensibles à différentes structures de population. Les méthodes prennent ainsi en compte différentes problématiques lors de leurs procédures de test. BayEnv et LFMM par exemple prennent en compte le flux de gène préférentiel entre populations. BayeScEnv ne peut pas prendre en compte explicitement<sup>1</sup> ce type de problèmes, mais est moins sensible aux problèmes d'hétérogénéité du génome créés par des taux de mutations hétérogènes ou de la sélection d'arrière-plan.

Cette diversité des signaux, des structures de population et autres effets de confusion nécessite donc une diversité de méthodes. Qui plus est, certaines méthodes sont développées pour affiner ou améliorer le fonctionnement d'autres méthodes existantes, comme Bayescan pour FDIST ou plus récemment, la méthode BayePass qui résout quelques problèmes dans l'implémentation de BayEnv (GAUTIER, 2015). Il existe maintenant trois méthodes différentes basées sur Bayescan : la méthode originale (Bayescan2), la méthode hiérarchique (Bayescan3, FOLL et al., 2014) et BayeScEnv. Ces trois méthodes possèdent des propriétés différentes qui leur donnent des utilités différentes. De par sa généralité (aucune variable environnementale par exemple) et incidemment son fort de taux de faux positifs, la méthode originelle est une très bonne méthode pour exclure les locus "aberrants" d'une analyse (de génétique des populations par exemple) sur laquelle ils pourraient avoir un fort effet. La méthode hiérarchique permet, outre de prendre des structures hiérarchiques en compte, l'étude de l'évolution parallèle entre différentes régions où peut vivre une espèce. Enfin, BayeScEnv permet d'isoler des signaux plus typiques d'une adaptation locale dus à un facteur environnemental, tout en prenant en compte d'autres sources potentielles de différenciation entre populations.

Finalement, il est nécessaire de rappeler que l'intérêt de ces méthodes ne réside pas dans leur capacité à détecter des locus sous sélection avec certitude : il est probable que le problème des faux positifs ne soit jamais réellement tout à fait résolu. Ces méthodes ne peuvent donc pas servir à considérer des "proportions de génome sous sélection", mais simplement à isoler des locus qui présentent un comportement suffisamment intéressant pour être considérés comme des locus "candidats" à la sélection. Ces études nécessitent donc d'être utilisées en complément d'autres analyses expérimentales.

---

<sup>1</sup>Bien que l'effet locus-spécifique doit contribuer à palier la sensibilité de cette méthode à ces problèmes.

# CHAPITRE 3

## L'expérience de jardin commun à l'ère génomique

---

### 1 Jardins communs et génotypage haut-débit

Comme expliqué dans l'Introduction (section 2.1), l'expérience de jardin commun est l'un des outils expérimentaux les plus utiles pour l'étude de l'adaptation locale. Pour rappel, elle permet de dépasser les problèmes posés par l'influence de l'environnement sur les phénotypes (plasticité phénotypique) en plaçant tous les individus étudiés dans le même environnement.

De nos jours, les expériences de jardin commun en écologie évolutive sont régulièrement utilisées pour effectuer des comparaisons de  $Q_{ST}$ - $F_{ST}$ , bien que ces comparaisons aient été sévèrement critiquées pour les problèmes qu'elles posent en termes de comparabilité des deux quantités (EDELAAR et al., 2011 ; EDELAAR & BJÖRKLUND, 2011) ou de la prise en compte de l'incertitude sur ces deux estimateurs (OVASKAINEN et al., 2011). Souvent, les marqueurs génétiques utilisés pour mesurer la  $F_{ST}$  dans ce type de d'expériences sont des marqueurs AFLP ou micro-satellites. Le génotypage haut-débit peut pourtant notablement dynamiser l'expérience de jardin commun en y apportant une dimension génomique ouvrant la porte à de nouvelles analyses (SAVOLAINEN et al., 2013). Combiner l'expérience de jardin commun avec les approches de scan génomique peut s'avérer aussi extrêmement intéressant (DE KORT et al., 2014).

L'utilisation du génotypage haut-débit et des avancées de la génomique des populations dans le cadre d'une expérience de jardin commun ouvre donc de nouveaux horizons expérimentaux, méthodologiques et scientifiques.

### 2 Présentation de l'ARTICLE 3

L'article suivant passe en revue les différentes opportunités qu'offrent les expériences de jardin commun à l'ère de la génomique moderne. Il ouvre quelques perspectives sur le type d'analyses qui peut être mis en place, spécifiquement lorsque l'on dispose de données génomiques denses (notamment scan génomiques).

#### 2.1 Contributions

J'ai écrit cet article avec l'aide d'Oscar Gaggiotti et d'Irène Till-Bottraud. Médéric Mouterde a participé à la rédaction de la partie sur les méthodes de génotypage haut-débit.

## 2.2 Résumé

De nombreux phénomènes sont facteurs de confusion (comme la dérive génétique ou l'histoire démographique) dans l'étude de l'adaptation locale. Quand des traits complexes sont impliqués dans l'adaptation locale, des phénomènes comme la plasticité phénotypique posent d'autant plus problème pour les biologistes évolutifs cherchant à étudier les relations complexes entre phénotype, génotype et environnement. Dans cette Perspective, nous suggérons que l'expérience de jardin commun, spécifiquement conçue pour s'affranchir des effets de la plasticité phénotypique, doit jouer un rôle dans l'étude de l'adaptation locale, même (voire surtout) à l'ère génomique.

Après une rapide revue des protocoles de génotypage haut-débit pertinents dans le contexte d'un jardin commun, nous explorons comment améliorer les analyses de jardin commun en utilisant des marqueurs génétiques denses et des méthodes statistiques récentes. Nous montrons ensuite comment, en combinant des approches de génomique des populations et d'étude par association pangénomique, il est possible d'aboutir à une étude efficace, exhaustive et intégrative de l'adaptation locale.

Spécifiquement, des résultats issus de la génomique (scans génomiques) et des phénotypes apportent des éléments de preuve indépendants concernant la possibilité d'un scénario d'adaptation locale, et les études d'association pangénomique permettent d'aller étudier les bases génétiques des traits adaptatifs.

Journal: *Heredity*, Early View (2015)

## ARTICLE 3

### Common garden experiments in the genomic era: new perspectives and opportunities.

Pierre de Villemereuil, Oscar E. Gaggiotti, Médéric Mouterde & Irène Till-Bottraud

#### Abstract

The study of local adaptation is rendered difficult by many evolutionary confounding phenomena (e.g. genetic drift and demographic history). When complex traits are involved in local adaptation, phenomena such as phenotypic plasticity further hamper evolutionary biologists to study the complex relationships between phenotype, genotype and environment. In this perspective paper, we suggest that the common garden experiment, specifically designed to deal with phenotypic plasticity has a clear role to play in the study of local adaptation, even (if not specifically) in the genomic era. After a quick review of some high-throughput genotyping protocols relevant in the context of a common garden, we explore how to improve common garden analyses with dense marker panel data and recent statistical methods. We then show how combining approaches from population genomics and genome-wide association studies with the settings of a common garden can yield to a very efficient, thorough and integrative study of local adaptation. Especially, evidence from genomic (e.g. genome scan) and phenotypic origins constitute independent insights into the possibility of local adaptation scenarios, and genome-wide association studies in the context of a common garden experiment allow to decipher the genetic bases of adaptive traits.

#### Introduction

Studying adaptation and the genetic bases of the adaptive traits is an ambitious but daunting enterprise, especially for complex traits that have a polygenic basis and are strongly influenced by the environment. Indeed, uncovering the evidence of genetic adaptation is almost always hampered by the pervasive effects of evolutionary phenomena such as genetic drift, phenotypic plasticity, complex demographic history and complex genetic architecture. In the particular case of local adaptation, evolutionary biologists have developed efficient tools to overcome these challenges and the common garden experiment is one of them. The rationale behind this protocol is to control for the effects of phenotypic plasticity and, to a certain extent, genotype-by-environment interactions by growing individuals from different populations in a common environment, and by using the quantitative genetics toolbox (see Box A3.1) to study the genetic bases of complex traits (e.g. life history, morphological and physiological traits).

Because it enables to unravel the genetic basis of complex phenotypes across various populations without the confounding effects of the corresponding environment, the common garden experiment is used to test for local adaptation signal in traits of interest such as life history traits (KAWAKAMI et al., 2011), phenology (BRACHI et al., 2013) and allometric relationships (GONDA et al., 2011). Local adaptation might be suspected because of the existence of an environmental gradient such a latitude (TORÄNG et al., 2015) or altitude (ALBERTO et al., 2011), or because of the existence of several contrasting environments, such as sea and fresh water (DEFAVERI & MERILÄ, 2014). Additionally, common garden experiments are also used to study the consequences of local adaptation for conservation (MCKAY et al., 2001) or even for ecosystem functioning (BASSAR et al., 2010). Despite its name, and although it has been used extensively with plants (LINHART & GRANT, 1996), this experimental approach can also be applied to a large variety of organisms including fish (BASSAR et al., 2010 ; DEFAVERI & MERILÄ, 2014), invertebrates (SPITZE, 1993 ; LUTTIKHUIZEN et al., 2003) and small mammals (BOZINOVIC et al., 2009). The main limitations to this experimental design are the ability to breed the species and to grow the produced offspring in laboratory or semi-natural conditions. Common garden experiments can also be used to study genotype-by-environment (G×E) interactions, by

implementing the same design in different environments. Although replicating common garden experiments is logistically challenging, the outcomes of such experiments are highly rewarding, as  $G \times E$  effects are likely common and very important in the wild (STINCHCOMBE, 2014).

Note finally that, although common garden experiments are closely related to reciprocal transplant experiments (which aim at testing local adaptation by showing that the average fitness of local individuals is higher than the average fitness of aliens, see e.g. ÅGREN & SCHEMSKE, 2012), there are important philosophical and practical differences between the two types of experiments. The difference is that reciprocal transplants are designed to prove local adaptation, whereas common gardens are designed to study the genetic bases of traits, regardless of whether they are adaptive or not. In practice, reciprocal transplants will typically create a differential survival, because the locals will survive better. This will be a confounding effect during the quantitative genetic analysis, because only the phenotypes of “fit” individuals are available. Common gardens, by contrast, are often designed to be “softer” on the individuals. Nevertheless, most of the elements in this article regarding common garden experiments can also be applied to reciprocal transplants, especially if one is interested in applying them to survival or some other measure of fitness.

To perform the quantitative genetics analyses of the studied traits, individuals of controlled families (i.e. group of individuals with known genealogy) are used. An average relatedness between individuals is derived from this known genealogy and allows to infer within-population additive genetic variance  $V_A$ , whereas effects due to the population of origin allows to infer the between-population additive genetic variance  $V_{\text{pop}}$ . This is so because all individuals share the same environment and, therefore, any average difference between populations must have a genetic origin. The residual variance  $V_R$  accounts for all other kinds of effects (e.g. environmental). These variance components can be used to estimate the heritability of the trait:

$$h^2 = \frac{V_A}{V_A + V_R} \quad (\text{A3.1})$$

It is also possible to estimate  $Q_{\text{ST}}$ , a standardized measure of genetic differentiation for quantitative traits (SPITZE, 1993; EDELAAR et al., 2011).  $Q_{\text{ST}}$  is defined as the ratio of among-population (additive) genetic variance  $V_{\text{pop}}$  over the total genetic variance (i.e. including the within-population additive

### Box A3.1: Quantitative genetics glossary

**Quantitative Genetics** Theoretical framework used to study the genetic basis of (mostly) quantitative polygenic traits. It uses relatedness between individuals to partition the phenotypic variance into (among others) genetic and non genetic components.

**Relatedness** Probability of shared ancestry (identity by descent, IBD) of any two homologous alleles sampled among two individuals. Can also be defined in terms of correlation of homologous alleles between two individuals when the reference population is the sample itself. Relatedness is indeed always defined according to a reference population (J. WANG, 2014).

**Additive genetic variance ( $V_A$ )** Variance component due to the additive effects of the alleles and genes responsible for the phenotype. Under general conditions (no epistasis, no inbreeding), this is the only component transmitted to the offspring generation.

**Dominance variance ( $V_D$ )** Genetic variance arising from interactions between alleles within each gene responsible for the phenotype. The dominance effect is perceptible only when comparing full-sibs and in presence of mild to strong inbreeding (WOLAK & KELLER, 2014).

**Parental effects** Direct or indirect effects of the parental phenotype on the offspring phenotype, apart from the genetic heredity of the phenotype. This includes, in particular, maternal energetic investment in offspring.

**Heritability** Proportion of the phenotypic variance genetically transmissible to the offspring generation within a population. Calculated as a ratio between  $V_A$  and the total phenotypic variance. The marker-based heritability is the proportion of phenotypic variance explained by the whole genetic marker panel, which is not necessarily equal to the true heritability.

**$Q_{\text{ST}}$**  Among-population genetic differentiation index. Ratio of the among-population additive genetic variance  $V_{\text{pop}}$  to the total additive genetic variance (calculated as  $V_{\text{pop}} + 2V_A$ ).

variance  $V_A$ ), and in the case of diploid species is given by:

$$Q_{ST} = \frac{V_{pop}}{V_{pop} + 2V_A} \quad (A3.2)$$

This parameter is a quantitative analogue of population genetics'  $F_{ST}$  and, under a hypothesis of neutrality, both should be equal. Hence, a common approach for distinguishing between neutral drift and local adaptation scenarios is to compare  $Q_{ST}$ 's and  $F_{ST}$ 's. Consequently, individuals from a common garden experiment are typically genotyped to compute  $F_{ST}$ .

Despite the advantages of common garden experiments, the study of local adaptation in non-model species during the last decade has been strongly driven by the study of genetic markers in natural populations (LUIKART et al., 2003). Typically, evolutionary biologists go to natural populations, sample tissue from the individuals and genotype them with high-throughput methods and then proceed with a genome scan analysis of selection (e.g. ECKERT et al., 2010 ; BOURRET et al., 2013 ; FISCHER et al., 2013). Although this method can be quite powerful, it has some limitations (e.g. false positives, no information on the adaptive phenotype). Several calls have been made to independently validate the results of such analyses (see BUEHLER et al., 2014, for a striking example), possibly using common garden or reciprocal transplant experiments (HOLDEREGGER et al., 2008 ; PARDO-DIAZ et al., 2015 ; RELLSTAB et al., 2015). Following these lines, this perspective paper addresses three main questions: where does the common garden experiment stand in the genomic era? In particular, what can common garden experiments bring to population genomics? Conversely, how can techniques from the genomic fields (e.g. high throughput genotyping and model-based inference of neutral evolution) extend the range and scope of common gardens?

It is important to note that population genomics aims at linking genotypes and environments through genome scans methods but often completely neglects to study the phenotypic traits under potential selection. There is much to gain by adding phenotypes into the equation (CUSHMAN, 2014). Yet, because phenotypic plasticity is hard to distinguish from local adaptation in wild populations, it seems useless, or at least dubious, to use phenotypes directly obtained in the field. This simple fact lies at the heart of common garden experiments and we suggest here that this approach is ideally suited to jointly study genotypes, phenotypes and environments, especially when they are combined with high-throughput genotyping and powerful statistical methods. After a short introduction to the different high-throughput genotyping methods available in the context of a common garden experiment, we will discuss how those methods and powerful statistical tools can rejuvenate this classical approach. Finally, we will discuss the complementarity between population genomics and common garden experiments, and how an integrative analysis can deepen our understanding of local adaptation.

## High-throughput genotyping in the context of a common garden

High-throughput genotyping defines any genotyping method yielding a large number of markers, thus providing a dense marker panel across the genome. Given the focus on non-model species in this paper, we consider as few as 10,000 independent markers as fairly “dense”, provided that the genome of the species is not too large. For example, 10,000 SNPs in a genome of size 100Mbp would represent approximately 3% of all SNPs if a SNP occurs every 300bp.

The most straightforward high-throughput genotyping method is whole genome sequencing (WGS). This method yields the largest possible number of markers, and offers the densest genotyping. However, this technique requires high DNA quality and quantity, bioinformatics computation power and, most importantly, access to genomic resources (e.g. genome assembly) within a relatively short phylogenetic range. The huge number of markers generated can also be problematic during the analyses because of high computation/memory requirements, high redundancy in information between linked markers and low signal to noise ratio. Still WGS is the ultimate high-throughput genotyping method, yielding up to millions of SNP markers throughout the whole genome. With a decreasing

cost and an increase in the number of species for which the whole genome has been sequenced over the years, it might soon become a recommended technique even for non-model species. A cheaper alternative to WGS are SNP genotyping chips, with most of the limitations above applying still.

For now, an approach likely to be best-suited for non-model species is genome representation sequencing. The overall principle of this approach is to sequence only restricted, but random, parts of the genome in order to decrease the sequencing effort, and hence the overall costs and computational efforts associated with genotyping. To do so, the above approaches mainly use DNA digestion by restriction enzymes followed by a ligation of tags and primers and PCR amplification. This is akin to the principle underlying AFLP genotyping (Vos et al., 1995). Here, however, the DNA fragments (or at least some of them) are partially sequenced (approx. 100 bp) using next generation technology such as Illumina HiSeq. This kind of approach includes the genotyping-by-sequencing method (GBS, ELSHIRE et al., 2011) and the family of Restriction-site Associated DNA sequencing methods (RADseq, MILLER et al., 2007 ; BAIRD et al., 2008).

The sequences obtained are then analysed using quality checks (i.e. selecting reads according to their sequencing quality, local coverage, availability over all or most individuals, etc.) and SNP calling pipelines, in order to identify SNP markers. Note that contrary to the AFLP approach, markers issued from RAD sequencing are preferentially issued from non-polymorphic restriction sites and are co-dominant. Alternatively, when more than one SNP is present on a 100bp sequence, they can be combined into a new marker with more than two alleles. The rationale behind this is that very close SNPs are likely to be strongly associated due to physical linkage, in which case, fewer but independent markers composed of more alleles are often preferable to strongly linked SNPs. Genome representation protocols can yield up to several hundreds of thousands of SNPs, but more typically tens of thousands. This can be achieved at a cost comparable or up to 10 times the cost of an AFLP analysis.

For all of the above, it is clear that NGS makes possible the generation of a very large number of markers for a moderate cost. When compared to AFLP markers, NGS marker panels are denser, the markers are codominant, and less arbitrary in their interpretation (i.e. no “binning” process), hence better in every way, except possibly for their cost. Micro-satellites, on the other hand, are very different: they usually provide very sparse panels (up to a few dozens of markers), but highly mutable and with a large allelic diversity. Although it has been argued that micro-satellites are better markers to infer relatedness (RITLAND, 2000), they typically yield smaller relatedness estimates than SNP or AFLP markers due to higher mutation rates (UPTMOOR et al., 2003 ; EL RABEY et al., 2013). They also yield smaller  $F_{ST}$  estimates (EDELAAR & BJÖRKLUND, 2011) for the same reason. Finally, although in theory more accurate than SNPs for the same number of loci, they typically yield one to two orders of magnitude less loci, hence they are less accurate in practice (UPTMOOR et al., 2003).

A key issue is the number of individuals that need to be genotyped. Our view is that ideally all individuals from the experimental garden(s) should be genotyped, because this opens the way towards the more refined or novel analyses detailed below. However, some of the analyses suggested here (e.g. genome scans) can be performed even when a sub-sample of individuals have been genotyped. DE KORT et al. (2014), for example, have sampled one individual *per* family in their common garden experiment to combine it with population genomics (i.e. genome scans) analyses. This cheaper sub-sampling procedure might be very attractive to researchers who are not interested in individual genotypes: i.e. neither in the relatedness inference, nor in the genome-wide association studies that are described below.

## Common gardens 2.0: new markers and new methods

We are certainly not the first to encourage the evolutionary biology community to switch towards NGS technology (LUIKART et al., 2003 ; SAVOLAINEN et al., 2013), and it is clear that such a “revolution” is already happening (reviewed in PARDO-DIAZ et al., 2015). However, we wish here to emphasise the interest of dense marker panels in the context of a common garden experiment.



As stated above, a study of the genetics of complex traits such as measured in common garden experiments strongly relies on the relatedness between individuals, which is often assumed, especially when individuals are siblings (e.g. HERNÁNDEZ-SERRANO et al., 2014). Yet contrary to the parent-offspring relationship, the relatedness between siblings varies: the commonly used value of 0.25 between half-sibs for example, is only an average, expected value. Hence using realised relatedness, inferred from molecular data, can allow for better estimates in the sense that (i) they are more robust to error in the kinship assessment (e.g. full-sibs instead of half-sibs) and (ii) they reflect more accurately the variation in relatedness between siblings. Better relatedness estimates are useful because they will improve the precision of the estimates of  $h^2$  and  $Q_{ST}$ . Note however that many markers are typically needed to obtain precise molecular estimates of relatedness (UPTMOOR et al., 2003). Dense markers provided by high-throughput genotyping naturally fulfill this requirement.

A large number of markers also allows the reconstruction of the family structure. Indeed even when relatedness is precisely estimated, the family structure (i.e. who is the mother/father of the individuals, which individuals are full- or half-sibs) is of utmost importance, in order to account for many confounding effects such as dominance (WOLAK & KELLER, 2014), parental effects (e.g. maternal WILSON et al., 2010) or selfing (GAUZERE et al., 2013). Note that maternal effects can also be accounted for by weighting seeds (in plants, ROACH & WULFF, 1987) or reduced by using F2 generations (ROACH & WULFF, 1987; MOUSSEAU & DINGLE, 1991). However, the possibility of using one of these methods will strongly depend on the studied species. According to A. G. JONES et al. (2010), brood size is one of the biggest limitations for parental reconstruction algorithms, due to issues of unsampled alleles when too few segregating individuals are available. With many markers, even with low levels of polymorphism (such as SNPs), this is no longer an issue, as it becomes possible to reconstruct a large-enough proportion of the parental genomes to obtain high certainties of assignment, even for small brood sizes. Now that efficient algorithms such as those implemented in COLONY (O. R. JONES & J. WANG, 2010; J. WANG, 2012), are available, the number of markers should not be a problem. This software allows reconstructing the family structure, as well as inferring parental genotypes, while accounting for selfing or genotyping errors. Indeed, one crucial issue for parental inference with a large number of markers is to include possible genotyping errors, which, if left unaccounted for, can severely bias the results (J. WANG, 2004).

The most innovative statistical method, especially designed to study common garden data, is probably the one developed by OVASKAINEN et al. (2011), which overcomes several problems associated with the classical  $F_{ST}$ – $Q_{ST}$  comparisons. In order to avoid clumsy comparisons between two noisy estimators, OVASKAINEN et al. (2011) conceived a model of neutral phenotypic differentiation between populations that is compared to phenotypic differentiation measured in a common garden experiment (i.e. the genetic differentiation linked to the phenotype). When suspiciously strong phenotypic differentiation is observed compared to the neutral expectation, a local adaptation hypothesis can be proposed. The neutral model of phenotypic differentiation is actually a combination of a within-population “animal model” (see KRUIK, 2004, for a description of the model) and an among-populations “ $F$ -model” see GAGGIOTTI & FOLL, 2010, for a description of the model of phenotypic evolution (KARHUNEN & OVASKAINEN, 2012). By doing so, this model allows for a multivariate genetic analysis to be performed, e.g. to infer genetic correlations and a  $G$  matrix. This is a perfect illustration of how models emerging from the field of population genomics (here the  $F$  model) can be used to dramatically improve the analysis of common garden datasets. This method has been implemented in the DRIFTSEL package (KARHUNEN et al., 2013). Using this method, KARHUNEN et al. (2014) demonstrated the presence of strong footprints of local adaptation in several populations of nine-spine stickleback (*Pungitius pungitius*).

## What is the use of common garden experiment in the genomic era?

It is well known in the domain of genome wide association studies (GWAS), which aim at uncovering the loci responsible for phenotypic variation, that such analyses should be performed with

extreme caution because of the potential effect of hidden population structure. Especially important are the combined effects of genetic drift and gene flow, and the confounding effect of phenotypic plasticity. However, both of the afore-mentioned problems can be overcome. Structure between populations structure can be accounted for by using appropriate models (e.g. NICHOLSON et al., 2002; BEAUMONT & BALDING, 2004) or methods (FRICHOT et al., 2013) from the genome scan literature. The second problem, on the other hand, is perfectly addressed by common garden experiments, which were specifically designed to control for phenotypic plasticity!

As a result, combining common garden experiments of non-model species with GWAS provides opportunity for multiple-population genome-wide association studies (BRACHI et al., 2013; SLAVOV et al., 2014). For a locally adapted trait, it would even be possible to differentiate markers explaining among-population phenotypic variability (by testing for among-population effects) from markers explaining within-population variability (by testing for within-population effects). The technique of *within-group centring* (DAVIS et al., 1961; van de POL & J. WRIGHT, 2009) could be used to this end. It simply consists in distinguishing between the mean-population effect and the within-population effect of each predictor of an association model, as follows:

$$y_{ij} \sim \mu + \beta_B \bar{x}_j + \beta_W (x_{ij} - \bar{x}_j) + u_j + e_{ij}, \quad (\text{A3.3})$$

where  $y_{ij}$  is the phenotype of individual  $i$  in population  $j$ ,  $x_{ij}$  is its genotype and  $\bar{x}_j$  the mean genotype in population  $j$ . The parameters  $\mu$ ,  $\beta_B$  and  $\beta_W$  are the fixed effects of the model. Note that the within-population effects can be tested independently by using a parameter  $\beta_W^j$  for each population  $j$ . The term  $u_j$  stands for any population structure correction and  $e_{ij}$  is the residual. This equation is simply an illustration of within-group centring and does not constitute a model *per se*. Accounting for population structure should help in distinguishing between neutral and selective scenarios for markers associated with between-population variability. As always (KORTE & FARLOW, 2013), the power of a GWAS to actually detect loci linked to the phenotypic variability strongly depends on the extent of linkage disequilibrium and the density of markers along the genome, in addition to the sample size. Hence the most useful, but most expensive, genotyping method for this kind of analysis is WGS. Note also that heterogeneity in recombination/mutation rates along the genome can generate false positives during such analyses (KORTE & FARLOW, 2013). Here the number of populations is also of importance, as it will determine the power to detect significance for the parameter  $\beta_B$ . Note that BRACHI et al. (2013) used a different approach of multi-scale (local to worldwide variation) analysis and found very different results depending on the studied scale of local adaptation. The approach that is probably the most typical of the genomic era is to scan genomes for signal of selection (mostly selective sweeps and local adaptation). Many methods have been developed in the last decades to detect local adaptation (BEAUMONT & BALDING, 2004; FOLL & GAGGIOTTI, 2008; COOP et al., 2010; BONHOMME et al., 2010; FRICHOT et al., 2013; GUILLOT et al., 2014; DUFORET-FREBOURG et al., 2014). Despite considerable efforts to account for population structures, these methods have been shown to display high error rates (LOTTERHOS & WHITLOCK, 2014; de VILLEMEREUIL et al., 2014). Hence validation of the results of a genome scan must always be done using independent tests. Gene ontologies and pathway analyses are the most common mean of checking these results. However, it has been suggested that common garden experiments might be a very efficient complement to those analyses (LEPAIS & BACLES, 2014; DE KORT et al., 2014; RELLSTAB et al., 2015).

Performing genome-scan analyses using common garden data can have many advantages. If a strong adaptive signal is detected both using both using genome scan methods (i.e. using genotypic and possibly environmental data) and the phenotypic data from a common garden experiment, that will constitute two independent piece of evidence favouring the hypothesis of local adaptation (HOLDEREGGER et al., 2008). As stated above, genome scan results need to be validated anyhow (PARDO-DIAZ et al., 2015; RELLSTAB et al., 2015), and performing a common garden experiment is an elegant way to do so. Whenever possible, combining genome scan approaches with common garden experiments is an efficient approach to the study of local adaptation. Moreover, by comparing the loci showing strong signals of differentiation and the loci associated with among-population phenotypic differentiation, it is possible to isolate candidate loci for local adaptation with very little information regarding the functional annotation of the species' genome. Third, using the environ-

mental information allows not only to identify the selected phenotypes (i.e. strongly differentiated genetically), but also to infer the environmental variable driving the selective pressure. In particular, if a locus is strongly associated with an environmental variable and with the among-population phenotypic differentiation, one might conclude that a relationship exists between the environmental variable and the phenotype (although only correlatively: each variable is a putative proxy for the real selective/selected variable).

An important problem, when performing genome-scan analyses directly on common garden individuals, is to correctly infer the source-population allele frequencies. The preferred way is simply to genotype the parents of the common garden individuals. But this is not always possible (e.g. genotyping the father for plants is impossible most of the time). In that case, allele frequencies inferred directly from the individuals should be accurate, as long as there is no sex-dependent allelic frequency bias. But the confidence in that inference will be overestimated by the fact that many related individuals were sampled. To account for this situation, a conservative solution is to calculate the allele frequencies based on the individuals of the common garden, but to consider that the sample size of these estimates are the number of parents that have generated the offspring. With this kind of data, all population-based methods (such as Bayescan, FOLL & GAGGIOTTI, 2008 or BayEnv, COOP et al., 2010) can be used. A second solution, if the confidence in parental genotypic reconstruction is high enough, is to directly use the inferred genotypes of the parents both to infer allele frequencies in the population, and directly as data for individual-based genome scan methods. Yet, in practice, these data will always be inferred with some uncertainty, and the consequences of ignoring this uncertainty during *post-hoc* analyses is unknown. Still, the interest of this approach is that individual-based methods (such as LFMM, FRICHOT et al., 2013 or PCAdapt, DUFORET-FREBOURG et al., 2014) can be used to analyse the data. A last solution is the one implemented by DE KORT et al. (2014), which consists in using only one individual *per* family. Although this solution requires a sufficiently large number of families for each population, it has the compelling advantage of simplicity and efficiency.

## Conclusion

Local adaptation is a play starring three actors: the environment, the phenotype and the genotype. The environment selects the phenotypes, which are (partly) determined by a number of genes. The evolutionary result is a change in allele frequencies of the polymorphic coding genes. Understanding the relationships between the three actors requires precise but large-scale measurements, rigorous experiments and powerful statistical methods. Because phenotypic plasticity is such a pervasive phenomenon and because it is nearly impossible to account for its effect on *in situ* phenotypes, phenotypes should never be directly compared between different populations, unless a case is made that the comparison is safe enough (low environmental contrasts or little phenotypic plasticity). In contrast, common garden experiments are ideally suited to perform such kinds of analyses, hence to study the phenotypic traits impacted by local adaptation.

Now that dense marker panels are obtainable for many individuals at a moderate cost, common garden experiments are expected to be performed more routinely. Of course, this is unless the biological characteristics (e.g. size, behaviour, generation time) prevent the applicability of this experiment. Common gardens could possibly even replace the field work required to obtain tissue samples for genotyping: as we mentioned, it would still allow for population genomics approaches, while guaranteeing independent validation through the study of phenotypes (PARDO-DIAZ et al., 2015; RELLSTAB et al., 2015), hence saving the cost of another genotyping campaign. As emphasised by LEPAIS & BACLES (2014), deciphering the genetic basis of local adaptation will only be accomplished by combining all the information yielded by dense marker panels, careful experiments and *in situ* sampling and observations. Replicating common garden experiments in different environments can also provide insight into complicated relationships between the three actors such as genotype-by-environment interactions.

High-throughput genotyping provides an abundance of genetic data. World-wide fine-scale databases (e.g. WorldClim, HIJMANS et al., 2005) and the advent of cheap *in situ* sensors also provide high quality environmental data. But collecting phenotypic data is still time-consuming, tedious and sometimes expensive. It thus seems that the last challenge that needs to be overcome is the development of high-throughput phenotyping allowing for a scaling-up and a more widespread use of common garden experiments.

## Aknowledgement

We thank the associate editor and three anonymous referees for their very thorough and relevant reviews, which considerably improved the focus and quality of this manuscript. PdV was supported by a doctoral studentship from the French *Ministère de la Recherche et de l'Enseignement Supérieur*. OEG was supported by the Marine Alliance for Science and Technology for Scotland (MASTS).

## 3 Message principal de l'article

### 3.1 L'apport du génotypage haut-débit

On pourrait naïvement croire que l'apport du génotypage haut-débit aux expériences de jardin commun pourrait se résumer à une simple augmentation du nombre de marqueurs disponibles, donc des estimations plus précises. En quelque sorte, il s'agirait d'un apport quantitatif qui ne changerait pas fondamentalement les possibilités d'analyse et ne justifierait donc que difficilement l'argent et l'énergie engagés dans ces protocoles.

Pourtant, le génotypage haut-débit ouvre effectivement de nouvelles perspectives d'analyse, et notamment dans le domaine des scan génomiques (que ce soit pour détecter de la sélection ou une association avec un phénotype). Les études d'association pangénomique sont régulièrement effectuées sur des populations dans un milieu contrôlé (élevage, serre) ou dont les facteurs environnementaux induisant de la plasticité phénotypique sont connus (chez l'Humain par exemple).

Effectuer de telles analyses d'association en milieu naturel est un exercice périlleux, dont les résultats risquent d'être très fortement influencés par la plasticité phénotypique. C'est pourquoi nous affirmons dans cet article que l'expérience de jardin commun possède tout son sens à l'ère génomique, notamment dès lors que l'on veut étudier les bases génétiques de traits adaptatifs.

Bien sûr, il serait contre-productif de prétendre que toute analyse de l'adaptation locale doit passer par une expérience de jardin commun, notamment si aucune analyse phénotypique n'est prévue et que seule la relation entre les gènes et l'environnement est étudiée, comme c'est typiquement le cas en génomique du paysage (RELLSTAB et al., 2015).

De même, si la base génétique des traits adaptatifs n'offre aucun intérêt particulier, il se peut que le génotypage haut-débit soit superflu dans une expérience de jardin commun. Tout dépendra, bien sûr, de la question biologique à laquelle l'expérience et l'analyse se proposeront de répondre.

### 3.2 Un nouveau statistique

Dans l'article, nous suggérons aussi que l'utilisation de nouvelles méthodes statistiques peut permettre de raffiner encore l'analyse des données de jardin commun, sans forcément nécessiter un génotypage haut-débit : c'est le cas, par exemple, du modèle d'OVASKAINEN et al. (2011). En effet, bien que les comparaisons  $Q_{ST}-F_{ST}$  puissent être utiles, elles sont entachées d'un grand nombre de difficultés statistiques, que le cadre théorique développé par OVASKAINEN et al. se propose de surmonter.

Dans le CHAPITRE 5, nous utiliserons le cadre théorique de ce modèle pour détecter de l'adaptation locale en utilisant des données environnementales. Cette pratique de détecter une relation entre deux variables en prenant en compte un modèle d'évolution neutre est courante à la fois dans le domaine des scan génomiques pour détecter la sélection (COOP et al., 2010) et dans le domaine des analyses comparatives phylogénétiques (GARAMSZEGI, 2014).

## 4 Un mot sur les traits non Gaussiens

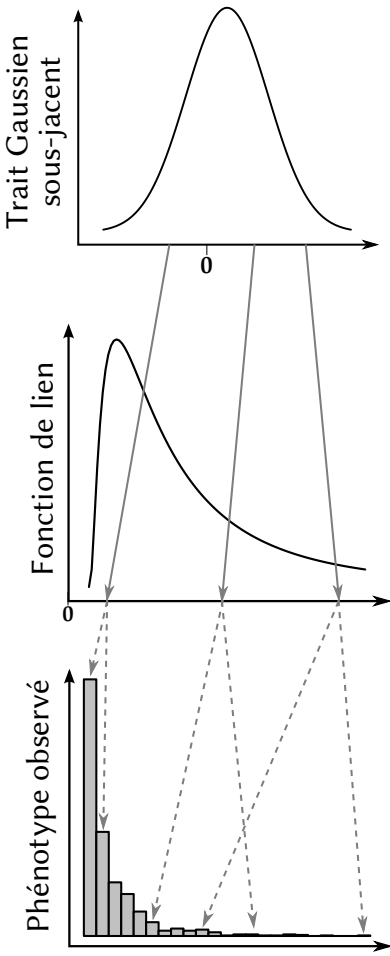


FIG. 8 : Schéma montrant la relation entre le trait Gaussien sous-jacent et le phénotype, en passant par le résultat de la fonction de lien dans un modèle linéaire généralisé.

La majeure partie des modèles d'évolution, comme celui d'OVASKAINEN et al. (2011) est pensée pour l'étude des traits Gaussiens. Cette hypothèse est à la fois faite pour des raisons pratiques, car ces modèles sont mathématiquement plus simples, et théoriques, car la génétique quantitative suppose l'émergence d'un phénotype Gaussien suite à l'effet additif des gènes (FISHER, 1918).

Pourtant, une très grande partie des traits d'intérêt pour l'écologie et l'évolution sont des traits qui ne suivent pas une distribution Gaussienne. C'est le cas notamment pour la survie, qui est un trait binaire, ou de traits liés à la reproduction (comme la taille de ponte), qui souvent suivent plutôt une loi de type Poisson. Puisqu'il est attendu que le résultat des effets additifs des gènes participant à la variabilité d'un caractère soit de distribution Gaussienne, des modèles hiérarchiques sont utilisés pour analyser ce type de traits. Dans ces modèles (nommés modèles linéaires mixtes généralisés), une variable  $l$  (un "trait sous-jacent" hypothétique) possède une distribution Gaussienne, qui est ensuite reliée au phénotype observé  $z$  par une fonction de lien  $g$  et une loi d'erreur appropriée  $\mathcal{D}$  (une loi de Poisson par exemple, voir FIG. 8) :

$$l \sim \mathcal{N}(\mu, \Sigma) \tag{3.1a}$$

$$\eta = g^{-1}(l) \tag{3.1b}$$

$$z \sim \mathcal{D}(\eta, \theta) \tag{3.1c}$$

Le cœur de ce modèle est donc le trait sous-jacent  $l$  plus que le phénotype observé. C'est sur ce premier que sont calculés les effets fixes linéaires et les variances des effets aléatoires (dont la variance génétique additive  $V_A$  et la variance inter-population  $V_{pop}$ ).

Dans un article rédigé pendant ma thèse, nous avons montré avec des collègues (voir Annexe S3) que, dans les modèles linéaires mixtes généralisés, la variance génétique additive du trait sous-jacent et celle du trait phénotypique observée était reliée par un paramètre  $\Psi$  (voir ÉQ. A6.18) :

$$V_A^{obs} = \Psi^2 V_A. \tag{3.2}$$

De manière intéressante, puisque la variance inter-population est aussi d'origine génétique additive, le même raisonnement s'applique naturellement et :

$$V_{pop}^{obs} = \Psi^2 V_{pop}. \tag{3.3}$$

La conclusion est que la  $Q_{ST}$  calculée sur l'échelle du trait phénotypique vaut :

$$Q_{ST}^{obs} = \frac{\Psi^2 V_{pop}}{\Psi^2 V_{pop} + 2\Psi^2 V_A} = \frac{V_{pop}}{V_{pop} + 2V_A} \tag{3.4}$$

Les deux types de  $Q_{ST}$  (sur l'échelle du trait sous-jacent ou du phénotype observé) sont donc égales. Ce résultat est dû à la linéarité des effets additifs. Le fait d'analyser des traits non Gaussiens, bien qu'il influe sur le calcul de l'héritabilité, n'influe donc pas sur le calcul de la  $Q_{ST}$ .

Quoiqu'il en soit, les effets de l'adaptation locale devraient se faire ressentir sur l'échelle du trait sous-jacent (qui reflète l'effet additif des gènes), bien que la sélection affecte le phénotype observé. Ainsi, tant qu'il n'est question que d'étudier la présence et les effets de l'adaptation locale, il est plutôt recommandé de travailler sur l'échelle du trait sous-jacent. Par contre, s'il s'agit de prédire les effets de la sélection sur les générations à venir, alors il est nécessaire de prendre en compte le fait que la sélection agit sur le phénotype observé et que son action doit se mesurer sur ce même phénotype (voir Annexe S3).

# **DEUXIÈME PARTIE**

---

**Adaptation aux conditions d'altitude chez la  
plante alpine *Arabis alpina***



# CHAPITRE 4

## Étude démographique *in situ* d'*Arabis alpina*

---

### 1 Une plante écologiquement particulière ?

#### 1.1 Une gamme altitudinale très large

Comme expliqué dans l'Introduction générale, *Arabis alpina* est une espèce artico-alpine capable de se développer sur une gamme altitudinale particulièrement large (de 800m à 3000m d'altitude). Ainsi, bien qu'étant une espèce majoritairement alpine et sub-alpine, elle peut descendre à des altitudes relativement faibles. Pourtant, redescendre le gradient altitudinal est souvent difficile pour les espèces alpines, qui sont généralement de particulièrement mauvaises compétitrices (CHOLER et al., 2001). *A. alpina* étant une plante pionnière, spécialiste des milieux ouverts en terrain instable (WHITTAKER, 1993), il est possible que son mode de vie lui permette d'éviter une partie de la compétition qu'elle pourrait subir à de plus basses altitudes, bien qu'elle reste en compétition avec d'autres espèces pionnières à ces altitudes (en particulier des annuelles).

De plus, les conditions environnementales sont réellement contrastées entre le milieu de l'étage montagnard (800m) et le haut de l'étage alpin (3000m). Le simple constat de la variation du paysage floral entre ces altitudes (forêts mixtes, puis de conifères, prairies et enfin végétation éparse) témoigne de ces variations. Bien sûr, celles-ci ne sont pas complètement linéaires avec l'altitude, et beaucoup dépendent des conditions locales (topologie, exposition), voire micro-locale (KÖRNER, 2003 ; KÖRNER, 2007). Pourtant, il est indéniable que les conditions biotiques et abiotiques auxquelles doit faire face *A. alpina* sont extrêmement variées et que l'ampleur de ces variations entraînent plutôt des variations en composition spécifiques (de BELLO et al., 2013).

Ce constat appelle donc à se poser la question suivante : comment *A. alpina* arrive-t-elle à coloniser des milieux dont la variabilité environnementale est si forte ?

#### 1.2 Possibilité d'adaptation locale

Pour qu'une espèce puisse coloniser des milieux extrêmement variés, qui peuvent imposer de fortes contraintes sélectives, il existe plusieurs possibilités, dont les deux majeures sont décrites dans l'Introduction générale : la plasticité phénotypique et l'adaptation locale. La plasticité phénotypique est, à priori, le plus efficace de ces deux mécanismes, puisqu'elle permet une réponse quasiment instantanée (sur une échelle de temps micro-évolutive) à l'environnement dans lequel se trouve la population. Pourtant, il s'agit aussi d'un mécanisme qui peut être coûteux (DEWITT et al., 1998). Pire,

puisque l'évolution de la plasticité elle-même dépend du contexte sélectif passé, elle peut s'avérer maladaptative (GHALAMBOR et al., 2007).

L'adaptation locale est un mécanisme beaucoup plus lent, mais qui entraîne une réponse sur le long terme des populations à l'environnement dans lequel elles se trouvent.

La possibilité d'adaptation locale dépend essentiellement de l'équilibre entre le flux de gènes, le potentiel adaptatif des espèces et la force de la sélection naturelle. En ce qui concerne le flux de gènes chez *A. alpina*, il semble être plutôt localisé. Dans une étude de paternité à large échelle, BUEHLER et al. (2012) ont détecté des parentés sur de longues distances (1km), mais la plupart étaient extrêmement localisés sur 10m autour de la plante. Ce type de flux de gènes très localisé mais à longue "queue de distribution" permet assez facilement l'apparition d'adaptation locale.

Un problème, par contre, est posé par la petite taille générale des populations d'*A. alpina*, ce qui peut limiter l'effet de la sélection. Cette petite taille est encore aggravée par l'autofécondation partielle des individus, augmentant encore plus l'effet de la dérive. Il se peut que ce problème soit compensé par deux phénomènes. Tout d'abord, si la dispersion est suffisante, alors l'effet de la dérive est amoindri par le fait que des allèles adaptatifs pour la population locale perdus par effet de dérive puissent être réintroduits par le biais du flux de gènes. Ensuite, et de manière plus mineure, la chute du taux d'hétérozygotes entraînée par l'autofécondation peut permettre d'éliminer plus facilement des allèles délétères récessifs de la population.

Il est possible que ces effets soient partiellement compensés par une banque de graine à long terme dans le sol, mais aucun travaux n'ont encore montré les caractéristiques des banques de graine chez cette espèce.

En bref, *A. alpina* se retrouve dans des milieux très différents, ce qui suggère une possibilité de forte plasticité phénotypique et/ou d'adaptation locale. Ce dernier scénario est rendu possible notamment par les caractéristiques du flux de gènes, mais pourrait être limité par les petites tailles de population. Notons que ces deux scénarios ne sont pas exclusifs l'un de l'autre.

### 1.3 Dynamique en milieu naturel ?

Avant de vraiment pouvoir avancer que la variabilité environnementale est liée à la gamme altitudinale d'*A. alpina*, il est important de tout d'abord vérifier à quel point elle possède un impact sur la dynamique des populations *in situ*. Dans ce chapitre, il s'agit justement d'apporter des informations biologiques sur cette plante encore peu étudiée en conditions naturelles, ainsi que déterminer les liens entre l'environnement et sa dynamique démographique. Tout d'abord, il peut être intéressant de faire le point sur les travaux sur cet organisme effectués au Laboratoire d'Écologie Alpine ces dernières années.

## 2 *Arabis alpina* au Laboratoire d'Écologie Alpine

### 2.1 Études de génétique des populations

L'étude de la plante alpine *A. alpina* a démarré au Laboratoire d'Écologie Alpine lors de la thèse de Bénédicte Poncet, débutée en 2006, qui a donné lieu à de nombreux travaux de génétique du paysage sur cette espèce (MANEL et al., 2010 ; PONCET et al., 2010 ; BUEHLER et al., 2013).

Ces travaux ont montré une forte relation entre l'environnement et les patrons de différenciation génétique chez cette espèce, spécifiquement l'influence de la température et de la précipitation

(MANEL et al., 2010). Une étude de scan génomique en France et en Suisse a pu détecter un certain nombre de locus AFLP sous sélection (9% sur 825 locus), principalement reliés à la température annuelle moyenne. Ces résultats suggèrent une présence probable d'adaptation locale. Ces travaux ont, en outre, montré que si la différenciation génétique entre les populations suisses était moyenne ( $F_{ST} = 0.13$ ), elle était relativement importante dans les populations du nord des Alpes ( $F_{ST} = 0.28$ , BUEHLER et al., 2013).

Enfin, une étude de séquençage massif a été menée pour obtenir un génome de référence pour cette plante (LOBRÉAUX et al., 2014). Ces travaux ont montré une assez forte homologie entre le génome d'*A. alpina* et celui de la plante modèle *Arabidopsis thaliana*.

## 2.2 Études de physiologie végétale

Des travaux concernant l'adaptation physiologique d'*A. alpina* au gel ont été menés au sein du laboratoire. Ces travaux ont montré de nombreuses adaptations physiologiques de cette plante au stress causé par des conditions de gel (KOLAKSAZOV et al., 2013 ; KOLAKSAZOV et al., 2014). De manière intéressante, les accessions utilisées comme "tolérantes au gel" pour ces travaux venaient d'une altitude élevée (col du Galibier, plus de 2500m) alors que les accessions "non tolérantes" venaient d'altitudes plus basses (Vercors, moins de 2000m, KOLAKSAZOV et al., 2013).

## 2.3 Début du suivi démographique

En 2007, lors de la thèse de Marco Andrello, il a été décidé de débiter un suivi démographique *in situ* sur 6 populations autour de Grenoble : trois populations dans le massif du Vercors et trois autour du col du Lautaret, ces 2 ensembles formant 2 gradients d'altitude distincts.

Dans chacune de ces populations, Marco Andrello a mis en place des quadrats contenant un nombre suffisant d'individus pour permettre un suivi démographique. *A. alpina* étant une espèce pérenne, ce suivi était individu-centré avec des mesures répétées sur les individus survivant les années suivantes, grâce à des étiquettes métalliques plantées dans le sol au pied de chaque individu du quadrat. Les nouvelles recrues étaient enregistrées chaque année pour procéder au même type de suivi. Différents types de traits phénotypiques, détaillés dans l'article suivant ont été enregistrés sur chaque individu.

Un des intérêts majeurs de ce suivi est que des sondes iButtons<sup>®</sup> ont été placées dans un coin de chaque quadrat pour mesurer, toutes les 3 heures, la température et l'humidité ambiante. Ces sondes étaient placées à peu près au niveau auquel pousse *A. alpina*. Ces données environnementales permettent donc d'avoir une idée assez précise des conditions environnementales que subissent les plantes et seront utilisées tout au long de ce chapitre et du suivant.

Au début de ma thèse, à l'été 2012, j'ai poursuivi ce suivi démographique, qui a ainsi été prolongé jusqu'en 2014 (la récolte des données environnementales s'est, elle, arrêtée en 2015). À l'aide de l'ensemble de ces années de suivi démographique incluant des données phénotypiques et environnementales, nous avons pu analyser les liens entre les taux vitaux (croissance, reproduction et survie) d'*A. alpina* et son environnement.

## 3 Présentation de l'ARTICLE 4

Cet article possède plusieurs buts. Tout d'abord, puisque l'on manque encore d'information concernant la biologie et la dynamique d'*A. alpina* dans son milieu naturel, il s'agira de fournir une des-

cription des facteurs de sa dynamique ayant un impact écologique (espérance de vie, survie et reproduction différentielle entre individus, etc...). Ensuite, il s'agira de décrire l'influence de la variabilité spatio-temporelle de l'environnement sur les taux vitaux des individus. Nous avons choisi, pour faciliter l'analyse, de séparer la variabilité temporelle au sein de chacun des sites d'une part, de la variabilité spatiale moyennée sur les différentes années du suivi d'autre part.

L'analyse des liens entre la variabilité environnemental spatiale et les taux vitaux nous intéresse particulièrement dans le cadre de cette thèse, puisque ces liens témoignent de la présence d'une contrainte de l'environnement sur les individus, qui induit des changements phénotypiques.

### 3.1 Contributions

Marco Andrello, Oscar Gaggiotti et Irène Till-Bottraud ont conçu l'étude et mis en place le protocole du suivi démographique. Marco Andrello et moi-même avons effectué les différentes campagnes de suivi (de 2007 à 2011 pour Marco Andrello et de 2012 à 2014 pour moi-même). Marco Andrello, Delphine Busson et moi-même avons analysé les données. Spécifiquement, Delphine Busson a procédé au traitement des données des sondes iButton<sup>®</sup>, la définition des variables thermiques et à une analyse préliminaire des variables thermiques et démographiques sous la direction de Marco Andrello tandis que j'ai effectué l'essentiel des analyses statistiques présentées dans cette version du manuscrit. Marco Andrello a rédigé l'essentiel du manuscrit, mais j'ai fortement participé à l'écriture du Matériel et Méthodes, des Résultats, ainsi qu'une partie conséquente de la Discussion. Marco Andrello, Irène Till-Bottraud, Oscar Gaggiotti et moi-même avons contribué à la réflexion scientifique et méthodologique depuis le début de l'étude (depuis 2012 pour moi).

### 3.2 Remarques

Cet article est en préparation pour *Journal of Ecology* et certaines des analyses restent préliminaires, notamment celles concernant la *compensation démographique* et la discussion sur l'évolution des traits d'histoire de vie.

La liste des auteurs est encore à définir, mais le consensus actuel place Marco Andrello en premier auteur et moi en second.

### 3.3 Résumé de l'article

Grâce à sa proximité génétique avec *Arabidopsis thaliana*, *Arabis alpina* (Brassicacée) est de plus en plus utilisée comme une espèce modèle pérenne pour des études d'évolution moléculaire adaptative. Nous avons étudié la démographie d'*A. alpina* dans six populations naturelles très différentes dans le degré de perturbation du sol, la pente et la végétation, et englobant l'entièreté de l'amplitude altitudinale de l'espèce.

Nous avons estimé trois taux vitaux (croissance, effort reproductif et survie) pour des plantes marquées individuellement et étudiées pendant un maximum de six ans (2008 – 2014). Nous avons caractérisé les conditions thermiques de chaque site à l'aide de différentes variables thermiques obtenues à partir de capteurs *in situ* de température en temps quasi-continu.

Bien qu'*A. alpina* soit décrite comme une espèce pérenne, l'espérance de vie était seulement de 1.82 ans et la plupart des plantes ont décédé avant de pouvoir produire des graines. La taille des plantes influait sur les trois taux vitaux. La température moyenne journalière possédait une relation positive avec la croissance et négative avec la survie. De plus, l'effort reproducteur covariait de manière négative avec la survie, suggérant un mécanisme de compensation démographique agissant le long du gradient altitudinal.

Ces résultats sont informatifs sur les pressions de sélection expérimentée par *A. alpina* dans les conditions naturelles et permettront d'aider à la conception d'études expérimentales et moléculaires sur l'adaptation locale chez cette espèce.

**Journal:** En préparation pour *Journal of Ecology*

Les *Supplementary Information* de cet article sont disponibles dans l'Annexe S4

## ARTICLE 4

### Population dynamics of *Arabis alpina* in the French Alps: evidence for demographic compensation?

Marco Andrello, Delphine Busson, Oscar E. Gaggiotti, Irène Till-Bottraud & Pierre de Villemereuil<sup>1</sup>

#### Abstract

Due to its genetic proximity with *Arabidopsis thaliana*, *Arabis alpina* (*Brassicaceae*) is increasingly used as a perennial model species in studies of molecular evolution and adaptation. We studied the demography of *A. alpina* in six natural sites widely differing in their degree of disturbance, slope and vegetation, and encompassing the full altitudinal range of the species. We estimated three vital rates (growth, reproductive effort and survival) for individually-marked plants, studied for six years (2008 – 2014). We characterized the thermic conditions of each site with different thermal variables obtained using *in situ* continuous-time data loggers. Although *A. alpina* is described as a perennial species, the average life expectancy was only 1.82 years and most plants died before setting seeds. Plant size was a strong predictor of all three vital rates. Mean daily temperature showed a positive effect on growth and a negative effect on survival. Furthermore, reproductive effort covaried negatively with survival, suggesting a mechanism of demographic compensation acting on an altitudinal gradient. Synthesis. These results are informative of the selective pressures experienced by *A. alpina* in natural conditions and will help design experimental and molecular studies of local adaptation in this species.

#### Introduction

The perennial herb *Arabis alpina* (L., *Brassicaceae*) is emerging as a new model organism in evolutionary developmental biology due to its proximity to *Arabidopsis thaliana* (KOCH et al., 2006). *A. alpina* is distributed in the Northern hemisphere and lives at elevations going from few meters above sea level in Northern regions (Scandinavia) to several thousand meters in African mountains. In the European Alps, *A. alpina* is restricted between 800 and 3000 meters (HUXLEY, 1986). Studies so far have focused on the genetic architecture of flowering and perenniality (R. WANG et al., 2009; BERGONZI et al., 2013), leaf senescence (WINGLER et al., 2012; WINGLER et al., 2015) and phenotypic traits as root hair and trichomes (CHOPRA et al., 2014). There is an ongoing effort in advancing the molecular knowledge of *A. alpina* by sequencing its genome (MELODELIMA & LOBRÉAUX, 2013; LOBRÉAUX et al., 2014; WILLING et al., 2015).

The ecology of *A. alpina* has been addressed mostly from the perspective of molecular ecology through studies of its genetic structure and phylogeny (ASSEFA et al., 2007; ANSELL et al., 2008; ANSELL et al., 2011; KARL et al., 2012; BUEHLER et al., 2012) and detection of loci under selection (PONCET et al., 2010; MANEL et al., 2010; MANEL et al., 2012; BUEHLER et al., 2012; ZULLIGER et al., 2013; BUEHLER et al., 2014) or with an evolutionary ecology perspective aimed at characterizing its mating system (TEDDER et al., 2015). A phenotypic study was however conducted through a transplant experiment aiming at detecting local adaptation (TORÄNG et al., 2015). Nonetheless, very few studies have tried to describe the population ecology of this species *in natura* and we lack basic knowledge on rates of survival and growth, reproductive effort and on the ecological factors driving the variation of vital rates in natural conditions.

---

<sup>1</sup>Cette liste est par ordre alphabétique, étant donné que l'ordre des auteurs n'est pas encore défini.

In this study, we investigated the demography of *A. alpina* in six field sites across two distinct mountain ranges in the French Alps, encompassing the entire altitudinal range of the species. Environmental variables were measured *in situ* to reflect local conditions. Gaining knowledge and understanding on the population ecology of *A. alpina* will help scientists working on this species to formulate hypothesis about the selective factors that are likely to drive the evolution of the species in the wild. For example, studies aiming at identifying genes under selection are easier if the selective factors are known (HANCOCK et al., 2011). While laboratory studies can reveal some potential selective factors, field conditions are complex and selective factors may differ from those that become important in laboratory conditions (BERGELSON & ROUX, 2010). This requires measuring vital rates (survival, reproductive effort and growth) *in natura*, describing their variation across individuals and geographical locations and understanding the ecological factors underlying such variations.

Vital rates of species vary in space and time in response to biotic and abiotic environmental conditions. Understanding how environmental conditions affect a species' vital rates is one of the main interests of population ecology. This is however challenging because of the complex relationships between an organism and its abiotic and biotic environment. For well-studied model species (e.g. *Arabidopsis thaliana* and some economically important species such as crops), there is some knowledge on the eco-physiological processes that determine survival rates, fecundity and growth. For non-model species, this knowledge is often absent.

In this case, a correlative approach can be employed: the correlation between some environmental variables and the vital rates of the species are tested and quantified. The choice of the variables is often done with very little a priori knowledge on the eco-physiology of the species. We took this approach to understand what environmental factors underlie the variation in vital rates (survival, fecundity and growth) in *A. alpina*. Specifically, we addressed three questions:

1. What are the demographic characteristics of natural populations of *A. alpina*?
2. What is the extent of variation in environmental conditions and population dynamics between different locations?
3. To what extent do environmental conditions affect population dynamics?

## Material & Methods

### Study species and study sites

The artico-alpine plant *Arabis alpina* (L., *Brassicaceae*) occurs in most European mountain ranges, Anatolia, northern and eastern Africa and subarctic North America (KOCH et al., 2006). *A. alpina* is often found in association with calcareous soil in disturbed habitats. In the European Alps, it is predominantly found in open and unstable grounds such as glacier forelands, scree slopes, rock ledges, and along foot paths and small streams in the montane, subalpine and alpine zones. It prefers moist sites with low vegetation cover. New seedlings germinate and establish throughout the growing season, plants flower for a few weeks and then flowers develop into fruits (siliques). The reproductive stems senesce and die while the plant overwinters as a rosette. Some degree of vernalisation is needed for the onset of floral development, which makes the plant perennial and iteroparous (R. WANG et al., 2009).

Six sites in three mountain ranges were studied (TAB. A4.1). Three sites (BRU, CHA and VIL) were located in the Vercors range, one site (LAU) was located in the Ecrins range near the Lautaret pass and two sites (GAL and PIC) were located in the Cerces range near the Galibier pass. For simplicity, we consider the Ecrins and Cerces sites as belonging to the same mountain range and refer to it as the Lautaret-Galibier mountain range. The maximal spatial distance between sites is 15 km in the Vercors and 4 km in the Lautaret-Galibier, while the average distance between mountain ranges is 65 km, thus supporting the choice of uniting LAU with GAL and PIC. The sites were located within a gradient encompassing the whole altitudinal range of *A. alpina* in the European Alps: the lowest site (BRU) was located at 930 m asl and the highest site (PIC) at 2930 m a.s.l. Besides elevation, the sites differed in aspect, soil type, slope, ground stability and vegetation (TAB. A4.1).

**TAB. A4.1:** Characteristics of the study sites. Latitude, longitude and elevation were measured with a GPS device. Elevation was further checked using the geographical coordinates on [www.geoportail.gouv.fr](http://www.geoportail.gouv.fr).

Mountain range	Site name	Site code	# plots	Elevation	Aspect	Latitude (N)	Longitude (E)	Soil type	Ground stability	Vegetation type
Vercors	Gorges du Bruyant	BRU	2	930	S	45.1534	5.6074	Rock scree, calcareous	Unstable	Understory
	Combe de Chaulange	CHA	3	1480	NW	45.0693	5.5913	Rock scree, calcareous	Stable	Shrubland
	Villard de Lans	VIL	3	1980	SW	45.0163	5.5700	Rock scree, calcareous	Unstable	Subalpine prairie
Lautaret-Galibier	Lautaret	LAU	3	2090	N	45.0268	6.3893	Stream side, schistes	Very unstable	Under Alnus viridis cover
	Galibier	GAL	3	2590	N	45.0587	6.4028	Shallow soil under rock cliff	Stable	Sparse
	Pic Blanc du Galiber	PIC	4	2930	E	45.0522	6.3838	Rock scree, schistes	Very unstable	Very sparse

**TAB. A4.2:** Used thermic variables. Variables in *italics* are those retained for the among-site comparison, the discriminant analysis and, except DD.2, as explanatory variables for the demographic variables. Subscript 1 and 2 indicate early and late season, respectively.

Variables	Description	Calculation
T.min.1, T.min.2	average daily minimal temperature	Mean of the daily minimal temperatures $T_{\min}$
<b>T.mean.1, T.mean.2</b>	average daily mean temperature	Mean of the daily mean temperatures $T_{\text{mean}}$
T.max.1, T.max.2	average daily maximal temperature	Mean of the daily maximal temperatures $T_{\max}$
<b>DD.1, DD.2</b>	accumulated growing degree days	Sum of the daily growing degree days $dd$
<b>Frost.1, Frost.2</b>	number of frost days	Number of days with at least one three-hourly measure below zero
Frost.e.1, Frost.e.2	number of frost events	A frost event is defined as a period of consecutive three-hourly measures below zero
<b>T.range.1, T.range.2</b>	average daily temperature range	Mean difference between daily maximal and minimal temperature, $T_{\max} - T_{\min}$



## Demographic study

In each site, we set up two to four permanent quadrats to contain a total of roughly 50 individual *A. alpina* plants. The study started in 2007 in Vercors and in 2008 in Lautaret-Galibier, and ended in 2014. The position of each plant was marked with a small metal tag nailed into the ground and charted to facilitate the census. Once a year, during summer, we recorded for each plant the vital stage (alive or dead), the number of reproductive and vegetative stems and the number of siliques per reproductive stem. Stems bearing siliques or flowers were considered reproductive, while rosettes and stems bearing only leaves were considered vegetative. We attempted to census the sites at the time of fruit maturation, but because of differences in phenology among plants and among individual stems of the same plant, some stems would often bear flowers and some siliques would not be completely developed. We nonetheless included flowers and immature siliques in the fruit count. A sheep herd trampled the LAU site in 2010 killing the majority of the plants. From census data, we computed the following variables for each individual and year  $t$ :

**Plant size** (year  $t$ ). The sum of reproductive and vegetative stems of the plant in year  $t$ .

**Growth** (from year  $t-1$  to year  $t$ ). The ratio of number of stems in year  $t$  over number of stems in year  $t-1$ .

**Reproductive effort** (year  $t$ ). The total number of siliques of the plant in year  $t$ .

**Survival** (from year  $t$  to year  $t + 1$ ). A binary variable indicating survival from year  $t$  to year  $t + 1$  (1 if the individual has survived and 0 if it has not).

To estimate seed set, we sampled 20 plants in each of three sites, VIL, LAU and GAL, outside but close to the permanent plots. We counted the seeds of each silique and averaged it over the siliques of the same plant (average seed set).

Plant age was determined only for individuals followed since their recruitment to the population, therefore from 2009 onwards. A power function describing the number of individuals as a function of age,  $N = ar^{\text{age}}$ , was fitted using nonlinear least squares in the R software. Life-expectancy was calculated as the average age at death across all cohorts.

## Thermal variables

We installed a temperature data logger (iButton<sup>®</sup> hygrochron, Maxim Integrated<sup>™</sup>) in each quadrat at about 10 cm above ground and sheltered it from direct sunlight with a square and opaque Plexiglas plate. The data loggers recorded air temperature at the nearest half-degree every three hours. Data were recovered once per year during the annual census. Measures from quadrats of the same sites were highly correlated ( $R^2 \approx 0.90$ ). Some data loggers were lost and for some others the data was unreadable due to battery exhaustion or other unknown causes, especially in the last years of the survey. These data loggers were replaced with new ones; however, this resulted in loss of data.

To replace missing data, we duplicated the data from the quadrats of the same site. When data were missing for all the quadrats of the site for a short period of time (23 occurrences), we reconstructed the missing data from nearby weather stations of the “Centre de Recherches sur les Ecosystèmes d’Altitude” (CREA – Mont-Blanc, <http://creamontblanc.org>): stations 207 and 209 for the Vercors sites, and 218 and 29-217 for the Lautaret-Galibier sites. These stations were on average 12 km away from the closest study site (TAB. S4.1 in Supplementary Information) and record temperature every 15 minutes. We fit a simple linear model with the temperature of the study site as response variable and the temperatures at ground level, 30 cm and 2 m above ground level of the two weather stations, as well as hour of the day, as explanatory variables. The models had a very good fit (average  $R^2 = 0.82$ ) and provided a reliable way to estimate the missing observations. To account for uncertainty in the estimates of the coefficients of the linear model, 10 replicated datasets were generated using values sampled from normal distributions defined by mean and standard error of the estimates of the coefficients

From the raw three-hourly measures, we calculated daily mean, minimal and maximal temperatures ( $T_{\text{mean}}$ ,  $T_{\text{min}}$  and  $T_{\text{max}}$ ). These data were used to define thermal variables for the vegetative season; we disregarded the non-vegetative season (autumn and winter) because plants are dormant and less sensitive to environmental conditions. The limits of the vegetative season were derived from temperatures through the curve of the growing degree days, defined as  $dd = (T_{\text{max}} + T_{\text{min}})/2 - T_{\text{base}}$ . We set  $T_{\text{base}} = 0$ , because arctic-alpine plants can grow at very low temperatures.

For each quadrat, we determined graphically when the cumulative curve of  $dd$  started to increase and when it started to stabilize and used these two dates as the limits of the vegetative seasons (FIG. S4.1 in Supplementary information). The vegetative season was further divided into two consecutive periods separated by the census date, to define thermal variables that had the same temporal structure as the census data. For simplicity, we refer to the first period as early season and the second period as late season. TAB. S4.2 (Supplementary information) lists the starting and ending dates of the vegetative season and the census dates for each quadrat and year. Three-hourly and daily temperatures were used to define seven thermal variables for the early and late seasons (TAB. A4.2).

## Statistical analysis

We analysed the thermal variables separately for the early and late season. Correlations between variables were measured by Pearson's linear correlation coefficient. We reduced the number of thermal variables used in the subsequent analyses by retaining only the ones showing a pairwise Pearson's correlation coefficient below 0.7 (TAB. A4.2).

Vital rates (growth, reproductive effort and survival) were analysed using generalized linear mixed models including plant size and the uncorrelated thermal variables as potential explanatory variables. Two distinct sets of analyses were performed, a site-specific analysis and a global analysis.

In the site-specific analysis, we analysed vital rates and environmental data separately for BRU, VII and GAL (the other sites had too few observations to provide meaningful results). Quadrat and year were included as random effects.

In the global analysis, we averaged vital rates, plant size and thermal variables over the years of the same site. Thus, we modelled the relationship between average thermal conditions and average vital rates. In the global analysis, we weighted the data points by the number of individuals used to compute the means to account for small sample sizes in some populations (e.g. PIC) and site was included as a random effect. In both the site-specific and the global analysis, the explanatory variables were standardized to facilitate the comparison of their effects.

Growth was modelled as a function of plant size in year  $t - 1$ , the environmental variables of the late season of year  $t - 1$  and the environmental variables of the early season of year  $t$ . This choice of variables respect the time frame of the demographic process of growth, which takes place during the time from the census of year  $t - 1$  until the census of year  $t$ . In the site-specific analyses, growth was log-transformed to achieve normality and variance homogeneity of the residuals, and plant size was log-transformed to obtain a log-linear relationship between plant size and growth, which corresponds to a power-function between size at time  $t - 1$  and size at time  $t$ . In both the site-specific and global analyses, the error structure was Gaussian.

Reproductive effort was modelled as a function of plant size in year  $t$  and the environmental variables of the early season of year  $t$ . In the site-specific analyses, the error structure was negative binomial to correctly model over-dispersed count data. In the global analysis, the error structure was Gaussian.

Survival was modelled as a function of plant size in year  $t$ , the environmental variables of the late season of year  $t$  and the early season of year  $t + 1$ , plus the interaction of all these variables with plant size. Ground stability was included as a categorical variable (TAB. A4.1) and the effect of sheep trampling in LAU in 2010 as a binary variable. In the site-specific analyses, the error structure was binomial. In the global analyses, the error structure was Gaussian.

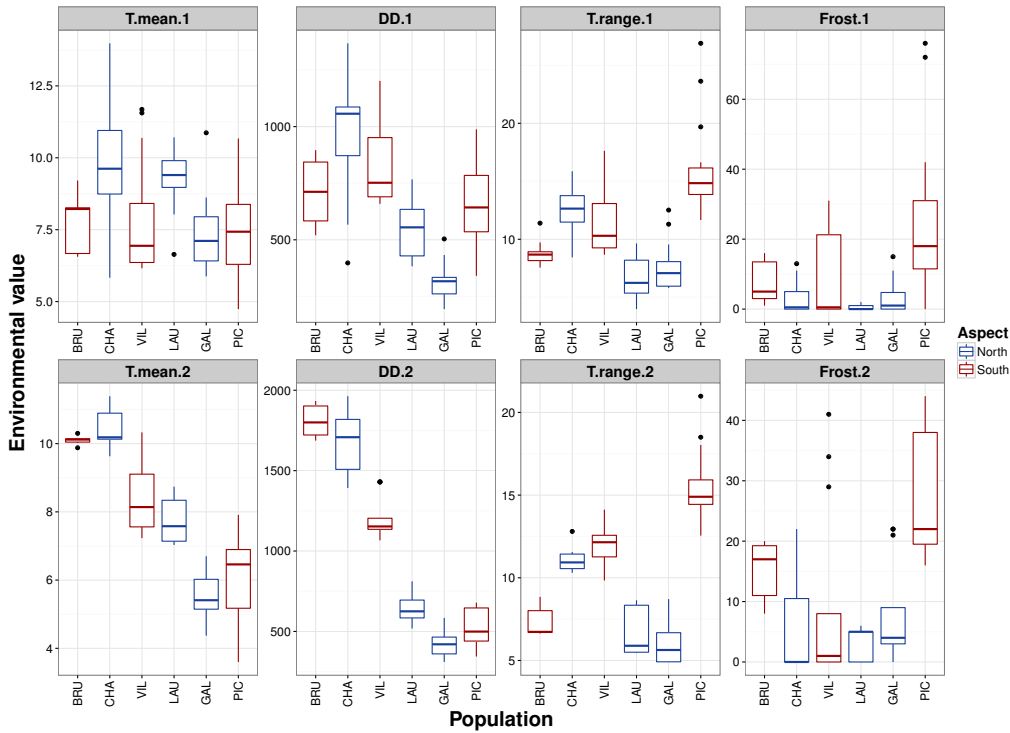


FIG. A4.1: The four thermal variables in the six sites for early (top row) and late (bottom row) season. Sites are ordered according to altitude. Blue: Northern aspect; Red: Southern aspect

Because of the uncertainty in the estimates of missing temperature measures by weather station measures, we used the 10 replicated datasets in a multiple imputation framework: we fit a model on each of the 10 dataset, averaged the estimates over models and computed the degrees of freedom according to BARNARD & RUBIN (1999). The efficiency of the multiple imputation framework was always above 0.99, indicating that 10 replicated datasets were largely sufficient to recover missing information.

Because the multiple imputation framework relies heavily on *t*-tests, we selected the best models using a sequential *t*-test approach: the full model was fitted and variables with non-significant estimates were discarded; a new reference model was then fitted, and non-significant variables were again discarded until all remaining variables were significant. We then tried to add back all variables and created a new reference model with significant variables. These steps were repeated until a situation was reached where no variable could be added or removed to the reference model, which was then considered to be the best model.

All analyses were performed in R with the packages *ade4* v.1.7 (DRAY & DUFOUR, 2007), *lme4* v.1.1 (BATES et al., 2015), *mice* v.2.22 (van BUUREN & GROOTHUIS-OUDSHOORN, 2011) and *mitml* v.0.2.

### Demographic compensation

Demographic compensation was tested using linear models between two vital rates at a time (average survival, average growth, average reproductive effort) and weighting the data points by the number of individuals used to compute the means to account for small sample sizes in some populations.

## Results

### Thermal variables

On the basis of pairwise correlation coefficients among the initial seven thermal variables for each period (TAB. S4.3 Supplementary information), we retained only four variables for early season

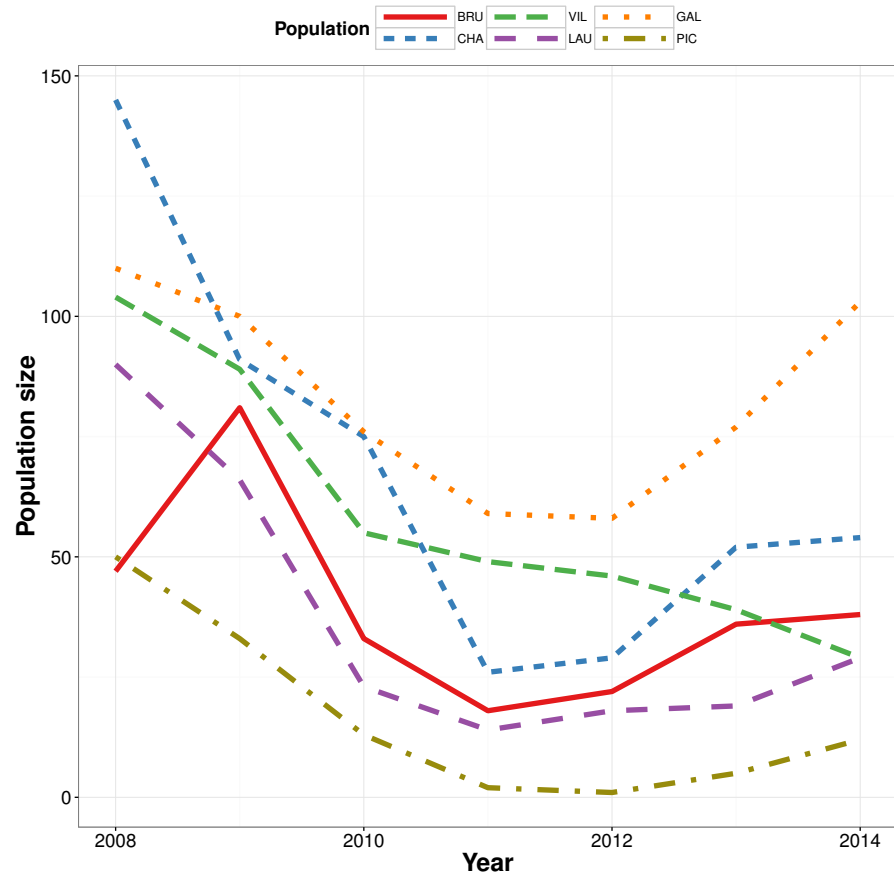


FIG. A4.2: Number of *Arabis alpina* individuals in each site for each year of study.

(T.mean.1, DD.1, Frost.1 and T.range.1) and three variables for late season (T.mean.2, Frost.2 and T.range.2) (Table 2). FIG. A4.1 shows the variation in thermal variables among sites.

Whereas variation among sites in the early season's mean daily temperature was significant but small (T.mean.1,  $\text{adj-}R^2 = 0.20$ ,  $F_{5,87} = 5.5$ ,  $p < 0.001$ ), the later season's mean daily temperature clearly varied with site (T.mean.2,  $\text{adj-}R^2 = 0.82$ ,  $F_{5,67} = 67.9$ ,  $p < 0.001$ ) and elevation (T.mean.2 =  $a + b\text{Elevation}$ ,  $a = 12.77 \pm 0.43^\circ\text{C}$ ,  $b = 0.002 \pm 0.0002$ ,  $\text{adj-}R^2 = 0.68$ ,  $p < 0.001$ ). The accumulated degree days were higher in the (lower elevation) Vercors sites than in the (higher elevation) Lautaret-Galibier sites, both in early (DD.1 =  $843 \pm 234\text{dd}$  in Vercors vs.  $481 \pm 229\text{dd}$  in Lautaret-Galibier;  $\text{adj-}R^2 = 0.37$ ,  $F_{1,95} = 57.7$ ,  $p < 0.001$ ) and late season (DD.2 =  $1449 \pm 289\text{dd}$  in Vercors vs.  $562 \pm 130\text{dd}$  in Lautaret-Galibier;  $\text{adj-}R^2 = 0.82$ ,  $F_{1,71} = 319.1$ ,  $p < 0.001$ ). The number of frost days was low on average, but PIC showed higher occurrences for the early season (Frost.1 =  $24 \pm 19$  days in PIC vs.  $5 \pm 8$  days in the other sites,  $t = 3.71$ ,  $p < 0.001$ ) and especially for the late season frosts (Frost.2 =  $34 \pm 14$  days in PIC vs.  $9 \pm 11$  days in the other sites,  $t = 4.57$ ,  $p < 0.001$ ). Temperature range exhibited moderate differences between sites ( $\text{adj-}R^2 = 0.60$ ,  $F_{5,87} = 29$ ,  $p < 0.001$  for the early season and  $\text{adj-}R^2 = 0.87$ ,  $F_{5,67} = 98$ ,  $p < 0.001$  for the late season) and ranged from  $9 \pm 1^\circ\text{C}$  in BRU to  $15 \pm 3^\circ\text{C}$  in PIC in the early season and from  $6 \pm 1^\circ\text{C}$  in GAL to  $16 \pm 2^\circ\text{C}$  in PIC in the late season.

PIC was thus the site experiencing the harsher conditions, i.e. large temperature ranges and high occurrence of frost events.

## Demography

VIL showed a decreasing trend in population size over 2008-2014 (FIG. A4.2). In the other sites, population sizes decreased from 2008 to 2011, then increased until 2014. At the end of the study,

population sizes of all sites except GAL were smaller than at the beginning. PIC nearly experienced extinction with only one individual remaining in 2012.

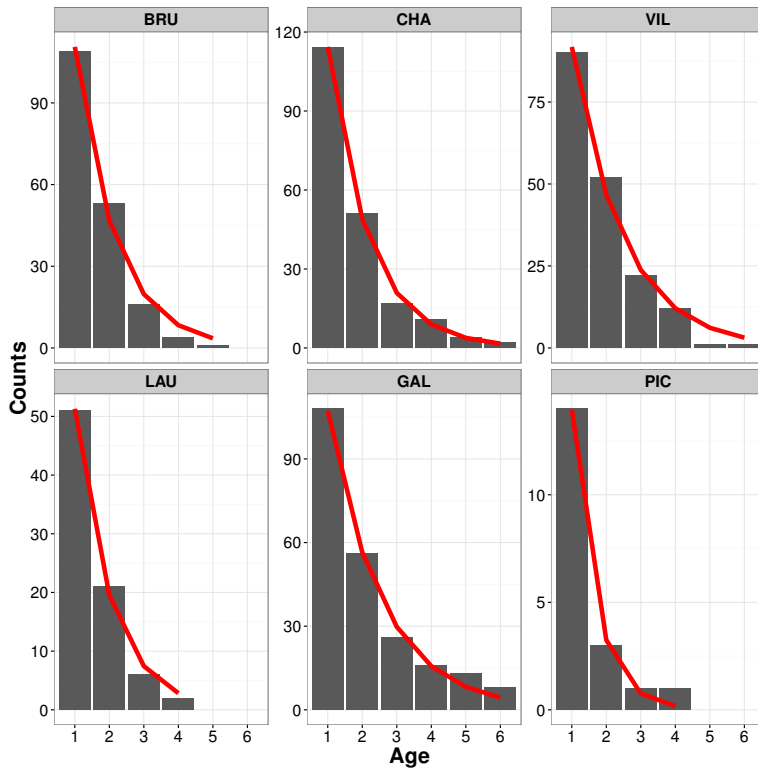


FIG. A4.3: Number of *Arabis alpina* individuals in the six sites for each age-class, pooled over the six years of study. The red line is an exponential curve fitted to the data:  $N = ar^{\text{age}}$

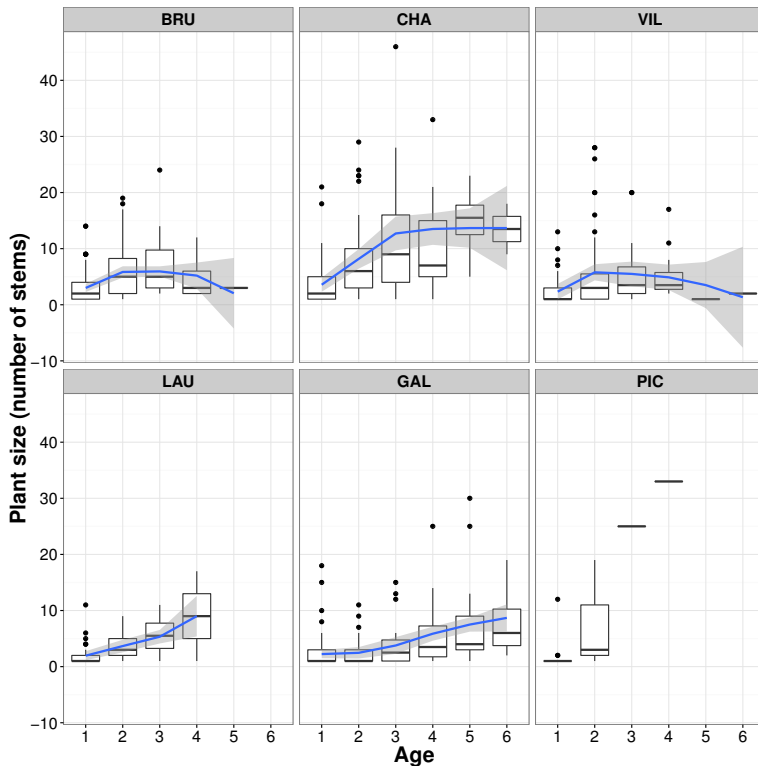


FIG. A4.4: Distribution of plant size per site and age-class. The blue line is a local LOESS smoothing and the shaded grey areas are 95% confidence intervals of the LOESS algorithm.

The age distribution of plants was similar across sites, with decreasing number of individuals with age (FIG. A4.3). The age distribution was well described by a power function  $N = ar^{\text{age}}$ , where  $r$  ranged between 0.23 (in PIC) and 0.53 (in GAL). Life-expectancy was 1.82 years (ranging from 1.42 in PIC to 2.09 in GAL). PIC was the site with shortest-lived plants, with 74% plants dying just after their first year, while the site with the longest-lived plants was GAL, where the plants dying just after their first year were only 48%.

Average plant size ranged between 3.0 stems in LAU to 6.5 stems in CHA and was 4.4 stems on average. Most plants were small: the 75% percentile of the size distribution across sites was 5 stems. The average size of plants increased with age (FIG. A4.4). However, in BRU and VIL the average size of older plants was smaller than the size of medium-aged plants and size reached a plateau in CHA.

Average survival rates increased with plant size, except in CHA, where the highest survival was observed for medium-sized plants. On average, 47% of plants died before becoming reproductive (ranging from 34% in BRU to 79% in PIC). Among reproductive plants, the average number of siliques per plant was 23.8 (ranging from 6.6 in GAL to 89.4 in PIC) and the average number of siliques per reproductive stem was 5.8 (ranging from 3.0 in GAL to 10.8 in PIC).

The average number of seeds per fruit was 25.1 and was not significantly different between the three studied sites ( $F_{2,52} = 0.036$ ,  $p = 0.96$ ).

**TAB. A4.3:** Effects of plant size and thermal variables on temporal variation in **growth** of *A. alpina* for three sites (BRU, VIL and GAL).

	Estimate	S.E.	<i>t</i> value	d.f.	<i>p</i> value
BRU					
(Intercept)	0.56	0.12	4.51	109.05	< 0.001
Log(Plant size)	-0.45	0.07	-6.47	109.05	< 0.001
T.mean.1	0.35	0.07	4.77	109.05	< 0.001
VIL					
(Intercept)	0.16	0.09	1.81	140.04	0.036
Log(Plant size)	-0.59	0.06	-9.24	140.04	< 0.001
DD.1	0.14	0.08	1.84	140.04	0.034
GAL					
(Intercept)	0.14	0.14	1.01	353.02	0.157
Log(Plant size)	-0.27	0.03	-7.96	352.73	< 0.001
Frost.2	0.09	0.04	2.35	350.72	0.010
T.range.2	-0.08	0.04	-1.81	350.38	0.035
Frost.1	-0.17	0.04	-4.76	351.43	< 0.001

**TAB. A4.4:** Effects of plant size and thermal variables on temporal variation in **reproductive effort** of *A. alpina* for GAL.

	Estimate	S.E.	<i>t</i> value	d.f.	<i>p</i> value
GAL					
(Intercept)	0.19	0.30	0.65	451.01	0.257
Plant size	1.29	0.12	10.98	451.01	< 0.001
T.range.1	0.30	0.18	1.71	451.01	0.044

### Site-specific analyses of vital rates

Growth was negatively dependent on plant size in the three sites (TAB. A4.3). Growth was positively associated to T.mean.1 in BRU and to DD.1 in VIL. These two variables were statistically linked in these two populations (Pearson's linear correlation coefficient  $\rho = 0.85$ ,  $p < 0.001$ , in BRU;  $\rho = 0.67$ ,  $p < 0.001$ , in VIL), indicating a similar pattern of growth variation in these populations. Growth in GAL had a totally different behaviour with a strong negative association with Frost.1, as well as a positive association with Frost.2 and a negative association with T.range.2.

Reproductive effort in GAL was positively dependent on plant size and T.range.1 (TAB. A4.4). Models for reproductive effort in BRU and VIL did not converge.

Survival showed the most complex pattern (TAB. A4.5). Notably, survival was positively correlated with plant size in VIL and GAL, but not in BRU. In fact, survival in BRU appeared to be only positively linked to one variable (T.mean.1). Inferences for VIL and GAL were more complicated with significant interactions with plant size, indicating that larger plants had a different sensitivity to thermal conditions than smaller plants.

### Global analyses of vital rates

Average growth depended strongly and negatively on plant size, indicating that relative growth decreased as plant size increased (TAB. A4.6). Average growth depended positively on T.mean.2 and negatively on T.range.2 with comparable effect sizes: indeed, growth was much larger in the warmer populations BRU and CHA, and smaller in the colder populations GAL and PIC. Notably,

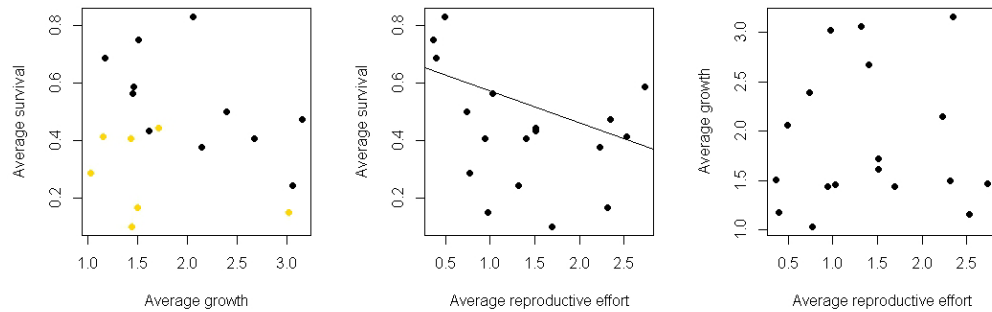
**TAB. A4.5:** Effects of plant size and thermal variables on temporal variation in survival of *A. alpina* for three sites (BRU, VIL and GAL)

	Estimate	S.E.	<i>t</i> value	d.f.	<i>p</i> value
<b>BRU</b>					
(Intercept)	-0.30	0.17	-1.76	166.04	0.040
T.mean.1	0.90	0.17	5.28	166.04	< 0.001
<b>VIL</b>					
(Intercept)	-0.11	0.33	-0.33	234.02	0.372
Plant size	0.54	0.17	3.16	234.02	0.001
T.mean.1	-0.54	0.17	-3.19	234.01	0.001
T.range.2	0.39	0.20	2.00	233.95	0.024
Plant size : T.mean.1	0.45	0.19	2.37	234.02	0.009
<b>GAL</b>					
(Intercept)	1.39	0.15	9.31	355.92	< 0.001
Plant size	0.46	0.18	2.54	355.87	0.006
T.range.2	-0.53	0.13	-4.00	355.58	< 0.001
DD.1	0.26	0.14	1.84	354.01	0.033
Plant size : Frost.1	0.47	0.21	2.25	355.84	0.012

**TAB. A4.6:** Effects of plant size and thermal variables on spatial variation in average vital rates (growth, reproductive effort and survival) of *A. alpina*

	Estimate	S.E.	<i>t</i> value	d.f.	<i>p</i> value
<b>Growth</b>					
(Intercept)	1.840	0.158	11.637	11.374	< 0.001
T.mean.2	0.467	0.140	3.338	11.373	0.003
T.range.2	-0.387	0.186	-2.076	11.373	0.031
<b>Reproductive Effort</b>					
(Intercept)	1.373	0.253	5.428	12.353	< 0.001
Plant size	0.412	0.151	2.727	12.353	0.009
<b>Survival</b>					
(Intercept)	0.450	0.045	10.022	8.460	< 0.001
Plant size	0.079	0.017	4.679	8.416	0.001
T.mean.2	-0.139	0.038	-3.685	8.432	0.003
Frost.2	-0.089	0.038	-2.336	8.380	0.023
T.mean.1	-0.065	0.023	-2.850	8.421	0.010
Frost.1	-0.082	0.031	-2.664	8.382	0.014

**FIG. A4.5:** Pair-wise correlations between average survival, average growth and average reproductive effort of *A. alpina*. Yellow points in the left panel corresponds to the LAU and PIC sites. The line in the middle panel is a fitted linear weighted regression.



growth depended only on late-season variables of the previous year, while early season variables of the current year were not significant.

Average reproductive effort depended negatively on plant size, all other effects being non-significant. GAL was the only site with consistently lower reproductive effort compared to the other sites (ANOVA between sites,  $F_{5,12} = 3.32$ ,  $p = 0.041$ ; GAL effect,  $t = -3.49$ ,  $p = 0.0045$ ).

Average survival was positively affected by plant size and negatively affected by T.mean.1, T.mean.2, Frost.1 and Frost.2. Average daily mean temperature in late season (T.mean.2) had the strongest effect. Indeed, this variable is an excellent predictor of average survival, when excluding LAU and PIC ( $R^2 = 0.82$ ), whereas the other variables allowed to account for the particular pattern of LAU (T.mean.1) and PIC (Frost.1 and Frost.2).

## Demographic compensation

Average survival was negatively, but not significantly, correlated to average growth ( $a = 0.75$ ,  $b = -0.11 \pm 0.07$ ,  $\text{adj-}R^2 = 0.08$ ,  $F_{1,16} = 2.41$ ,  $p = 0.14$ , FIG. A4.5). The correlation was slightly stronger when LAU and PIC were removed from the analysis ( $a = 0.81$ ,  $b = -0.13 \pm 0.08$ ,  $\text{adj-}R^2 = 0.13$ ,  $F_{1,12} = 2.97$ ,  $p = 0.11$ , FIG. A4.5). The relationship between average survival and average reproductive effort was negative and significant ( $a = 0.68$ ,  $b = -0.11 \pm 0.05$ ,  $\text{adj-}R^2 = 0.18$ ,  $F_{1,16} = 4.66$ ,  $p = 0.04$ ). However, the pattern was not linear and better described as a U-shaped (see FIG. A4.5). The relationship between average growth and average reproductive effort was not significant ( $\text{adj-}R^2 = 0.01$ ,  $F_{1,16} = 1.2$ ,  $p = 0.29$ ).

## Discussion

### Demography

The six natural sites in which we have studied the demography of *Arabis alpina* differed strongly in their abiotic conditions. Slope, aspect, ground stability and thermal conditions varied greatly among sites. Despite these differences, the six sites showed some similarities in demography, particularly in the temporal variation of plant number during our study (FIG. A4.2).

Although *A. alpina* is described as a perennial (R. WANG et al., 2009 ; BERGONZI et al., 2013 ; WINGLER et al., 2015), life expectancy was only 1.8 years on average. We observed variation in life expectancy, with some plants persisting for the entire duration of our study (i.e. up to 6 years), but at least half of the plants (up to 74% in PIC) died in their first year, with the consequence that half of the plants (up to 79% in PIC) never set seed. This high mortality is not an artefact of high



seedling mortality, which is expected to be naturally high, because plants were marked only when they had several developed leaves.

Life-history theory predicts that age at first reproduction decreases as mortality increases (ROFF, 2001), therefore earlier reproducers should be favoured in sites as PIC where mortality is high. In many alpine species, the timing of reproduction is regulated by vernalisation requirements and the genes controlling the plant responses to vernalisation have been studied in detail in *A. alpina* (R. WANG et al., 2009 ; BERGONZI et al., 2013). If functional genetic variation exists at these genes, alleles relaxing vernalisation requirement or conferring completely annual habits may be favoured in sites characterized by high plant mortalities. Yet another selective pressure would be the increase in resources limitation with altitude, implying a life-history response in which growth and reproduction are slowed down and adult survival and lifespan is increased (KÖRNER, 2003), as is apparently the case in the population GAL. As a consequence, the fitness advantage of an annual mutant should be evaluated considering all the components of the life-cycle and in particular the advantages of growing larger to produce more seeds, and the increasing survival rates at larger sizes (HORVITZ & SCHEMSKE, 2002). Indeed, for some years in PIC only a single large individual, bearing hundreds of siliques, was reproductive. More information regarding adaptive characteristics (e.g. fitness and genetics of life-history traits) in PIC population would be required to disentangle both evolutionary scenarios.

Average plant size increased with age. In some sites, average plant size showed a decreasing or stabilising pattern at older ages. This was the result of year-to-year variation in the number of stems of individual plants combined with a small sample size of long-living plants. Indeed, individual plant size in a given year was the sum of the number of stems developed in the same year and the number of stems overwintering from the previous year. Reproductive stems always die at the end of the reproductive season and cannot overwinter, while vegetative stems can overwinter but are subject to winter stochastic mortality. For many individuals, the obligate mortality of reproductive stems very often led to a decrease in plant size, because the new stems were not sufficient to replace the old ones. As a result, size variation was widespread in our data, which combined to issues of small sample size of long-living plants, sometimes caused this negative trends (in BRU and VIL, this trend is caused by <10 individuals).

We defined plant size as the number of stems without considering variation in biomass among stems, which was important but would have been too time-consuming and invasive to measure. Thus, patterns of variation in plant size should not be interpreted as a proxy for biomass variation or fitness. Nonetheless, the number of stems was useful to explain variation in survival, growth and reproductive effort. We thus advise using the number of stem as a simple and cost-effective metric to predict individual plant demographic rates of *A. alpina*, but we also encourage further studies of alternative metrics that may be better correlated to vital rates (CASWELL, 2001), such as plant diameter or the sum of stems weighted by the number of leaves.

Reproductive effort, measured as the number of siliques per plant, was highly variable and depended on plant size. Large plants could bear hundreds of siliques. The number of siliques per stem was less variable and ranged between 3 and 10. Each fruit bore on average 25 seeds and the number of seeds per fruit showed larger variation between individuals than between sites (only three sites studied).

## Effect of thermal variables on vital rates

Explaining variation in plant growth, survival and reproduction on the basis of topographical and geographical site characteristics is challenging because plant life is the result of many interacting factors such as elevation, aspect and microtopography (KÖRNER, 2007). This is why we decided to use local environmental data collected with data loggers located at vegetative height (a few cm above ground). This avoids the confusion of using environmental data from weather stations located many km away, which may not be representative of the environmental conditions experienced by the plants. Nonetheless, when looking at short periods of time, temperatures measured in nearby

weather stations were correlated to temperatures measured *in situ* and could be used to reconstruct missing data within the vegetative season.

We used thermal variables estimated on periods that are relevant for the vegetative cycle of the plants rather than relying on calendar-based measures. Studies on alpine plants often use calendar-based variables such as monthly mean temperatures or annual minimal temperature to infer biogeographic patterns of species occurrence (AUSTIN & VAN NIEL, 2011) or local adaptation (MANEL et al., 2012). While these data may lead to significant correlations at large spatial scale, it is difficult to interpret them at the scale of individual sites because the same calendar month corresponds to different phenological phases in different sites. Our choice of defining the vegetative phase on the basis of *in situ* temperature instead of calendar dates proved more realistic.

Using these data, we studied the relationships between vital rates (growth, reproductive effort and survival) and thermal variables. The effects of thermal variables on temporal variation in vital rates were not consistent across sites. We found comparable results for only two sites and one vital rate (growth rate for BRU and VIL, both located in the Vercors mountain range). Relationships between survival and thermal variables were highly site-specific. As survival is an integrative trait (i.e. influenced by many features of the individuals) and strongly submitted to local environmental and stochastic factors (e.g. random death due to accident, predation or disease), it might be expected that its variations are to be highly influenced by the local conditions.

Spatial variations in vital rates exhibited a clearer pattern. Spatial variation in average growth and survival rates were dominated by the mean daily temperatures of late season (T.mean.2), which strongly covaried with elevation (Pearson's linear correlation coefficient  $\rho = -0.90$ ,  $p < 0.001$ ). Growth rate was further affected by late season's mean daily temperature range (T.range.2). This indicates that early environmental conditions have little effect on plant growth and, therefore, that growth mostly happens during the late season. However, this relationship may also be the result of smaller between-site variation in early season than in late season thermal variables (FIG. A4.1). In addition to T.mean.2, average survival was affected by early season's mean daily temperature (T.mean.1) and by the number of frost days in the early and late season (Frost.1 and Frost.2); a refined analysis showed that these latter variables explain average survival for the two sites not following the T.mean.2 trend (LAU and PIC). More specifically, the relationship between growth and T.mean.2 was positive, whereas the relationship between survival and T.mean.2 was negative.

A negative relationship between survival and growth would imply demographic compensation (DOAK & W. F. MORRIS, 2010; VILLELLAS et al., 2015). Demographic compensation is implicated in shaping species distribution limits, where decreases in one vital rate at the species edge are compensated by increases in another vital rate. Over an altitudinal range or an environmental gradient, the mechanisms should be similar. We observed a negative relationship between survival and growth, but it was not significant, because of the small number of sites in our study. However, should this pattern be general, it would mean that *A. alpina*, at the specific level, follows a known rule of temperate alpine ecology, namely that high elevation plants tend to have higher survival and lower growth rates than lower elevation plants (KÖRNER, 2003), because of physiological (e.g. lower temperature, lower partial CO<sub>2</sub> pressure) and ecological (e.g. shorter reproductive seasons) processes. We observed, however, a significant negative relationship between survival and reproductive effort.

In this context, the case of the GAL site is interesting. Despite not being the highest site, it displays strong signs of the syndrome described above, since it had the highest survival rate and the second lowest growth rate and the lowest reproductive effort. Moreover, GAL was the only site showing a lower average reproductive effort than the other sites, and reproductive effort was not related to thermal variables. Despite similar temporal patterns of population dynamics among sites during the study (FIG. A4.2), vital rates had a highly site-specific behaviour and the relationships between vital rates and thermal variables do not appear to be fully generalizable across sites. However, an effect of the late season's mean daily temperature seems to emerge from the analysis of spatial variation in average vital rates (at least for growth and survival), although site-specific behaviour

again hampers strong conclusions regarding this phenomenon (especially for survival). More sites would be needed to ensure the generality of the observed trade-off between growth and survival. Yet, the possible existence of this adaptive compromise already raises the issue of whether it has a evolutionary origin (local adaptation) or not (phenotypic plasticity). An answer to this question would require in depth studies including common garden experiments combined with genome scans and association studies.

## **Acknowledgements**

We wish to thank Serge Aubert, Rolland Douzet and Bénédicte Poncet for helping us find the populations, Philippe Choler for interesting suggestions on the thermal determination of the vegetative season, Anne Delestrade for providing us with the weather station data and all the people helping with the field work, in particular Stefan Willhoit, Blaise Tymen, Elsa Jullien, Jérémy Camazzola, Anne-Lise Bartalucci, Lucas Hemery, Julie Chauvin, Florian Alberto, Loïc Chalmandrier, Aymeric Pilleux, Charles Pouchon, Thomas Goeury, Tristan Cumer, Perrine Augrit and Justine Bisson. MA and PdV were funded by a PhD scholarship by the French Ministry of Research.

## 4 Quelques données plus anecdotiques

Certaines des observations et données recueillies sur le terrain n'ont pas figuré dans cet article. L'une de ces observations les plus importantes est qu'*A. alpina* est capable de reproduction clonale. Elle possède en effet un organe, ressemblant à un petit rhizome, mais aérien, qui lui sert à projeter de nombreuses tiges loin de sa racine pivot principale.

De ce fait, *A. alpina* possède un port très complexe, fait de tiges très longues, parfois très excentrées par rapport au reste de la plante. Ces tiges (principalement des tiges reproductives) et les petits rhizomes possèdent des racines adventives qui permettent à la plante de "s'étaler" horizontalement de manière très efficace. Quand, par accident, une coupure apparaît au niveau de ces rhizomes, les deux parties de la plante sont capables (du fait de la présence des racines adventives) de survivre indépendamment. Ce phénomène résulte donc effectivement en une reproduction clonale de la part d'*A. alpina*. Quelques occurrences de ces reproductions clonales sont survenues durant le suivi démographique. Cependant, la détection de ce type de phénomène reste occasionnelle et repose sur des preuves indirectes (proximité des plantes, reste du rhizome scindé...). Il reste donc difficile de quantifier l'ampleur de ce type de reproduction dans les populations naturelles.

Une autre observation, juste mentionnée dans l'article, est que les terrains sur lesquels poussent *A. alpina* sont très instables. Sans qu'il soit possible de le quantifier précisément, une proportion importante de la mortalité dans les quadrats était causée par des glissements ou éboulis de très faible ampleur. Des glissements et éboulis de plus fortes ampleurs (à l'échelle du quadrat entier) ont eu lieu dans quelques populations : ces événements étaient nombreux pour les populations PIC et LAU (toutes deux présentes sur un terrain schisteux, très instable). Pour la population LAU, lors du nettoyage du site en 2015, un des quadrats, situé sous un aulne vert au bord d'une rivière, avait été totalement ravagé par un glissement de terrain projetant l'aulne dans la rivière.

Que ces deux sites aient une très faible survie moyenne et ne suivent pas la tendance entre température moyenne et survie pourrait être lié à cette instabilité très forte (liée à la nature schisteuse du terrain). Cependant, il faudrait plus de populations pour réellement tester cette hypothèse.

## 5 Conclusions principales de l'article

### 5.1 Sur la biologie générale d'*Arabis alpina*

Les données générales sur les taux vitaux, sont, en elles-mêmes extrêmement intéressantes. Par exemple, alors qu'*A. alpina* est effectivement une plante pérenne, on réalise que son espérance de vie est tout de même réellement courte (1,82 ans) et que 51% des individus ne vivent qu'une année (tandis que certains ont survécu à l'intégralité du suivi démographique). La reproduction aussi est très inégale, avec une forte proportion des individus (47%) qui n'ont pas produit de fruits lors de leur existence.

Ces résultats décrivent assez bien l'écologie d'*A. alpina* comme une plante pionnière, clairement à stratégie  $r$  (faible survie, forte reproduction, avec un renouvellement rapide des populations, PIANKA, 1970). La reproduction étant largement inégale, et les individus qui survivent jusqu'à l'année suivante se reproduisant plus que ceux qui ne survivent pas, il est clair que certains individus participent de manière démesurée aux générations suivantes, ce qui pourrait avoir de fortes conséquences en terme de taille efficace des populations et de consanguinité.

Un résultat intéressant porte sur la taille des plantes. Les courbes de “croissance” des individus sont extrêmement variables, avec des gains ou des chutes du nombre de tiges qui peuvent être très impressionnantes. En effet, si la croissance individuelle (conditionnellement au fait que l’individu survive) est en moyenne de 1.75 tiges par an, ce chiffre varie entre 26 tiges perdues et 37 tiges gagnées en un an (données non montrées dans l’article).

Ceci est dû au fait que la vaste majorité des tiges, chez *A. alpina*, sont des tiges reproductives qui meurent à la fin de la saison de reproduction. La plante croît et décroît donc considérablement entre l’hiver et l’été, d’autant plus qu’elle a produit de tiges reproductives.

Ce constat de perte considérable de biomasse à l’automne soulève la question du stockage des ressources chez cette espèce durant l’hiver. Il se trouve que presque rien n’est connu à ce sujet. Cependant, on pourrait avancer deux hypothèses : les réserves pourraient se trouver soit dans les feuilles des tiges végétatives (qui, elles, survivent à l’hiver), soit dans ces organes qui ont l’apparence de petits rhizomes.

## 5.2 Sur la relation entre *Arabis alpina* et son environnement

La variabilité environnementale mesurée sur le terrain reflète bien le contraste entre les milieux dans lesquels on retrouve *A. alpina*, comme suggéré dans l’introduction de ce chapitre.

Les études des relations entre les taux vitaux et l’environnement ont montré que ceux-ci (à l’exception, peut-être de l’effort reproducteur) étaient influencés par les variations spatiales et temporelles de l’environnement. En ce qui concerne les variations spatiales, nous avons pu montrer que la température moyenne “tardive” (en fait, très fortement corrélée à la température moyenne totale,  $r = 0.98$ ) possède une forte influence sur la croissance et la survie des plantes. Il semble en effet y avoir un compromis entre la croissance et la survie avec une forte croissance, mais faible survie dans les milieux chauds et réciproquement. En outre, la population dans le milieu le plus froid (GAL) possède la plus faible croissance, la plus forte survie, mais aussi la plus faible reproduction.

Ce qui est intéressant dans ces résultats, c’est qu’ils correspondent aux patrons interspécifiques observés avec l’augmentation de l’altitude et donc la diminution de la température moyenne (KÖRNER, 2003). Ceci suggère très fortement qu’il s’agit là de variations phénotypiques adaptatives en fonction de l’environnement, donc une possible adaptation locale.

Bien sûr, il est, en l’état des choses impossible de dire avec certitude qu’il s’agit d’adaptation locale, puisque ce résultat pourrait être tout aussi bien expliqué par de la plasticité phénotypique. Pour pouvoir réellement étudier ce phénomène et ses bases génétiques, il est nécessaire de recourir à une expérience de jardin commun.

# CHAPITRE 5

## Étude de l'adaptation locale chez *Arabis alpina* à l'aide d'une expérience de jardin commun

---

### 1 Plasticité phénotypique ou adaptation locale ?

#### 1.1 Le problème du phénotype *in situ*

Dans le CHAPITRE 4, nous avons vu qu'il existe un lien entre les taux vitaux (spécifiquement, la croissance et la survie) et l'environnement des populations étudiées. Notamment, les plantes possèdent une plus forte croissance et une moins bonne survie dans les milieux plus chauds. Ces relations peuvent être reliées à des prédictions d'écologie alpine (KÖRNER, 2003) concernant les patrons inter-spécifiques variant avec l'altitude (et donc, généralement, la température moyenne).

L'ensemble de ces considérations tend à faire penser que la variation phénotypique observée (au moins sur la croissance et la survie) en milieu naturel est d'origine adaptative. Étant donné que nous n'avons pas pris en compte l'effet de la dérive entre populations, nous ne pouvons cependant pas totalement exclure la possibilité d'une origine neutre, par dérive.

Or, tout porte à croire que la dérive au sein de nos populations est très importante, étant donné les petites tailles de population, le régime de reproduction (autofécondation partielle) et la forte reproduction différentielle entre individus. En outre, vu la distance entre nos populations et les caractéristiques du flux de gènes (BUEHLER et al., 2012), cette dérive doit très vraisemblablement entraîner une très forte différenciation génétique neutre entre populations. Enfin, nous nous situons dans une région où de forts taux de différenciation entre populations ont été enregistrés (BUEHLER et al., 2013).

Bien sûr, avec des données génétiques, il serait possible de prendre en compte cette dérive neutre entre populations, et de montrer, le cas échéant, que les variations observées ne sont pas d'origine neutre. Pourtant, cela ne consisterait toujours pas une preuve du scénario d'adaptation locale.

Comme expliqué à plusieurs reprises, deux mécanismes, en effet, sont possibles pour expliquer une variation non-neutre du phénotype avec l'environnement : la plasticité phénotypique et l'adaptation locale. Il est important ici de noter que les deux mécanismes sont potentiellement adaptatifs (au sens donné dans l'Introduction générale). C'est bien évidemment le cas de l'adaptation locale. C'est aussi le cas de la plasticité phénotypique en ceci qu'elle constitue en elle-même un trait phénotypique, au moins partiellement génétiquement déterminé et très vraisemblablement sélectionné parce qu'il augmente la valeur sélective des individus. Pour autant, la démonstration du côté adaptatif de la plasticité phénotypique est moins directe que la détection d'adaptation locale. Or, la plasticité

phénotypique n'est pas nécessairement adaptative puisqu'elle est issue d'une histoire sélective particulière qui peut entraîner des réponses maladaptatives, spécifiquement dans un environnement inédit (GHALAMBOR et al., 2007).

L'étude adaptative de ce mécanisme demande donc de travailler dans des conditions environnementales définies avec une précaution particulière, et de pouvoir relier les réponses phénotypiques observées à la valeur sélective des individus *in situ* (pour en démontrer le caractère adaptatif). Le tout, bien sûr, en prenant en compte la contribution de l'adaptation locale à la variabilité des traits phénotypiques étudiés.

Pour toutes ces raisons, nous tenterons de démontrer l'existence d'une adaptation locale dans ce chapitre, mais nous nous contenterons de simplement quantifier l'existence d'une plasticité phénotypique sans en discuter la nature adaptative. Puisque ces deux phénomènes peuvent provoquer des réponses similaires en conditions naturelles, il est d'abord important de réussir à les séparer expérimentalement.

## 1.2 Faire la part des choses : le jardin commun

La raison de la confusion entre l'effet de la plasticité phénotypique et celui de l'adaptation locale est que les deux mécanismes prévoient un lien statistique entre le phénotype et l'environnement : le premier, parce que des mécanismes développementaux, de régulation ou comportementaux provoquent une "adéquation"<sup>1</sup> immédiate (à l'échelle micro-évolutive) du phénotype des individus à leur environnement ; le second, parce que la sélection passée, dans un environnement plus ou moins constant a provoqué cette même "adéquation" en modifiant les fréquences alléliques des gènes impliqués dans l'expression du trait phénotypique. En bref, le premier est un phénomène rapide, le second est un phénomène lent, bien que les deux mènent potentiellement au même type de réponse.

Ainsi, la meilleure façon de séparer les deux phénomènes est-elle de jouer sur cette différence au niveau des "temps de réponse". Pour ce faire, il faut donc changer l'environnement des individus, ou pour être plus précis, rendre cet environnement comparable entre les individus de différentes populations pour "lisser" l'effet de la plasticité phénotypique. En effet, en plaçant des individus dans un même environnement<sup>2</sup>, la réponse "immédiate" de la plasticité phénotypique sera similaire entre tous ces individus.

Les effets de l'adaptation locale, étant eux d'origine purement génétique, seront toujours différents selon les provenances respectives des individus. C'est le principe de l'expérience de *jardin commun* : en plaçant tous les individus dans le même environnement, il est possible de considérer toute différence entre populations comme étant d'origine génétique. Cependant, ces effets d'origine génétique seront aussi influencés par la dérive, et donc par la différenciation génétique neutre entre populations. Il est donc important de rejeter la possibilité d'une différenciation neutre avant de s'avancer sur une hypothèse d'adaptation locale.

---

<sup>1</sup>Au sens donné dans l'Introduction générale.

<sup>2</sup>Le tout n'étant pas simplement de prendre des individus et de les déplacer, mais de faire en sorte que tous les individus n'aient connu que le même environnement *durant toute leur vie*. Ceci afin d'éviter des effets développementaux précoces.

## 2 Méthodes statistiques et moléculaires utilisées

Comme mentionné dans le CHAPITRE 3, il existe actuellement de nouvelles techniques statistiques et moléculaires qui peuvent être utilisées dans le cadre d'une expérience de jardin commun, et dont nous allons avoir l'utilité dans ce chapitre.

### 2.1 Méthodes statistiques

Comme présenté dans l'ARTICLE 3, nous allons faire usage de la méthode d'OVASKAINEN et al. (2011) pour prendre en compte la différenciation phénotypique attendue par dérive neutre. Pour autant, la statistique de test  $S$  développée par ces auteurs entraîne un test légèrement trop prudent (OVASKAINEN et al., 2011). Or, dans notre cas, le faible nombre de populations (six populations seulement), ainsi que la forte différenciation neutre attendue entre populations font que, même de manière générale, un test par comparaison  $Q_{ST}$ - $F_{ST}$  pourrait très difficilement s'avérer positif.

Pour palier à ce problème, et comme expliqué dans l'article ci-dessous, nous avons décidé d'opter pour une méthode plus puissante de test d'association entre l'environnement de la population d'origine et le phénotype mesuré en jardin commun. Pour prendre en compte l'effet de la dérive, le modèle d'OVASKAINEN et al. (2011) sera utilisé comme modèle nul. Ce protocole de test, bien que plus spécifique que les tests mentionnés ci-dessus (puisque'il ne teste qu'une variable environnementale), nous permet, pour peu que l'on ait accès aux variables pertinentes, d'obtenir des résultats significatifs dans notre situation.

Or, grâce aux analyses sur les phénotypes *in situ*, nous savons que certaines des variables environnementales utilisées dans le CHAPITRE 4, bien qu'elles ne soient pas exhaustives, sont du moins pertinentes pour certains des traits que nous allons étudier.

### 2.2 Méthodes moléculaires

Le but de l'analyse suivante est non seulement de démontrer (si possible) l'existence d'une adaptation locale, mais aussi de permettre l'étude, le cas échéant, des bases génétiques de cette adaptation. Pour cela, nous utiliserons les méthodes de scan génomique de sélection abordées dans les CHAPITRES 1, 2 et 3, ainsi qu'une étude d'association comme mentionné dans l'Introduction générale et le CHAPITRE 3. Cependant, pour pouvoir utiliser de telles méthodes, il est préférable d'avoir accès à un panel de marqueurs suffisamment dense dans le génome (voir ARTICLE 3).

Pour obtenir un panel de marqueurs suffisamment dense, nous avons décidé de recourir au *double digest Random Amplified DNA sequencing* (ddRADseq, PETERSON et al., 2012) dont le protocole est expliqué dans l'Introduction générale. L'avantage de ce protocole, c'est qu'il est possible d'obtenir un très grand nombre de marqueurs co-dominants avec un coût relativement réduit. En outre, puisque nous avons accès à un génome de référence pour *A. alpina* (WILLING et al., 2015), nous pouvons rapporter nos marqueurs à ce génome et utiliser le contexte génomique (en particulier, si la région est dans un gène ou non). Le protocole ddRADseq a dû être légèrement modifié pour tenir compte de problèmes d'inhibiteurs des polymérases, assez spécifique aux plantes.

Bien sûr, la densité des marqueurs en utilisant le ddRADseq est toute relative. Elle est très supérieure à ce que l'on peut obtenir grâce à des AFLP, par exemple, mais aussi très inférieure à la densité offerte par du séquençage complet de génome.



### 3 Présentation de l'ARTICLE 5

Dans l'article suivant, nous allons présenter les résultats de l'expérience de jardin commun que nous avons menée sur nos six populations d'*A. alpina*. Puisque nous avons mis en place l'expérience dans deux jardins différents, il sera possible de quantifier la plasticité phénotypique présente sur les traits phénotypiques mesurés. En outre, en utilisant les sondes *in situ* dans nos six populations, il sera possible de tester une relation entre l'environnement dans la population d'origine et les traits phénotypiques mesurés dans nos jardins. Puisque l'influence de la plasticité phénotypique aura été lissée entre populations grâce au protocole expérimental, nous pourrions mesurer la différenciation phénotypique d'origine génétique entre populations.

Enfin, en utilisant les données moléculaires, il sera possible de tester s'il existe une relation entre l'environnement de la population d'origine et le trait phénotypique mesuré, tout en prenant en compte la possibilité d'une origine neutre de ces différences.

Dans un second temps, nous essaierons de détecter des gènes candidats (potentiellement sous sélection et associés à un trait phénotypique) qui pourraient être impliqués dans ces phénomènes d'adaptation locale.

#### 3.1 Contributions

L'expérience a été conçue par Irène Till-Bottraud et moi-même. La mise en place de l'expérience, le phénotypage et l'analyse des données ont été menés par moi-même, avec l'aide d'Oscar Gaggiotti et Irène Till-Bottraud. Le génotypage, ainsi qu'une partie de l'analyse bioinformatique, ont été conduits par Médéric Mouterde. L'article a été rédigé par moi-même, avec l'aide d'Oscar Gaggiotti et Irène Till-Bottraud.

#### 3.2 Remarques

Cet article est une ébauche, en préparation, possiblement pour *Molecular Ecology*. Certaines analyses (études d'association, prise en compte de l'incertitude sur le modèle neutre) ne sont pas tout à fait abouties. Les résultats fondamentaux et leur discussion sont cependant fixés dans l'ensemble.

Notons que cette expérience, dans cette forme finale, est le résultat de nombreux rebondissements, le plan original étant de génotyper des individus du suivi démographique. Malheureusement, la quantité et la qualité des échantillons récoltés au début du suivi démographique ont fait qu'après presque un an de travaux, ce génotypage a dû être abandonné. Entre temps, suite à une proposition d'Irène, j'avais pris en charge cette expérience de jardin commun. C'est donc tout naturellement que le génotypage s'est reporté sur les individus de ce jardin commun offrant des perspectives d'analyses très intéressantes.

#### 3.3 Résumé

Quand des populations, en contact à travers un flux de gène modéré, subissent des conditions environnementales différentes, cela peut mener à un phénomène évolutif nommé adaptation locale (c'est à dire une adaptation aux conditions environnementales locales). Les écosystèmes alpins sont enclins à de tel phénomènes, puisqu'ils sont caractérisés par des conditions environnementales très tranchées sur de courtes distances géographiques. Parmi les espèces de plantes alpines, *Arabis alpina* est caractérisée par une forte gamme altitudinale, allant de 800m jusqu'à 3000m d'altitude.

Grâce à une expérience de jardin commun, répliquée dans deux sites différents, et combinée à une collecte de données environnementales *in situ* et d'un génotypage haut-débit, nous montrons que des traits phénotypiques clés de cette plante (longueur totale de fruits, hauteur, croissance) présentent un signal significatif d'adaptation locale lié à la température moyenne des populations d'origine. Cette significativité est testée contre une hypothèse de dérive neutre, en utilisant l'information issue des marqueurs moléculaires et un modèle rigoureux de différenciation phénotypique sous une hypothèse de neutralité.

En utilisant deux sites expérimentaux, nous montrons que ces traits adaptatifs sont aussi caractérisés par la présence d'une plasticité phénotypique considérable.

Grâce à la densité de marqueurs moléculaires offerte par le génotypage haut-débit, nous utilisons des méthodes de scan génomique pour détecter de la sélection et d'études d'association avec nos traits phénotypiques. Ceci nous permet de proposer une liste de 13 gènes chez *A. alpina* qui montrent à la fois un signal de sélection et sont associés à un trait phénotypique adaptatif.

Globalement, nos résultats prouvent l'existence d'une sélection le long d'un gradient altitudinal pour des plantes plus petites et plus compactes, qui croissent plus lentement et produisent moins de fruits. Ces résultats à l'échelle d'une espèce sont très similaires aux patrons interspécifiques attendus en écologie alpine.

**Journal:** En préparation pour *Molecular Ecology*

Les *Supplementary Information* de cet article sont disponibles dans l'Annexe S5

## ARTICLE 5

### Local adaptation allows the alpine plant *Arabis alpina* to inhabit within a wide altitudinal range.

Pierre de Villemereuil, Médéric Mouterde, Oscar E. Gaggiotti & Irène Till-Bottraud

#### Abstract

When populations in contact through a mild gene flow experience contrasted environment, local adaptation, i.e. adaptation to the local environmental conditions, is a possible evolutionary output. Alpine ecosystems are likely to promote such a phenomenon, as they are characterised by highly contrasted environmental conditions within a very short distances. Among alpine plant species, *Arabis alpina* is characterised by a very wide altitudinal amplitude, ranging from 800m to 3000m of elevation. Using a replicated common garden experiment, combined with *in situ* environmental data collection and high-throughput genotyping, we show that key phenotypic traits of the plant such as total fruit length, growth and height display significant signs of local adaptation along the gradient of average temperature of the original population. This significance is tested against neutral drift, using molecular markers information and a rigorous model of phenotypic differentiation under a neutrality hypothesis. Using the two different experimental sites, we show that these adaptive traits are also characterised by a considerable phenotypic plasticity. We used high-throughput genotyping to obtain dense markers in *A. alpina* genome and performed genome scans to detect signal of selection and association studies with phenotypic traits. A short list of 13 genes in *A. alpina* showed a significant signal of selection and were associated with an adaptive phenotypic trait. Overall, our results show selection, along the altitudinal gradient, for smaller, more compact plants, that are growing slower and producing less fruits. These intraspecific results mirror interspecific patterns that are well described in alpine ecology.

#### Introduction

Local adaptation arises when populations in contact through mild gene flow experience contrasted environmental conditions: if the environment imposes natural selection and if some adaptive potential exists in the populations, evolution is expected to favour traits values that increase the fitness of individuals in their local environment. As a result, individuals have a better fitness in their local environment than individuals from other populations (KAWECKI & EBERT, 2004). Conceptually, local adaptation is a natural experiment of micro-evolution, which informs us on how species evolve in response to their environment. It has also important implications for conservation and response to global change (AITKEN et al., 2008 ; ALBERTO et al., 2013).

The study of local adaptation is however made complex by the confounding influence of genetic drift, population structure and phenotypic plasticity (KAWECKI & EBERT, 2004). This latter phenomenon is particularly limiting, because the plastic response of the organisms to the environmental conditions hampers the comparison of phenotypes in different environments and hence the study of the influence local adaptation might have on them. Common garden experiments (KAWECKI & EBERT, 2004 ; SAVOLAINEN et al., 2013), in which offspring of individuals from the studied populations are grown in similar environmental conditions allow to overcome this limitation. In this experimental setting, the phenotypes become comparable between populations and differences in mean population phenotypes are considered of genetic origin.

Yet genetic differences might still have different possible causes, the two main ones being genetic drift and local adaptation. To distinguish between the two, it is necessary to include molecular information in the analysis to distinguish between these two evolutionary phenomena (SAVOLAINEN

et al., 2013). Only when the neutral hypothesis of genetic drift is ruled out can we suppose that local adaptation might be responsible for the observed phenotypic differences between populations.

In the context of common garden experiments, this is most often done by using  $Q_{ST}$ - $F_{ST}$  comparisons (SPITZE, 1993 ; LEINONEN et al., 2013). The  $Q_{ST}$  is defined as the ratio of between-population genetic additive variance  $V_B$  over the total genetic additive variance (i.e. including the within-population additive variance  $V_W$ ) :

$$Q_{ST} = \frac{V_B}{V_B + 2V_W}. \quad (A5.1)$$

In the context of neutral evolution by drift only, the value of the  $Q_{ST}$  is expected to be equal to the value of the molecular  $F_{ST}$  (SPITZE, 1993).

Common garden experiments (and the related protocol of reciprocal transplants) and  $Q_{ST}$ - $F_{ST}$  comparisons have become one of the most useful tool to study local adaptation (SAVOLAINEN et al., 2013). It was successfully used to detect local adaptation in a wide range of plant species ranging from annual model plants such *Arabidopsis thaliana* (STINCHCOMBE et al., 2004 ; BRACHI et al., 2013 ; KAWAKAMI et al., 2011) and non model perennial herbs (KAWAKAMI et al., 2011) to forest trees (MORGENSTERN, 2011 ; ALBERTO et al., 2011 ; ALBERTO et al., 2013). It was also used to investigate the extent of local adaptation in the invasive potential of species (KOLLMANN & BANUELOS, 2004 ; LEGER & RICE, 2007 ; COLAUTTI et al., 2010). Most of the times, local adaptation is studied in populations sampled along an environmental gradient (KAWAKAMI et al., 2011 ; ALBERTO et al., 2011).

Despite this success (LEINONEN et al., 2013), the  $Q_{ST}$ - $F_{ST}$  comparison framework has been repeatedly criticised for the lack of link between mutation rates in molecular and phenotypic data (EDELAAR et al., 2011 ; EDELAAR & BjÖRKLUND, 2011) and failures to account for statistical uncertainty (EDELAAR et al., 2011 ; OVASKAINEN et al., 2011). Another drawback with this framework is that it conveys little information regarding the selective origin of the phenotypic differentiation, i.e. the environmental variable driving local adaptation on the studied phenotypic trait. Fortunately, a more rigorous model of neutral evolution expected at the phenotypic level was recently developed by OVASKAINEN et al. (2011). In this framework, molecular data are used to compute a population-level variance-covariance matrix of neutral phenotypic evolution (KARHUNEN & OVASKAINEN, 2012), which is then used to test whether phenotypic differentiation between populations is greater or lower than expected under drift (KARHUNEN et al., 2013). This framework still does not make use of environmental information (but see KARHUNEN et al., 2014), but does provide a neutral model of neutral evolution against which such relationships between selective environment and phenotypic trait variations measured in common garden could be tested.

Finally, the detection of adaptive traits is often not the sole aim of evolutionary biology studies, another aim being unravelling the genetic architecture of those traits (PARDO-DIAZ et al., 2015). Quite naturally, the study of genetic architecture is influenced by the same confounding effects as the study of the phenotypic traits themselves : especially, phenotypic plasticity and population structure can generate a large amount of false positives. For non model species, the best method to avoid issues due to phenotypic plasticity is to compare phenotypes in a common environment. As a consequence, it is useful to combine common garden experiments with genome scans and genome-wide association methods to detect the loci potentially responsible for local adaptation (de VILLEMEREUIL et al., 2015).

Because they are characterised by strong environmental contrasts over small geographic distances (KÖRNER, 2003), alpine ecosystems and alpine species, especially sessile organisms such as plants, are likely to be submitted to local adaptation. Among the plants characteristic of the European Alps, *Arabis alpina* is remarkable for its wide altitudinal amplitude ranging from 800m to 3000m of elevation. This range extends from the middle of the montane zone to the top of the alpine zone : these zones, and the subalpine zone in between, represent very different environments with contrasting temperature, humidity, radiation, partial CO<sub>2</sub> pressure and snow cover conditions. As a consequence, elevation forms a stress gradient along which environmental conditions harshen dramatically (KÖRNER,

2003), although many other factors modulate this effect (KÖRNER, 2007). From 800m to 2000m, the major environmental transition is the transition from deciduous to coniferous forest. Above 2000m, elevation has an even stronger impact on the specific and functional diversity of plant communities, with a transition from interspecific competition to facilitation (CHOLER et al., 2001), as well as on plant physiology and morphology, with, for example, the extreme case of cushion plants in the alpine and nival zones (BOUCHER et al., 2012 ; ROQUET et al., 2013).

Hence very different species are typical of these various environmental conditions. This begs the question of how the species *Arabis alpina* is able to grow, survive and reproduce in such a wide altitudinal amplitude. Two main characteristics of the species might explain this: either this species is very plastic and is able to change its phenotype according to some environmental cue, or there is strong local adaptation. Of course, these hypotheses are not mutually exclusive and both might contribute to the wide altitudinal range of this species.

To study the scenario of local adaptation in the case of *Arabis alpina*, we conducted a common garden experiment using six populations covering the natural altitudinal range (900m – 3000m) of the species in the French Alps. To characterise the process of local adaptation in this species, we combined phenotypic, genotypic and *in situ* environmental data in statistical models accounting for the effects of population structure and genetic. In order to assess the extent of phenotypic plasticity and genotype-by-environment effects, we conducted the experiment in two different environments that represent the montane and subalpine zones. Using dense genotypic data, we conducted both genome scan analyses to detect selection and association studies to explore the genetic architecture of our measured phenotypes. Combining both of these genomic analyses with information available on the closely related plant *Arabidopsis thaliana* allowed us to detect candidate genes potentially involved in local adaptation.

## Material & Methods

### Species and population

#### ► Species

*Arabis alpina* is a short-lived perennial arctic-alpine plant growing on open, unstable grounds. It is characterised by a wide altitudinal amplitude (800m – 3000m of elevation above sea level in the Alps), entomogamous pollination and autochorous seed dispersal. In the French parts of the Alps (where this study was conducted), it is also characterised by partial selfing (ANSELL et al., 2008).

#### ► Populations

The six studied populations covered much of the natural altitudinal range of the species (900m to 3000m) and were localised in two different massifs: three populations were from the Vercors massif (900m to 2000m) and three populations were located near the Lautaret pass (2000m to 3000m). As explained in ARTICLE 4, an individual-based demographic survey was performed on these populations from 2008 to 2014. Population characteristics are summarised in TAB. A5.1<sup>1</sup>.

#### ► Environmental data

Environmental data was collected using temperature and humidity sensors (iButton® Maxim Integrated™) were collected every three hours in each population from 2008 to 2015. The sensors were placed above the ground (around the canopy height of *Arabis alpina*) and protected from the sun. Using these semi-continuous data, we were able to define, for each population and each year, the start of the growing season as the date when degree-days (above 0°C) started to accumulate and the end of the growing season as the date when degree-days stopped accumulating. The start and end of the seasons were determined either by a trend in daily mean temperature or by snow cover. Using

---

<sup>1</sup>See also TAB. A4.1.

**TAB. A5.1:** Summary of the characteristics of the 6 studied populations. *Avg. Temp.*: average daily mean temperature; *Temp. Range*: average daily temperature range; *Avg. Hum.*: average daily mean humidity index; *Season Length*: average length of the reproductive season (in days); *Freezing days*: average number of days below 0°C

ID code	Massif	Elevation	Aspect	Avg. Temp.	Temp. Range	Avg. Hum.	Season Length	Freezing days
BRU	Vercors	930m	South	9.1°C	7.3°C	102.6%	251.6	18.6
CHA	Vercors	1480m	North	10°C	11.2°C	87%	238.3	31.8
VIL	Vercors	1980m	South	8.2°C	11.7°C	96.6%	217.5	47
LAU	Lautaret	2090m	North	8.2°C	6.1°C	100%	136.2	6.1
GAL	Lautaret	2590m	North	6.2°C	6.6°C	92.4%	120.7	16.3
PIC	Lautaret	2930m	South	6.8°C	15.2°C	86.5%	158.2	62.1

the GPS location points, we estimated the elevation and aspect (North- or South-facing side) of each population.

From the semi-continuous series, we summarised the environmental conditions in each population by five variables (see TAB. A5.1):

**Average Temperature** Average of the daily mean temperature during the growing season.

**Temperature Range** Average of the daily temperature range during the growing season.

**Average Humidity** Average of the daily mean humidity during the growing season.

**Season Length** Number of days between the start and the end of the growing season.

**Freezing days** Number of days for which a negative temperature was recorded during the growing season.

In order to describe the spatio-temporal environmental variations, we performed a Principal Component Analysis (PCA) and a discriminant analysis on the yearly fluctuations of these variables using either the population or the year as a discriminant factor. These analyses were conducted using the *ade4* R package (DRAY & DUFOUR, 2007).

## Common garden experiment

### ► Plant collection

We collected maternal progenies from the six natural populations. To do so, during summer 2012, we put bags around maturing fruits of plants located in these populations. The seeds were collected when fruit matured. During the spring of 2013, we germinated seeds in the lab. We then planted the seedlings in the two different gardens.

### ► Experimental setting

One garden was located near the Vercors Regional Natural Park House and the second garden was located at the Joseph Fourier Alpine Botanic Garden at the Lautaret pass. The experimental gardens were composed of three and five blocks of 100 (5×20) plants respectively. To avoid any kind of border effect, we planted 54 *Arabis alpina* individuals around the 100 monitored individuals within each block.

### ► Phenotypic traits

We phenotyped the common garden individuals for different traits at the height of the reproductive season. We measured the *total fruit length* as the total number of fruits multiplied by the average fruit length (measured over 5 fruits). In the Vercors garden, the plants were too big to record the actual number of fruits, so we recorded the number of reproductive stems and estimated the average number of fruits *per stem* on 10 stems.

We recorded different morphological measurements such as *basal height* (height to the highest leaf of the “rosette” part of the plant), *vegetative height* (height of the highest leaf) and *reproductive height* (height of the highest flower corolla). The individual *surface area* was estimated using the measure of two orthogonal diameters and approximation of the area as an ellipse. We estimated the *growth rate* as the ratio between the individual area in 2015 and 2014 (Lautaret garden individuals only). We categorised the vegetative habit of the plants of the individuals into four different *morphotypes*: “Sparse rosettes”, “Intermediate”, “Numerous Rosettes”, “Compact rosettes” (see Section S5.1 in Supplementary Information).

We computed *flowering time* as the number of weeks between the end of snow cover in the garden and the first observation of an open flower. Because the season started particularly early in 2015, many plants were already flowering at the first visit.

All of these phenotypic traits, except for the morphotype, were recorded during 2014 in Vercors and during 2014 and 2015 in Lautaret. At the end of 2014 for Vercors and 2015 for Lautaret, we pulled the surviving plants out to weight their *aerial biomass* and dried them to measure their *dry biomass*.

Finally, we recorded mortality at the beginning and the end of each summer. *Survival* was computed as the number of surviving years. We considered that plants that were pulled out for biomass measurement would have survived until the following year.

During the summer of 2014, in Lautaret, a strike of white rust (*Albugo candida*, BAKA, 2008) severely infected the individuals with dramatic consequences on their growth, reproduction and survival. We recorded plants displaying symptoms of sickness during this summer. The white rust targeted more specifically the local populations (i.e. the three populations from the Lautaret massif,  $F_{1,442} = 23.8, p < 2.10^{-6}$ ).

## Genotyping

In July 2014, we collected leaf samples on the 204 surviving individuals in Lautaret. We extracted DNA from these samples using the Qiagen DNeasy Plant Mini kit. We then used a double digest RAD sequencing protocol (PETERSON et al., 2012) using the *ecoRI* and *mspI* restriction enzymes. Fragments between 150bp and 600bp were pair-end sequenced on 125bp using an Illumina HiSeq sequencer.

We mapped reads on the *Arabis alpina* reference genome (NCBI, GenBank, Accession JNGA00000000, Version 1, <http://www.ncbi.nlm.nih.gov/nucore/JNGA00000000.1>) and further processed them using the Stacks pipeline (CATCHEN et al., 2011 ; CATCHEN et al., 2013). After filtering for allelic coverage above 2 for all individuals, missing rate below 30% and minor allelic frequencies above 1%, we performed SNP calling using Stacks. To avoid issues due to strong linkage between SNP on the same reads, we used multiallelic sequence polymorphism which we hereafter refer to as “RAD haplotypes”, or for the sake of simplicity, loci.

Finally, we removed some genotyped individuals due to strong rates of missing values, missassignment and aberrant clustering behaviour. In the end, we retained 3,528 loci on 152 individuals, with on average 25.3 individuals *per* population and 3.1 individuals *per* family.

## Statistical analyses of the phenotypic traits

### ► Population and family structures

We checked population structure using the clustering algorithm sNMF (FRICHOT et al., 2014). To infer the ancestral genetic additive variance of the traits, we estimated population-level coancestry (OVASKAINEN et al., 2011) using the Admixture F-model implemented in the R package RAFM (KARHUNEN & OVASKAINEN, 2012). To be precise, we ran RAFM separately on each massif dataset and combined the matrix estimates for both massifs into a bigger matrix assuming a coancestry of 0 between massifs. We did so because of the particular hierarchical structure of our data, which the F-model has difficulty to account for (EXCOFFIER et al., 2009).

We inferred the family structure from molecular data, separately for each population and accounting for the partial selfing and hermaphroditism in *Arabis alpina*. This was done using the COLONY software (O. R. JONES & J. WANG, 2010).

Using both of these inferences and following OVASKAINEN et al. (2011), we constructed a matrix of population-level relatedness (hereafter called  $\mathbf{B}$ ) and a matrix of within-population relatedness (hereafter called  $\mathbf{W}$ ). The overall relatedness between individuals since the assumed ancestral population (see OVASKAINEN et al., 2011) is thus the sum of both matrices:  $\mathbf{A} = \mathbf{B} + \mathbf{W}$ .

### ► Analysis of the phenotypic traits

Since only a subset of the individuals was genotyped, focusing only on these individuals might result in a great loss of power. Instead, we performed two kinds of analyses. The first, hereafter referred to as “Genotyped analysis” includes only the genotyped individuals, hence only individuals from the Lautaret garden.

The estimated random effect variances in this kind of analysis includes the between-population genetic variance  $V_B$  (inferred using the covariance matrix  $\mathbf{B}$ ), the within-population genetic variance  $V_W$  (inferred using the covariance matrix  $\mathbf{W}$ ), the maternal effect variance  $V_M$  (inferred using maternal identity) and the block effect variance  $V_{\text{block}}$ .

The second kind of analysis, hereafter referred to as “Global analysis” includes all individuals, from both gardens. The estimated random effect variances in this kind analysis includes the between-population genetic variance  $V_B$  again, the family effect variance  $V_F$  (which, in absence of genotypic information, includes both  $V_W$  and  $V_M$  above with unknown weighting), a garden-by-population (a.k.a. genotype-by-environment) interaction variance  $V_{G \times E}$  and again a block effect variance  $V_{\text{block}}$ . Note that, in the case of growth rate, the genotype-by-environment had no meaning since we only had data for the Lautaret garden individuals, hence it was removed.

For both kinds of analyses, we also tested three potentially confounding fixed effects. The “garden” effect tested whether the phenotypes were different in the two gardens, the “year” effect tested whether phenotypes differed between measurements in 2014 and in 2015 in the Lautaret garden and the “white rust” effect tested the effect of the white rust on the phenotype in Lautaret.

These analyses were conducted using the MCMCg1mm R package (HADFIELD, 2010). Point estimates given in this paper are the mean for the fixed effects and the median for variances and variance ratios. For all models, convergence was checked graphically and using the Heidelberger and Welch’s test (HEIDELBERGER & WELCH, 1981), and length of runs and thinning were set up so as to obtain an effective size above 1000 for all parameters of the model.

### ► Testing for local adaptation

In order to test for local adaptation, we tested whether each one of the five *in situ* environmental variable had an effect on the phenotype measured in the common garden(s). Because the matrix  $\mathbf{B}$  captures the expected effect of genetic drift, any significant effect of the environment could be considered as significant against the hypothesis of drift.

We decided to use only relationships with the environment as a test for local adaptation as our small number of populations would make any other test framework conservative (e.g. the  $S$ -test, OVASKAINEN et al., 2011). Using the “genotyped analysis”, we computed  $Q_{ST}$  values (SPITZE, 1993), but did not test it against  $F_{ST}$  values, as this testing framework suffers from many issues (EDELAAAR et al., 2011; OVASKAINEN et al., 2011). Instead, we compared  $Q_{ST}$  values with or without the effect of significant environmental effect to quantify how well these effects account for between-population genetic variance.



## Genome scans and association studies

### ► Genome scans to detect selection

To detect loci that show a pattern of local adaptation, we used two different methods. The first method was BayeScEnv (de VILLEMEREUIL & GAGGIOTTI, 2015), which tests for a relationship between genetic and environmental differentiation, while accounting for various confounding effects such as genetic drift and genome heterogeneity. We set the prior probability for the neutral model to 0.1 and the prior preference for “local adaptation” to 0.5. The environmental differentiation was defined as the Euclidean distance to the average value among all populations. BayeScEnv controls for the false discovery rate by yielding  $q$ -values as a test statistic (STOREY, 2003).

The second method we used was LFMM (FRICHOT et al., 2013), which tests for a linear pattern between individual genotypes and an environmental variable, while accounting for population structure using latent factors. We set the number of factors  $K$  to 6 (our number of populations and preferred number of clusters) and performed 10 runs of the algorithm. The  $z$ -values yielded by the different runs were combined using Stouffer’s method (STOUFFER et al., 1949). A genomic inflation factor (DEVLIN & ROEDER, 1999) was computed using the resulting  $z$ -values, which were then transformed into corrected  $p$ -values. To control for false discovery rate, the distribution of  $p$ -values were further transformed into  $q$ -values using the STOREY & TIBSHIRANI (2003) algorithm.

We ran each of the above methods seven times, using each of five environmental variables described above, plus population elevation and aspect. Given the overall small power of this analysis, we considered  $q$ -values as significant with a threshold of 0.1.

### ► Association studies

We performed an association study to link our genotypic markers to the phenotypes we measured. Because relatedness-based methods such as EMMA (KANG et al., 2008) are ill-designed for the study of strongly differentiated populations, we decided to resort to LFMM (FRICHOT et al., 2013), which is better suited to account for population structure with strong drift-induced between-population structure. As a consequence, our test framework does not use the genotypes to predict the phenotypic traits, but the reverse. To be used as predictors, traits were centred and scaled. Ordered categorical and binary phenotypes such as morphotype and survival were transformed into integer values before centring and scaling. The settings of the analysis and post-analysis were identical to the above, we used e.g.  $K=6$  and 10 runs, controlled for genomic inflation and transformed the  $p$ -values into  $q$ -values. Again, we considered  $q$ -values as significant with a threshold of 0.1.

Because it was not possible to use phenotypes from both years, 2014 and 2015, we used year 2014 in which more measurements were available.

### ► Finding candidates genes

Loci that were found associated to one of the measured phenotypes and with a significant signal of selection were extracted from both of the analyses above. For such loci lying within an annotated genic region, we performed Blast queries (ALTSCHUL et al., 1997) against *Arabidopsis thaliana* protein database (LAMESCH et al., 2012). We considered only significant hits as homology when the maximum “bit score” was above 200 and the percentage of identity above 60%. As no guarantee of functional homology comes with the molecular homology, this was solely done to gain insight into possible candidates genes for local adaptation, and not as a functional gene ontology validation.

## Results

### Analysis of environmental variables

The PCA on the environmental variables separated the temperature amplitude variables (temperature amplitude and freezing days) on the first axis (40.1% of explained variance), the temperature trend variables (average temperature and length of season) on the second axis (31.3% of explained

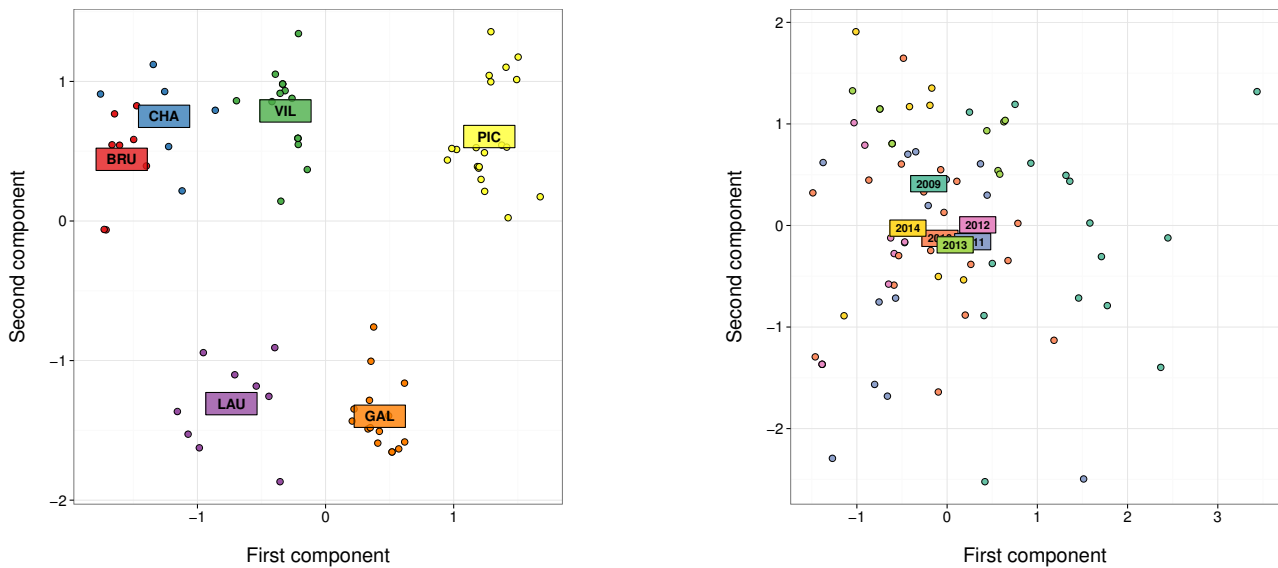


FIG. A5.1: Projections of the environmental conditions after discriminant analysis according to the local population (left) or the year of measurement (right) on the two first axes. Each point corresponds to one year in one quadrat within the population.

variance) and the average humidity on the third axis (19.7% of explained variance). The first two axes could be respectively related to aspect (correlation  $\rho = 0.65$ ,  $p = 3.13 \cdot 10^{-10}$ ) and elevation (correlation  $\rho = 0.89$ ,  $p < 2.10 \cdot 10^{-16}$ ), though the first axis was also related to elevation (correlation  $\rho = 0.36$ ,  $p = 0.0019$ ).

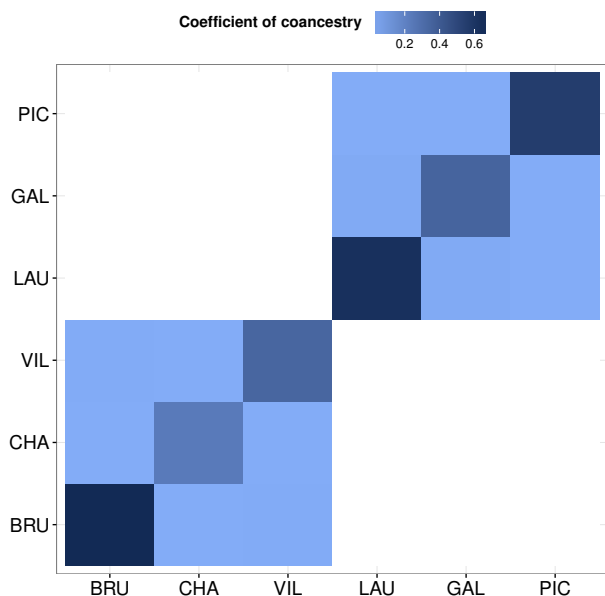
The discriminant analyses show that environmental conditions varied widely across populations (FIG. A5.1, left), but much less so across years (FIG. A5.1, right). In both cases though, randomisation Monte Carlo tests were significant ( $p < 0.001$  for both analyses, with 1000 randomisations), indicating a non-random clustering according to both populations and years.

The discriminant analysis on the populations (FIG. A5.1, left) shows a greater environmental differentiation between the populations of high elevation near the Lautaret pass (LAU, GAL, PIC), whereas populations from the Vercors massif (BRU, CHA, VIL) were more similar to each other.

These latter populations were characterised by a long growing season and high average temperatures, but narrow daily temperature range (TAB. A5.1). In relation with its high elevation and Southern aspect, PIC was strongly characterised by a wide daily temperature range and number of freezing days. LAU and GAL were characterised by both narrow daily temperature range and short growing season. BRU and LAU were furthermore characterised by high humidity. Despite not being the highest population (2500m, whereas PIC is at 3000m), GAL was the coldest population, due to its Northern aspect (TAB. A5.1).

## Neutral population structure

The cross-entropy criterion used in sNMF (FRICHOT et al., 2014) showed strong population clustering, with a most likely number of population  $K = 6$  (FIG. S5.3) and little sign of gene flow between populations (FIG. S5.4). The variance-covariance matrix  $\mathbf{B}$  estimated by RAFM (KARHUNEN & OVASKAINEN, 2012) is well-aligned with these results, with very little coancestry between populations (FIG. A5.2). The estimated drift varied also considerably between populations: BRU, LAU and PIC were the populations experiencing the strongest drift (FIG. A5.2, diagonal elements). As a result, the neutral genetic differentiation between populations was very strong: the WEIR & COCKERHAM (1984) value of the  $F_{ST}$  was estimated to be 0.60.



**FIG. A5.2:** Variance-covariance matrix **B** (population-level coefficients of coancestry) estimated by RAFM. Darker blue stands for higher values in the estimated coefficient of coancestry. White areas were not estimated, but set to zero for the analysis.

**TAB. A5.2:** Results for the **global analysis** for the morphotype (Type), total fruit length (TFL), survival, growth, flowering time (FT), basal height (H. base), vegetative height (H. veg.), reproductive height (H. repro.), area, dry biomass (Biom. dry) and the dry/fresh biomass ratio (Biom. ratio). The table shows the significant confounding and environmental effects (possibly in interaction with the garden).

Trait	Confounding effects	Environment effects	Estimate	Estimate (Vercors)	Estimate (Lautaret)
Type	—	Avg. Temp.	-1.34 (-2.11,-0.58)	—	—
TFL	Garden, White rust	Avg. Temp.	0.81 (0.02,1.64)	—	—
Survival	Garden	—	—	—	—
Growth	—	Avg. Temp.	0.34 (0.06,0.6)	—	—
FT	Year	—	—	—	—
H. base	Garden, Year	Garden : Aspect	—	-0.08 (-0.3,0.15)	-0.2 (-0.4,0.01)
H. veg.	Garden, White rust, Year	Avg. Temp.	0.24 (0.05,0.45)	—	—
H. repro.	Garden, White rust, Year	Avg. Temp.	0.22 (0.04,0.4)	—	—
Area	Garden, Year	Garden : Season length	—	0.49 (0.02,0.97)	0.07 (-0.36,0.57)
Biom. dry	Garden	—	—	—	—
Biom. ratio	Garden	—	—	—	—

## Analysis of the common garden phenotypic traits

### ► Global analysis

The analysis using all the individuals (global analysis) allowed us to find several significant environmental effects linked to the phenotypes measured in the gardens (TAB. A5.2). Of the 11 traits studied, seven (morphotype, total fruit length, growth, the three height measures and area) showed significant relationships with an environmental variable.

In most cases (morphotype, total fruit length, growth reproductive and vegetative height), the significant variable was the average temperature. The estimates for this environmental effect depict larger, less compact plants that are growing faster and reproducing more with higher temperatures. In the other two cases (basal height and area), the environmental effects (aspect and season length, respect.) were significant in interaction with the garden (i.e. only in the Lautaret and Vercors garden, respectively).

Comparing the intra-class correlation coefficients for the between-population variance  $V_B$  with or without environmental effects (FIG. A5.3, top panels) showed that, for the morphotype and the vegetative and reproductive heights, the environmental effect explained a large amount of this variance.

Apart from the morphotype and flowering time, all traits significantly differed between gardens, demonstrating the presence of phenotypic plasticity in these traits. However, for both of those traits, the absence of significant effect might be due to (or compensated by) the presence of large  $G \times E$  effects (FIG. A5.3, bottom-right panel). In all cases, the family effect contributed little to the total variance (FIG. A5.3).

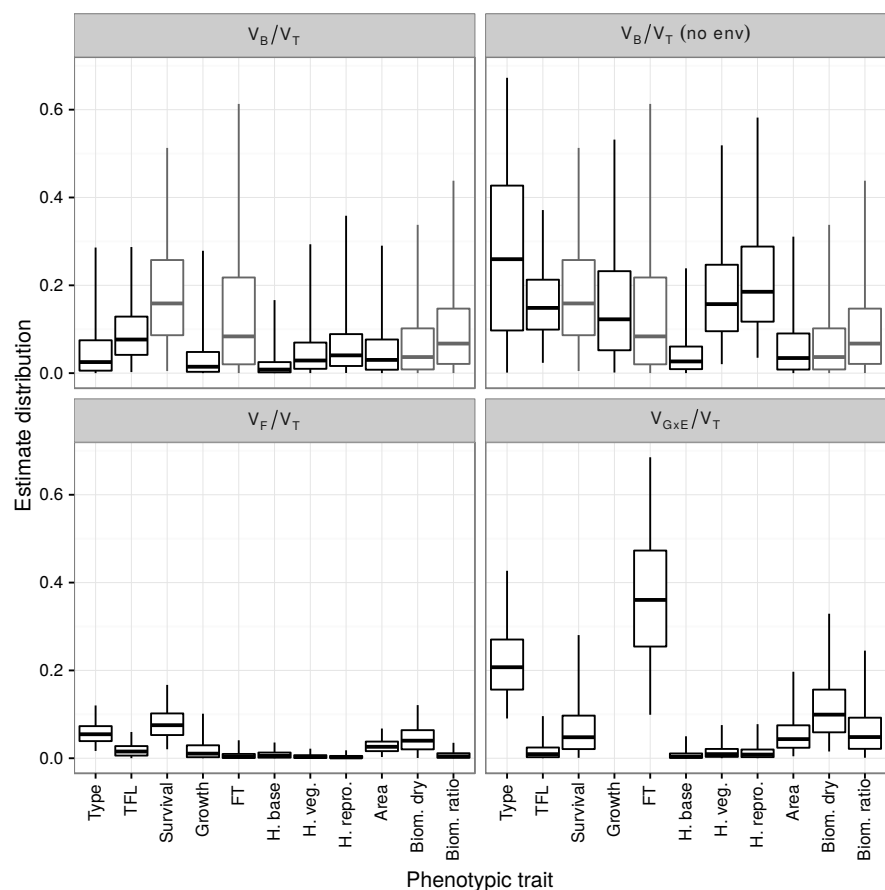


FIG. A5.3: Results from the global analysis: intraclass correlation coefficients (ICC, i.e. ratio of the effect variance to the total variance  $V_T$ ) for the 11 phenotypic traits. ICCs corresponding to the between-population genetic variance  $V_B$  is shown with (top-left panel) or without (top-right panel) the environmental effect. When no environmental effect was significant, both estimates are identical and thus displayed in grey. Bottom panels show the ICCs corresponding to the family effect variance  $V_F$  (bottom-left panel) and to the garden-by-population effect variance  $V_{G \times E}$  (bottom-right panel). The boxes and whiskers correspond to the 50% and 95% inter-percentile intervals respectively, the middle corresponds to the point estimate.

**TAB. A5.3:** Results for the **genotyped analysis** for the morphotype (Type), total fruit length (TFL), survival, growth, flowering time (FT), basal height (H. base), vegetative height (H. veg.), reproductive height (H. repro.), area, dry biomass (Biom. dry) and the dry/fresh biomass ratio (Biom. ratio). The table shows the significant confounding and environmental effects (possibly in interaction with the garden).

Trait	Confounding effects	Environment effects	Estimate
Type	—	Avg. Temp.	-1.24 (-2.22,-0.53)
TFL	White rust	Avg. Temp.	0.77 (0,1.59)
Survival	—	—	—
Growth	—	Avg. Temp.	0.43 (0.09,0.77)
FT	Year	—	—
H. base	Year	Aspect	-0.22 (-0.42,-0.03)
H. veg.	Year	Avg. Temp.	0.23 (0.05,0.45)
H. repro.	Year	—	—
Area	Year	—	—
Biom. dry	—	—	—
Biom. ratio	—	—	—

**TAB. A5.4:** Number of significant loci after the association study for the morphotype (Type), total fruit length (TFL), survival, growth, flowering time (FT), basal height (H. base), vegetative height (H. veg.), reproductive height (H. repro.), area, dry biomass (Biom. dry) and the dry/fresh biomass ratio (Biom. ratio).

	Type	TFL	Survival	Growth	FT	H. base	H. veg.	H. repro.	Area	Biom. dry	Biom. ratio
Nb. locus	3	13	18	587	13	39	17	46	27	3	10

### ► Genotyped analysis

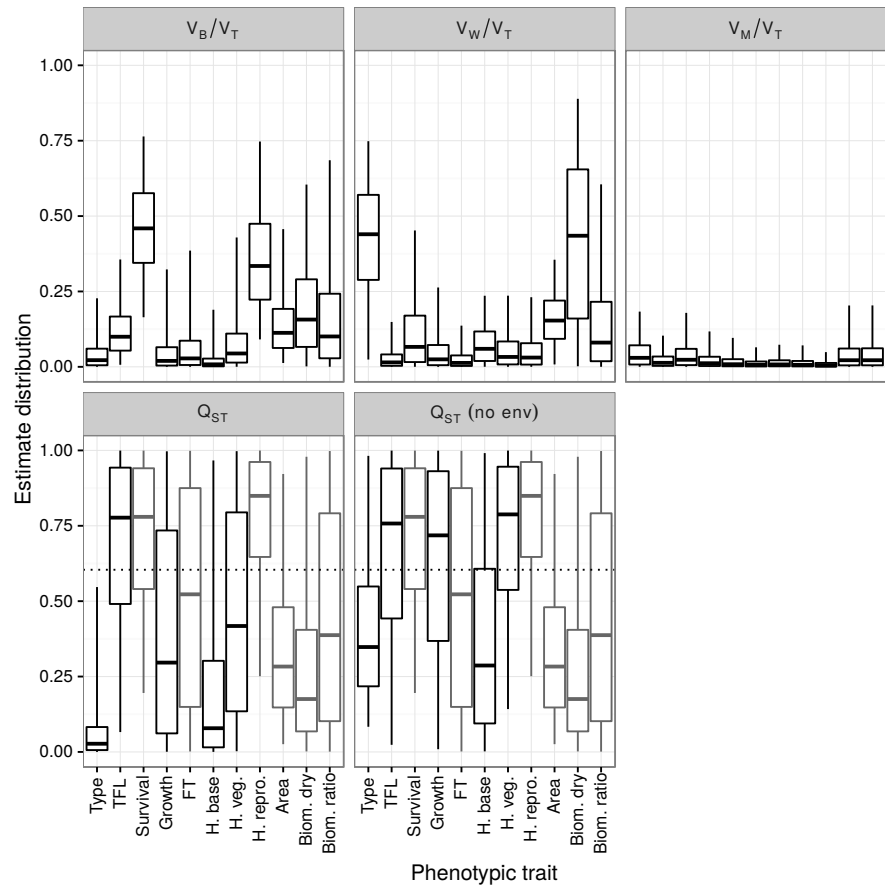
In the analysis using only genotyped individuals (thus only from the Lautaret garden), we identified five traits with a significant effect of an environmental variable (TAB. A5.3). These traits and their effect were very similar to the inferences in the global analysis. Despite not being significant, average temperature was very marginal for reproductive height ( $p = 0.055$ ). Fitting the significant variables partly diminished the  $Q_{ST}$  estimates (FIG. A5.4, bottom panels), but only dramatically so for the morphotype and almost imperceptibly so for the total fruit length.

Again, despite the absence of significant relationship with the studied environmental variables, both the ICC corresponding to  $V_B$  and the  $Q_{ST}$  were high for survival (FIG. A5.4), suggesting a strong (possibly non neutral) phenotypic differentiation, which we fail to explain using our environmental variables.

Despite a strong signal of local adaptation, the morphotype was also one of the phenotypic traits with the greatest proportion of variance explained by within-population genetic variance (FIG. A5.4, top-middle panel). The dry biomass variance was better explained by the within- than by the between-population genetic variance. This resulted in a small  $Q_{ST}$  value (FIG. A5.4), but the uncertainty around this estimate is too large to conclusively suggest potential balancing selection.

Finally, for all traits, maternal effects explained very little of the total variance (FIG. A5.4, top-right panel).

**FIG. A5.4:** Results for the **genotyped analysis**: Intraclass correlation coefficients (ICC, i.e. ratio of the effect variance to the total variance  $V_T$ ) and  $Q_{ST}$  estimates for the 11 phenotypic traits. ICCs shown correspond to the between-population genetic variance  $V_B$  (top-left panel), the within-population genetic variance  $V_W$  (top-middle panel) and the maternal effect variance  $V_M$  (top-right panel). Bottom panels show the  $Q_{ST}$  estimates when environmental effects are fitted in the model (bottom-left panel) and without environmental effect (bottom-right panel). When no environmental effect was significant, both estimates are identical and thus displayed in grey. The boxes and whiskers correspond to the 50% and 95% inter-percentile intervals respectively, the middle corresponds to the point estimate. The dotted line correspond to the estimated value of the  $F_{ST}$ .



**TAB. A5.5:** Number of significant loci for the two methods used and the seven tested environmental variables: elevation (Elev.), aspect, average temperature (Avg. Temp.), daily temperature range (Temp. Range), average humidity (Avg. Hum.), length of growing season (Season Len.) and number of freezing days (Freezing).

	Elev.	Aspect	Avg. Temp.	Temp. Range	Avg. Hum.	Season Len.	Freezing
BayeScEnv	11	34	10	37	51	36	36
LFMM	46	49	40	67	65	30	66

## Genome scans results

### ► Association studies

The association study identified between 3 and 587 loci significantly associated with one of the 11 phenotypic traits (TAB. A5.4). The number of significant loci associated to growth (587) was very high, especially when compared to the other traits. Hence, we considered this association as being strongly compromised with false positives and decided to not further interpret this result. In total, 150 loci were found significantly associated with at least one trait, among which 58 were located in a genic region.

### ► Genome scans

Genome scans methods detected between 10 and 67 significant loci, depending on both the method and environmental variable used. In total, 249 loci were found significant using at least one environmental variable, with much overlap between the variables. BayeScEnv yielded almost always less significant loci than LFMM. Based on the small number of populations and the characteristics of each method (de VILLEMEREUIL & GAGGIOTTI, 2015), this could be due to BayeScEnv being less powerful than LFMM under this conditions. There was little overlap between the output of the two methods with only 4 loci found significant by both methods.

**TAB. A5.6:** List of the 13 genes detected as selected and associated with adaptive phenotypic traits in *Arabis alpina*, and if available, their homologues in *Arabidopsis thaliana*. See TAB. A5.2 and TAB. A5.5 for the meaning of the abbreviations.

<i>Arabis alpina</i>			<i>Arabidopsis thaliana</i>	
Gene ID	Environment	Trait	Gene ID	Name/Function/Comment
120	Aspect, Temp. Range, Freezing	H. repro., H. veg.	—	—
1190	Avg. Hum., Freezing	H. repro.	—	Numerous homologues with genes of the pentatricopeptide-repeat protein family
1327	Elev., Avg. Temp., Temp. Range, Avg. Hum., Season Len., Freezing	H. repro.	—	—
1688	Temp. Range, Freezing	H. veg.	AT1G14970	O-fucosyltransferase family protein
3899	Aspect, Avg. Temp., Avg. Hum.	Type	AT1G60500	DRP4C (expressed in egg cells only)
3900	Aspect, Season Len.	H. base	AT1G60500	DRP4C (expressed in egg cells only)
11661	Aspect, Temp. Range, Avg. Hum., Season Len., Freezing	H. repro.	AT2G39290	PGP1 (phosphatidylglycerol biosynthetic process in chloroplast and mitochondria)
12168	Freezing	H. repro.	AT2G43700	LECRK-V.4 (kinase activity)
15932	Aspect, Freezing	Survival	AT4G02660	BCHA2 (signal transduction)
17383	Elev., Temp. Range, Avg. Hum., Freezing	H. repro.	AT4G12010	Disease resistance protein
18233	Temp. Range, Freezing	H. base	AT2G24960	unknown protein
20812	Avg. Temp., Season Len.	H. base, H. repro.	—	—
26173	Elev., Aspect, Temp. Range, Freezing	TFL	AT5G24300	SS1 (starch synthesis)

### ► Candidate genes

We excluded growth from the analysis because of its large number of hits during the association study. We also excluded flowering time, area and the two measures of biomass because they did not display signs of local adaptation. Using both results from the association studies and genome scans, we drew a list of 40 loci which comprised 13 genes (see TAB. A5.6). Among these 13 genes, 10 had significant homologues in the *A. thaliana* genome (two of them had the same homologue).

Only two of the homologues had unknown function in the *A. thaliana* species. However, very few of the genes had a precisely known biological process in the gene ontologies of the TAIR database. Only two genes, AT5G24300 (starch synthesis) and AT4G12010 (disease resistance) were involved in obvious biological and ecological processes. The latter was associated with reproductive height, which was influenced by the white rust (TAB. A5.2). This might explain why a disease resistance protein got associated with reproductive height.

The Blast output for the gene 1190 in *A. alpina* had a large number of hits (2,156) mainly with members of the pentatricopeptide-repeat protein family, although with a maximum “bit score” of only 197. Despite its recent discovery (LURIN et al., 2004), this family is known to be involved in RNA post-transcriptional processes in the chloroplasts and mitochondria (SCHMITZ-LINNEWEBER & SMALL, 2008). In fact, we found four genes (1190, 11661, 15932, 26173) which homologues’ products are located in the chloroplast or the mitochondria, which is a significant enrichment compared to the whole genome of *A. thaliana* (44.4% against 11.5%,  $\chi^2 = 6.61$ ,  $p = 0.01$ ).

## Discussion

### Patterns of local adaptation

Using an approach combining common garden phenotyping, high-throughput genotyping and *in situ* environmental data, we were able to detect patterns of local adaptation and produce insight into their potential genetic bases. Several of the studied phenotypic traits were statistically linked with local environmental conditions. The morphotype, vegetative and reproductive heights were linked with the average temperature, indicating a selection for smaller and more compact plants in cold conditions. This selection could result from physiological limitations due to lower temperatures and/or lower partial CO<sub>2</sub> pressure (KÖRNER, 2007). The signal of selection accounted for much of the inter-population variance for those traits, as illustrated by the influence of accounting for average temperature on  $V_B$  (FIG. A5.3) and the  $Q_{ST}$  (FIG. A5.4).

Total fruit length and growth were also linked with local average temperature, indicating a selection for higher reproductive output in warmer conditions. This result is consistent with previous findings in natural populations (see ARTICLE 4), which found that growth rate covariates positively with “late” average temperature during the reproducing seasons. Despite such a relationship not being significant *in situ* for the reproductive effort, the coldest population (GAL) still displayed a consistently low reproductive effort when compared to the other populations (see ARTICLE 4). Also, such a local adaptation on reproductive effort was found in a latitudinal study in Spanish and Swedish populations of *A. alpina* (TORÄNG et al., 2015). Note that the average temperature did not account for much of the inter-population variance in the case of the total fruit length (Figs. A5.3 and A5.4). The high  $Q_{ST}$  value for this trait (FIG. A5.4, bottom panels) however suggests that the remaining variance to be explained might also be partly of adaptive origin.

Two other traits displayed a signal of local adaptation but only in one of the gardens. Basal height was associated with aspect, but only in the Lautaret garden, whereas area was linked to season length, but only in the Vercors garden.

That both results are significant in only one garden might stem from an interaction between the expression of the phenotypes and the experimental garden: plants were *c.* 10 times larger in the Vercors garden than in the Lautaret one (average individual surface area of 12,442mm<sup>2</sup> over 1,266mm<sup>2</sup> respectively, for the year 2014,  $F_{1,522} = 406$ ,  $p < 2.10^{-6}$ ).

This was most likely due to a warmer environment as well as more fertile soil conditions in the Vercors garden. As a consequence, plants were more limited in growth in the Lautaret garden, masking possible trends regarding the individual surface area, whereas the overall large number of stems in the Vercors potentially masked any trends in the Vercors garden (heights being measured as a maximum, their distribution is heavily modified by the number of stems). Note that, because season length is strongly correlated with average temperature ( $\rho = 0.82$ ,  $p < 0.048$ ), the relationship between the area and season length is consistent with the pattern of smaller, more compact plants in cold conditions. The relationship between basal height and aspect, however, testify of the presence of another gradient of selection involving the aspect rather than the elevation.

The scenario of local adaptation for these traits is rendered even more likely by the fact that spatial environmental variability is much greater than the temporal one, as shown by the discriminant analyses. This preponderance of space over time is indeed a prerequisite for local adaptation (KAWECKI & EBERT, 2004). Note however that, because we used a F1 generation rather than a F2, adaptive maternal effects could explain the obtained results along with local adaptation (ROACH & WULFF, 1987). Although we cannot totally exclude this scenario, the fact that maternal effects were negligible for all traits (see FIG. A5.4, top-right panel) is a strong indication that this might not be the case.

The case of survival is ambiguous. Despite trends suggesting local adaptation in survival (e.g. large  $Q_{ST}$  value), we could not find an environmental variable to which it was associated. This can be explained by the fact that the populations with the lowest survival were LAU (2000m of elevation) and PIC (3000m of elevation), whereas GAL (2500m of elevation) had a survival rate comparable



to the other populations. Once again, these results can be related to the study on *in situ* survival (see ARTICLE 4), that found a very strong negative trend with the “late” average temperature for all populations except for LAU and PIC.

## Phenotypic plasticity and genotype-by-environment interaction

There was considerable phenotypic plasticity between the two experimental gardens. Plants in the warmer, more humid and more fertile garden in the Vercors were *c.a.* 10 times larger area and had *c.a.* 16 times more fruits. This plasticity was captured in our analysis by the Garden and Year effects in the “Global analysis” (see TAB. A5.2). The Garden effect was significant for all traits but two (morphotype and flowering time) and Year was significant for flowering time, individual surface area and the three height measures.

We also measured whether there was genotype-by-environment interaction ( $G \times E$ ) by measuring the population-by-garden random effect variance  $V_{G \times E}$ . The contribution of this effect to the total variance was considerable for morphotype and flowering time and negligible for the three kinds of heights and total fruit length.

Hence, the two traits for which the Garden effect was not significant (morphotype and flowering time) also were those with the greatest contribution of  $V_{G \times E}$  to the total variance. However, running a model without  $G \times E$  resulted in a significant Garden effect for the flowering time only. This suggests that a strong  $G \times E$  is masking a slight effect of the environment alone in flowering time, but not for the morphotype. This is sensible, as morphotype is an integrated trait which is probably not much influenced by the environment.

Overall, grouping together the Garden and  $G \times E$  effects under the denomination of “phenotypic plasticity” (*sensu lato*), every studied traits showed clear signs of plasticity. Yet, it was still possible to detect patterns of local adaptation to the average temperature in both gardens for the morphotype, total fruit length and vegetative and reproductive heights. This indicates that both local adaptation and phenotypic plasticity plays a role in allowing the wide altitudinal range of *Arabis alpina*. Because distinguishing both phenomena in natural populations is virtually impossible, it is, however, extremely difficult to quantify the relative role of both mechanisms, which might be further confounded by drift.

Yet, if we consider that the signal of local adaptation is sufficiently strong for the morphotype for example, we could quantify those in our experimental settings using the corresponding effects: drift is measured by  $V_B$  which has a contribution of 3.65% to the inter-population variance, local adaptation is measured by the variance due to the average temperature effect, which has a contribution of 62.6% to the inter-population variance and finally, plasticity is measured by the  $G \times E$  effect which has a contribution of 29.9% to the inter-population variance. In that particular case, we could infer, using many hidden assumptions (mainly that our  $G \times E$  is an accurate measure of the naturally occurring one), that local adaptation is mostly driving the inter-population differentiation for the morphotype.

Regarding phenotypic plasticity, the case of flowering time is interesting in that this trait is calculated as the number of weeks between the end of snow cover in the focal experimental garden and the recorded date of flowering. Hence, since the end of snow cover in the Lautaret garden was two months later than the end of snow cover in the Vercors garden, it is true that flowering date is plastic between gardens, but not flowering time. Flowering time, however, was significantly plastic between years 2014 and 2015 in the Lautaret garden. Indeed, although the end of snow cover was 17 days earlier in 2015 than in 2014, the flowering dates were roughly similar. This result might, however, be an artefact due to the fact that some plants were already flowering at the first visit in 2015.

## Detection of candidate genes for local adaptation

Using high-throughput genotypic data, we were able to perform two kinds of genome scans: using association studies, we were able to isolate loci significantly linked to phenotypic traits and using genome scan methods to detect selection, we were able to identify loci significantly associated with some environmental variables. Combining both conditions, we isolated 40 loci that displayed association with both a phenotypic and an environmental variable among which 13 laid within a gene.

Some of these genes had homologous counterparts in the genome of *Arabidopsis thaliana*. Of course, even when the function in *A. thaliana* is known, this does not imply that the function is conserved in *A. alpina*, but it is nonetheless a strong indication regarding the potential function of the gene's product in the latter species. Interestingly, this short list was significantly enriched in genes which products that are located within chloroplasts and/or mitochondria.

Among the homologous gene that have a known function in *A. thaliana*, the most promising is AT5G24300/SS1, which is involved in starch synthesis in *A. thaliana* and associated with total fruit length in *A. alpina*. Starch synthesis is indeed a key process in the storage of energy both directly to grow fruits and to store energy in the seeds. Because it was not possible to correct for the effect of white rust in our analyses, it is possible that some genes involved in disease resistance (AT4G12010), found under selection and associated with the reproductive height, are “false positives”. That is, in the sense that there are not, in fact, associated to reproductive height. However, they still represents interesting candidate genes for disease-driven local adaptation.

Regarding the other genes with known function, the link between their function in *A. thaliana* and the phenotypic trait they are associated with in *A. alpina* is more distant. Some, however, were found to be involved in interesting biological patterns. The *A. thaliana* knockout mutant of the gene AT1G14970, for example, which was found associated with vegetative height in our study, displays a comparable silique length than Columbia wildtype at 20°C, but significantly longer ones at 28°C (HSIEH, 2009). The fact that AT1G14970 is involved in temperature sensing was proposed to explain this phenomenon (HSIEH, 2009).

As another example, the gene AT2G43700 has been found to have an effect on nitrogen use efficiency («Use of CAD genes to increase nitrogen use efficiency and low nitrogen tolerance to a plant» 2015). This could be related to nitrogen depletion in high elevation soils (KÖRNER, 2003).

However, the link between the morphotype and a gene (AT1G60500/DRP4C) only expressed in egg cells *A. thaliana* is more ambiguous. This link is especially puzzling because the DRP4 subfamily seems to be homologous to the animal antiviral Mx proteins, although the function of DRP4C is still unknown (HONG et al., 2003). The link between AT4G02660/BCHA2 (a transduction protein in the chloroplast) and survival is also unclear, although this gene has been found to be expressed in trichome cells (WIENKOOP et al., 2004).

Finally, the fact that some loci that are significant for both selection and association with a phenotypic trait do not lay within a genic region does not mean that they are not good candidates for further study. Their status is simply more unclear than for genes with homologues in *A. thaliana*, as there is no supplementary information to give insight into their potential functions.

It is important to stress here that we do not consider homology with a gene with a consistent function in *A. thaliana* as a confirmatory result that our candidates are indeed true positives. This would require functional validation in *A. alpina* and replication using other populations (BUEHLER et al., 2014), which is beyond the scope of this paper.

## Conclusion

Using phenotypic, genotypic and environmental data, we were able to detect local adaptation in six populations of *Arabidopsis alpina* and link it to potential selective environmental variables. To do so, we did not use the classical  $Q_{ST}$ - $F_{ST}$  framework, but a parametric environmental analysis accounting for genetic drift. This allowed us to detect significant signs of local adaptation despite a small number of populations and a very strong neutral genetic differentiation between populations ( $F_{ST} = 0.60$ ). This

framework of analysis can be related to the common practice in phylogenetic comparative methods (PCM Felsenstein, 1985 ; Garamszegi, 2014), whereby evolutionary biologists test for biologically constrained trends (e.g. allometry) or adaptive evolution while accounting for evolution along the phylogeny. The idea of using PCM to account for inter-population drift rather than inter-specific evolution was already suggested (Stone et al., 2011), but here, we used an explicit and rigorous framework to account for genetic drift (Ovaskainen et al., 2011).

Overall, our results show that the main environmental factor driving local adaptation in *Arabis alpina* is the average temperature (which decreases with elevation, but is also influenced by aspect). The evolutionary consequences of colder temperature in this species are to produce smaller and more compact plants, with a slow growth and which produce less fruits. Such a trend has often been described for inter-specific variations along an altitudinal gradient. Short and compact stature, for example, is known to help plants to decouple their temperature from atmospheric cold temperatures, hence helping to keep photosynthetic activity sufficiently efficient, as is illustrated by cushion plants (Körner, 2003). Slow growth, lower productivity and higher survival are also typical traits of alpine plant species (Körner, 2003). Our results are well aligned with these ecological expectations, with the notable exception of survival. These results overall suggest that local adaptation is a key feature that allows *A. alpina* to display such a large altitudinal distribution, enabling this species to overcome, to some extent, the environmental limitations imposed by the altitudinal gradient.

The evolutionary mechanism of this local adaptation remains to be investigated. Among the three traits most related to fitness (growth, total fruit length and survival), the directional selection was positively correlated for two (growth and total fruit length), but no sign of opposite directional selection has been found for the last one (survival).

It seems likely that local adaptation is due a genetic assimilation (Waddington, 1953 ; Waddington, 1959 ; Pigliucci et al., 2006) of physiological and resource limitations in high elevation conditions (lower temperatures, lower partial CO<sub>2</sub> pressure and lower soil fertility). Note, however, that the afore-mentioned compromise between growth and survival was shown in the wild populations (see TAB. 4).

Finally, we isolated a list of 17 genes that displayed both a signal of selection due to one of the environmental variable studied and a significant association with one of the studied phenotypic traits. This list of candidate genes constitutes the foundations for further studies of the genetic bases of local adaptation in *Arabis alpina* with the ultimate objective of understanding the mechanisms that lead to its wide altitudinal range.

## Acknowledgements

We are indebted to Fabrice Roux who suggested to record the morphotype as a phenotypic trait, as well as to Jinliang Wang, Matthew Wolak, Markku Karhunen and Otso Ovaskainen for their help during the analysis. We are also grateful to Delphine Rioux and Thibaut Capblancq who were of great help for the genotyping. We thank all the people who helped with the common garden experiment including Aymeric Pilleux, Pascal Salze, Amandine Deschamps, Quentin Rousseau, Perrine Augrit, Sofia Karabatsos, Charles Pouchon, Thomas Goeury, Mónica Cecilia Corona, Tristan Cumer, Loïc Chalmandrier, Alex Farajdo and Justine Bisson. PdV was supported by a doctoral studentship from the French *Ministère de la Recherche et de l'Enseignement Supérieur*. OEG was supported by the Marine Alliance for Science and Technology for Scotland (MASTS).

## 4 Adaptation locale le long d'un gradient altitudinal

### 4.1 Un syndrome phénotypique

Les résultats de notre étude de jardin commun ont donc montré qu'il existe bien de l'adaptation locale chez *A. alpina* et que celle-ci était majoritairement liée à la diminution de la température moyenne avec l'altitude (ainsi que d'autres facteurs comme l'exposition). Les traits phénotypiques impliqués étaient le morphotype, la croissance, la longueur totale de fruits ainsi que la hauteur de la plante (végétative et reproductive).

Comme expliqué dans l'article, ces résultats montrent que les *A. alpina* d'altitude (ou du moins des milieux froids) possèdent une tendance génétique à être plus petites et compactes, à croître plus lentement et à produire moins de fruits. Ce "syndrome" est à mettre en relation avec celui qui ressort de l'analyse des traits phénotypiques *in situ* : les plantes des milieux plus froids ont une croissance plus lente et une meilleure survie. Dans le cas de la population la plus froide, la production de fruits est aussi significativement diminuée.

Si l'on prend l'ensemble de ces données en compte, il est indéniable qu'il existe un "syndrome" phénotypique d'*A. alpina* lié à la température moyenne (et/ou à l'altitude) et que ce syndrome possède une base génétique, qui est d'origine adaptative. Nous pouvons donc en déduire plusieurs choses. Tout d'abord, le gradient d'altitude, au moins de par la baisse de température, exerce bien une pression de sélection sur les populations d'*A. alpina*, ce qui était notre attendu, compte-tenu de la forte variabilité environnementale qu'il entraîne.

Ensuite, il existe (ou a existé) un potentiel adaptatif (c'est à dire de variance génétique sur des traits adaptatifs clés) suffisant chez *A. alpina* pour répondre à cette pression, ce qui engendre le patron d'adaptation locale que nous venons de détecter. En outre, le flux de gènes modéré effectivement observé chez cette espèce n'a pas été suffisant pour contrer les effets de l'adaptation locale.

Enfin, la matrice de variance-covariance génétique (ou matrice G) ne semble pas avoir été contraignante sur l'évolution de ce syndrome. Plus précisément, les corrélations génétiques entre les différents traits (morphotype, croissance, taille et effort reproducteur) ne semblent pas avoir été orientées contre le gradient de sélection. Malheureusement, nous ne pouvons conclure sur cette hypothèse que de manière indirecte (nous observons de l'adaptation locale cohérente sur un ensemble de traits, donc elle n'a pas été empêchée), mais le trop faible nombre d'individus disponibles ne nous permet pas d'aller étudier directement ces corrélations génétiques<sup>1</sup>.

Bien sûr, il est loin d'être rare de trouver de l'adaptation locale le long d'un gradient altitudinal (BYARS et al., 2007 ; GIMENEZ-BENAVIDES et al., 2007 ; GONZALO-TURPIN & HAZARD, 2009 ; HAUTIER et al., 2009 ; ALBERTO et al., 2011 ; LUQUET et al., 2015). En outre, certaines des caractéristiques du syndrome détaillé ci-dessus ont été retrouvées dans d'autres études de jardin commun. Chez *Festuca eskia*, par exemple, la survie augmente et la reproduction diminue avec l'altitude (GONZALO-TURPIN & HAZARD, 2009). HAUTIER et al. (2009) ont aussi montré que, chez *Poa alpina*, les plantes de plus haute altitude étaient plus petites et possédaient un effort reproducteur réduit. La convergence entre ces deux études et la nôtre tendrait à suggérer que ce syndrome serait une réponse attendue aux contraintes que posent les conditions d'altitude dans les milieux tempérés européens.

---

<sup>1</sup>L'estimation de ces matrices requiert en effet un très grand nombre d'individus (de l'ordre du millier) dont les relations d'apparentement sont connues. Notre jeu de données est donc largement insuffisant pour un tel exercice.

## 4.2 Bases génétiques de l'adaptation

À l'aide des données phénotypiques et génotypiques de cette étude, nous avons pu analyser les bases génétiques de cette adaptation. Pour ce faire, nous avons combiné des méthodes de scan génomique (association phénotypique et détection de sélection), afin de détecter des marqueurs montrant à la fois des signes d'association avec un trait phénotypique dont nous avons pu montrer qu'il était adaptatif et des signes de sélection avec une ou plusieurs variables environnementales.

Cette étude a été rendue possible, entre autres, par l'utilisation du protocole de génotypage haut-débit ddRADseq, qui nous a permis d'obtenir un panel de marqueurs relativement dense dans le génome d'*Arabidopsis alpina*. Ce génome ayant une taille d'environ 350Mpb (LOBRÉAUX et al., 2014), nos 3528 marqueurs permettent d'avoir un marqueur toutes les 100 000 paires de bases en moyenne. Cette densité reste cependant limitée comparée à celles qui peuvent être obtenues par un séquençage complet. Il n'est pas vraiment pertinent, par exemple, d'utiliser le déséquilibre de liaison entre marqueurs dans nos analyses (que ce soit dans des analyses statistiques ou dans des graphiques de Manhattan, par exemple). En outre, de nombreux gènes ne sont pas couverts par nos marqueurs, limitant donc l'exhaustivité de notre étude.

Pourtant, nous avons pu établir une liste de 40 locus montrant à la fois des signes de sélection et d'association avec un trait phénotypique. Parmi ces 40 locus, 13 se trouvent à l'intérieur d'un gène chez *A. alpina*. En se limitant aux variables environnementales fortement liées à l'altitude (altitude, température moyenne et longueur de la saison de croissance), cette liste tombe à 18 locus, dont 7 se trouvent à l'intérieur d'un gène (1327, 3899, 3900, 11661, 17383, 20812, 26173).

À l'aide de la séquence du génome de référence disponible pour *A. alpina* (WILLING et al., 2015), nous avons pu effectuer un alignement local des séquences des gènes annotés dans ce génome sur une banque de gènes disponible pour l'espèce modèle *Arabidopsis thaliana* afin d'obtenir quelques informations sur le rôle possible de ces gènes chez *A. alpina*. Bien que l'homologie de la séquence ne garantisse pas l'identité de la fonction, la fonction du gène chez *A. thaliana* reste un indice fort de la fonction potentielle de son homologue chez *A. alpina*. Pour autant, la plupart des fonctions retrouvées chez *A. thaliana* sont peu éclairantes sur la fonction que peuvent avoir ces gènes candidats chez *A. alpina* (cf. TAB. A5.6). Comme expliqué dans l'article, le lien le plus "logique" est celui retrouvé entre le gène SS1 qui participe à la synthèse de l'amidon et la longueur totale de fruits. Pour les autres cas, il est possible d'envisager des liens plus ou moins forts à l'aide des informations scientifiques publiées. Tout ceci est discuté plus amplement dans la discussion de l'ARTICLE 5.

Il s'agit ici plutôt de revenir sur les faits suivants. Premièrement, en prenant en compte les résultats du CHAPITRE 1, il n'est bien sûr pas exclu que notre liste de gènes candidats contiennent des faux positifs. Nous avons cependant limité ce risque en combinant différents critères, provenant de différentes sources d'information (environnement et phénotype). La fonction des homologues détectés chez *A. thaliana* permet aussi de pondérer le probable intérêt de nos candidats (par exemple, SS1 semble a priori un meilleur candidat que LECRK-V.4).

Deuxièmement, en lien avec nos efforts pour limiter les faux positifs, cette liste reste relativement courte. Cela permet à des études ultérieures de pouvoir confirmer ou infirmer individuellement chacun de ces gènes candidats. Ces études pourront être basées sur une validation fonctionnelle de ces gènes chez *A. alpina* (études par mutants *knock-out*, par exemple) ou une étude de génétique du paysage de ces gènes, utilisant plus de populations.

Troisièmement, les locus qui n'appartiennent pas à des régions géniques annotées chez *A. alpina* ne sont pas nécessairement à considérer comme des faux positifs. Ils peuvent en effet se trouver dans des régions géniques non annotées, en déséquilibre de liaison avec un gène<sup>1</sup>, ou encore participer à la régulation *trans* de l'expression d'un autre gène.

## 5 Du lien entre le phénotype *in situ* et en jardin commun

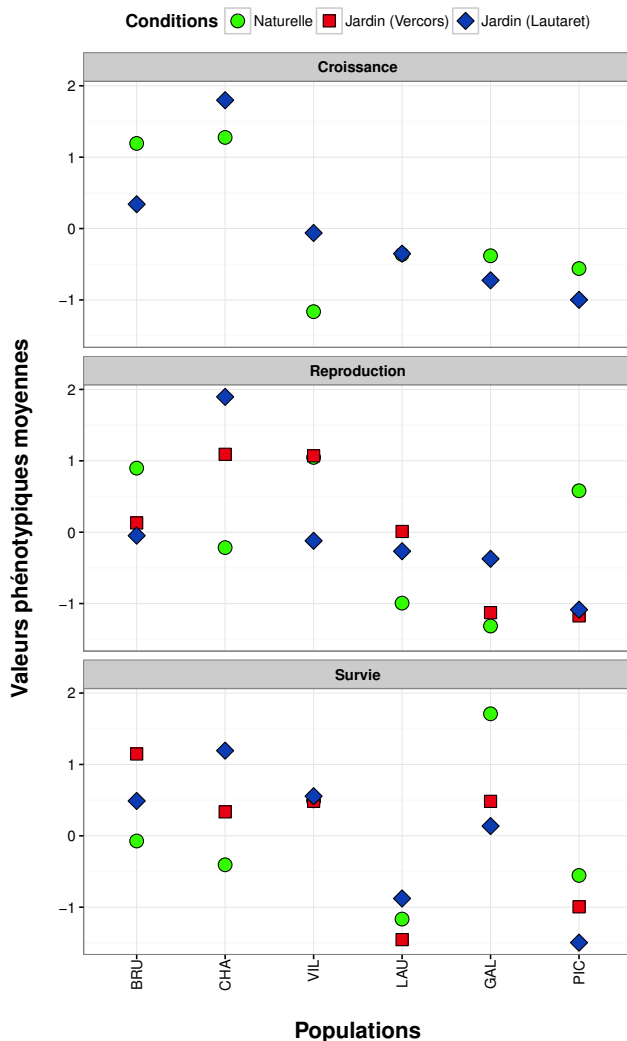


FIG. 9 : Comparaison entre des valeurs moyennes de traits phénotypiques (par population) entre trois différentes conditions : en milieu naturel (rond vert), dans le jardin du Vercors (carré rouge) et dans le jardin du Lautaret (losange bleu). Les traits ont été standardisés pour faciliter la comparaison.

sance et la survie (FIG. 9). Les variations entre jardins expérimentaux semblent cependant moindre

Bien qu'il ait été légèrement abordé, il est intéressant de revenir sur le lien entre le phénotype *in situ*, mesuré dans l'ARTICLE 4 et le phénotype en jardin commun analysé dans l'ARTICLE 5. Pour ce faire, le mieux est d'analyser les moyennes phénotypiques pour chaque population, dans leur milieu naturel et dans chacun des deux jardins (FIG. 9).

Avant d'analyser la FIG. 9, il est important de préciser à quel point les traits phénotypiques sont comparables entre le milieu naturel et les jardins expérimentaux. Pour plus de cohérence avec les études présentées dans les deux articles, nous avons conservés la manière de mesurer les traits phénotypiques propres à chacune de ces études. Ainsi, entre le milieu naturel et les jardins expérimentaux, la croissance est-elle mesurée respectivement à partir du nombre de tiges totales et à partir de l'aire ; la reproduction respectivement comme le nombre de siliques et comme la longueur totale de fruits et la survie respectivement comme la probabilité de survivre à l'année suivante et la survie cumulée le long de l'expérience.

Pour être comparés sur la même échelle, les traits ont tous été centrés-réduits.

Malgré les différences de calcul dans les différents milieux, on remarque une forte correspondance des phénotypes moyens entre les conditions naturelles et expérimentales, notamment en ce qui concerne la crois-

<sup>1</sup>Notre critère de présence au sein d'un gène était en effet relativement strict. En réalité, un marqueur peut ne pas se trouver à l'intérieur d'un gène, mais y être lié. Tout dépendra du déclin du déséquilibre de liaison le long du génome. Malheureusement, la densité de nos marqueurs ne nous permet pas de l'étudier avec une précision suffisante.

que les variations entre chaque jardin et les conditions naturelles. Pour le cas particulier de la reproduction, il est important de noter que la moyenne de la population PIC est largement tirée vers le haut par le fait qu'un seul gros individu représente la population pour les années 2011 et 2012 du suivi. Quoiqu'il en soit, il s'agit du trait où la relation entre les phénotypes en conditions naturelles et expérimentales est la plus mauvaise.

Que le lien entre les valeurs moyennes phénotypiques en conditions naturelles et expérimentales soit notable est particulièrement informatif. D'abord, cela suggère que, bien que la présence d'une plasticité phénotypique soit indéniable (notamment en comparant les deux jardins), l'effet de l'adaptation locale explique relativement bien à elle seule la variabilité phénotypique inter-population observée dans les conditions naturelles (au moins dans le cas de la croissance et de la survie). Ensuite, dans le cas de la survie justement, il avait été difficile de conclure à la présence d'adaptation locale sur ce trait uniquement sur la  $Q_{ST}$ , puisque les relations avec l'environnement n'étaient pas significatives. La relation assez forte (sans pourtant être significative dû au faible nombre de points) entre ce trait phénotypique en conditions naturelles et expérimentales est un signal de plus qu'il existerait bel et bien de l'adaptation locale sur ce trait, qu'il serait difficile de modéliser simplement à l'aide de nos données environnementales.

## 6 Petit retour sur la méthodologie

### 6.1 De l'incertitude autour des estimations

Dans l'ARTICLE 5, la FIG. A5.4 montre non seulement les estimations pour la  $Q_{ST}$  mais aussi l'incertitude autour de celle-ci. On peut remarquer que cette incertitude est considérable, puisque l'intervalle de crédibilité à 95% couvre quasiment l'intégralité du domaine des possibles (entre 0 et 1).

Cette incertitude possède différentes origines. Tout d'abord, l'estimation de la variance d'un effet aléatoire se fait toujours avec une forte dose d'imprécision, donc d'incertitude. Lorsque plusieurs variances sont impliquées dans le calcul d'un paramètre (héritabilité ou  $Q_{ST}$ , par exemple), les incertitudes liées aux différentes variances se cumulent. Ensuite, l'utilisation de modèles linéaires mixtes généralisés entraîne une incertitude supplémentaire (de VILLEMEREUIL et al., 2013) due au fait que les estimations se déroulent sur l'échelle sous-jacente (voir FIG. 8). Enfin, le faible nombre d'individus utilisés ici (152) aggrave encore tous ces problèmes.

Ce qu'il y a d'intéressant dans la façon dont ont été estimées les incertitudes dans cette étude est qu'il s'agit de distributions postérieures bayésiennes. L'avantage de ces distributions est qu'elles reflètent toute la complexité de cette incertitude, c'est à dire qu'elles ne reposent pas sur une approximation Gaussienne de l'incertitude autour du trait en question<sup>1</sup>. Or, en constatant l'étalement des distributions entre 0 et 1, l'importance de certaines queues de distribution et surtout l'asymétrie de celles-ci, il est clair qu'une approximation Gaussienne ne peut pas tout à fait tenir sur un paramètre tel que la  $Q_{ST}$  (ou l'héritabilité d'ailleurs).

Il serait bon de faire des analyses propres pour le démontrer, mais il se peut que des comparaisons  $Q_{ST}$ - $F_{ST}$  fréquentistes, prenant en compte l'incertitude sur la  $Q_{ST}$  par une approximation Gaussienne soient des tests légèrement trop laxistes parce qu'ils sous-estimeraient l'incertitude autour de l'estimation de la  $Q_{ST}$ , spécifiquement lorsque la  $F_{ST}$  et/ou la  $Q_{ST}$  se rapproche des bords (0 et 1).

---

<sup>1</sup>L'inconvénient est qu'elle est sensible à l'a priori fourni. Dans notre cas, il se peut que les valeurs de  $Q_{ST}$  soient légèrement sous-évaluées à cause de l'influence de cet a priori.

## 6.2 De l'incertitude autour du modèle neutre

Un autre problème d'incertitude est posé par la façon dont nous estimons et utilisons notre "modèle neutre" dans l'article ci-dessus. Ce modèle neutre est principalement représenté par une matrice de variance-covariance  $\mathbf{B}$  entre populations, qui nous sert à modéliser la covariation attendue entre populations sous l'hypothèse de dérive pure. Ce modèle possède des bases théoriques solides (OVASKAINEN et al., 2011) et la façon dont nous l'utilisons (comme modèle nul pour le test d'une régression linéaire) n'est pas sans lien avec les analyses comparatives phylogénétiques (GARAMSZEGI, 2014).

Pourtant, la littérature autour des analyses comparatives phylogénétiques a montré que ne pas prendre en compte l'incertitude sur le modèle nul lors d'un test de significativité pouvait entraîner de légers problèmes de faux positifs (de VILLEMEREUIL et al., 2012). Il serait donc nécessaire de réitérer ces analyses en prenant en compte cette incertitude (KARHUNEN et al., 2013 ; KARHUNEN et al., 2014), ce qui n'a pas (encore) été fait ici, essentiellement par manque de temps. Étant donné le grand nombre de locus utilisés ici, l'incertitude ne devrait cependant pas être très forte et sa prise en compte ne devrait pas changer qualitativement nos résultats.

## 6.3 Scan génomique et scan génomique

Dans l'Introduction générale, nous avons pris soin de différencier deux types de "scans génomiques" : ceux dont le but était de détecter de la sélection d'une part, et ceux dont le but était d'associer certains locus à un trait phénotypique d'autre part. À priori, à part l'idée de "scanner le génome", ces deux types de méthodes n'ont pas grand-chose en commun.

Pourtant, elles partagent certains points communs. Elles sont toutes sensibles à la structure de population, d'une manière ou d'une autre. De plus, elles entraînent un très grand nombre de tests multiples et donc l'utilisation des concepts statistiques associés à ce genre de cas (taux de fausse découverte,  $q$ -valeur, facteur d'inflation génomique).

Dans l'article ci-dessus, nous avons montré qu'il est possible de pousser le parallèle légèrement plus loin, puisque nous avons utilisé une méthode de détection de sélection comme une méthode d'association phénotypique. La méthode utilisée, LFMM, tente d'associer significativement la variabilité génétique à un locus particulier avec la variabilité d'une variable environnementale testée (tout en corrigeant pour la structure de population). Que l'on associe le locus à l'environnement ou l'environnement au locus ne fait statistiquement aucune différence : il s'agit simplement de montrer qu'il existe une co-variation des deux variables, et non un effet de l'une sur l'autre. De même, associer un locus à un trait phénotypique ou l'inverse ne fait pas non plus de différence. Ce qui fait une différence, c'est que l'on va considérer que l'un est connu (l'effet fixe, le prédicteur) et l'autre mesuré avec une certaine erreur, ou possédant une variabilité intrinsèque (la variable réponse).

Dans notre cas, le génotype possède une distribution qu'il est possible de modéliser, et que l'on se doit de modéliser d'une manière ou d'une autre, afin de prendre en compte la structure de population. Le phénotype, par contre, ne suit pas forcément une distribution Gaussienne, comme beaucoup de nos traits phénotypiques. Ainsi, il peut être plus facile de modéliser la distribution du génotype que celle du phénotype.

Cette raison, combinée au besoin d'une méthode statistique capable de prendre en compte la structure de nos populations (entre massifs, entre populations et au sein des populations), nous a fait



choisir d'utiliser LFMM comme méthode d'association<sup>1</sup>, bien qu'elle soit à la base une méthode de scan de sélection. Il y aurait bien sûr de la place pour améliorer cette démarche, possiblement avec une autre méthode que LFMM.

---

<sup>1</sup>Merci à Olivier François pour la suggestion !

# **DISCUSSION GÉNÉRALE**

---

# 1 Discussion de l'ensemble des résultats sur *Arabis alpina*

## 1.1 Résumé des patrons phénotypiques observés

Les analyses présentées dans le CHAPITRE 4 nous ont permis d'obtenir des informations importantes sur la plante alpine *Arabis alpina*. Tout d'abord, nous avons pu constater, que bien que cette plante soit effectivement pérenne, elle possède une espérance de vie en milieu naturel extrêmement courte (1,82 ans). En outre, la survie et la reproduction peuvent être extrêmement différentes entre individus, avec environ la moitié des individus qui meurent avant toute reproduction. Compte-tenu de l'état des connaissances avant le début de cette étude (l'idée dominante étant qu'*Arabis alpina* possédait un rythme de vie plutôt lent, typique des plantes alpines), ces résultats sont plutôt surprenants.

L'un des résultats les plus intéressants du CHAPITRE 4 reste la relation entre la température moyenne "tardive" et les taux vitaux. Nous avons en effet pu montrer que la croissance et la survie variaient en sens inverse avec ce paramètre environnemental<sup>1</sup>.

La population la plus froide et à la saison la plus courte, GAL, montre en outre, une reproduction significativement diminuée par rapport aux autres. L'environnement pourrait donc avoir une influence sur la reproduction en plus de la croissance et de la survie.

Dans les analyses présentées dans le CHAPITRE 5, nous avons pu montrer des relations significatives entre la température moyenne et des traits phénotypiques mesurés dans un environnement commun : la croissance, la reproduction, la taille et le port de la plante. Étant donné que notre analyse prend en compte la variabilité phénotypique attendue sous une hypothèse de dérive neutre, ces relations significatives témoignent du caractère adaptatif de ces traits vis-à-vis des conditions de température.

La relation entre la température moyenne et la survie mesurée dans nos jardins expérimentaux n'était pas significative, bien que la variabilité entre populations de ce trait semblent être liées à la variabilité observée en milieu naturel (FIG. 9, CHAPITRE 5). La très forte variabilité sur la survie observée, notamment entre les trois populations du Lautaret (LAU, GAL et PIC), mais cohérente dans les différentes conditions de mesure (milieu naturel, jardin du Vercors, jardin du Lautaret) permet de suggérer qu'il existe une forte composante génétique au déterminisme de la survie entre ces populations. Ceci, combiné à la tendance avec la température observée en milieu naturel et la forte  $Q_{ST}$  mesurée dans nos jardins expérimentaux, suggère en retour (sans que l'on ne puisse l'affirmer définitivement) qu'il existe de l'adaptation locale sur ce trait.

L'ensemble de ces résultats nous a mené, à la fin du CHAPITRE 5, à parler d'un syndrome phénotypique lié à la température moyenne (et/ou à la pression partielle en  $CO_2$  et/ou à la durée de la saison et/ou aux radiations solaires...). Ce syndrome voudrait que les plantes des milieux plus froids soient plus petites et plus compactes, croissent plus lentement, se reproduisent moins et survivent plus longtemps (avec tous les bémols nécessaires pour cette dernière caractéristique). Il se trouve que ce syndrome correspond à un attendu écologique fort.

---

<sup>1</sup>Bien sûr, ce paramètre de température moyenne peut agir comme une variable substitutive à toute variable environnementale covariant globalement avec l'altitude (pression partielle en  $CO_2$ , radiations solaires, longueur de la saison végétative ou même propriétés du sol). Lorsque nous parlerons de relation avec la température, ce sera toujours en prenant implicitement en compte cette possibilité.

## 1.2 Analyse écologique de ces patrons adaptatifs

Le syndrome décrit ci-dessus est cohérent avec les patrons inter-spécifiques observés le long des gradients altitudinaux (KÖRNER, 2003). La différence ici est qu'il s'agit d'un patron intra-spécifique modelé par l'adaptation locale. D'une certaine manière, les attendus de macro- et de micro-évolution se rejoignent ici devant la même contrainte sélective (ou des contraintes proches). Il est donc possible d'analyser ces contraintes à l'aide des connaissances établies en écologie alpine sur les patrons de diversité fonctionnelle (de BELLO et al., 2013).

Par exemple, de manière générale, il existe un lien, dans les milieux subalpins et alpins, entre la température (ou l'altitude) et de nombreux traits fonctionnels des communautés de plantes : les plantes de haute altitude ont tendance à être plus petites, avec des feuilles épaisses et à avoir une reproduction plus faible (de BELLO et al., 2013).

Le cas du morphotype est particulièrement intéressant, puisque le morphotype "Rosettes compactes" (FIG. S5.2, C), caractéristique des populations de plus haute altitude (GAL et PIC) possède un port très fortement apparenté aux plantes en coussins (tissu végétatif compact, réparti de manière quasiment sphérique autour du centre de la plante). Ces plantes constituent une forme de vie très typique des milieux alpins (KÖRNER, 2003). Des études récentes ont montré que l'évolution de la forme de vie en coussin a joué un rôle important dans la radiation du genre *Androsace* (BOUCHER et al., 2012 ; ROQUET et al., 2013).

L'origine écologique et biologique de ce syndrome est connue. La petite taille et surtout le port compact font partie d'une stratégie morphologique pour découpler la température de la plante de la température atmosphérique (notamment grâce aux radiations solaires par beau temps KÖRNER, 2003) afin de maximiser l'efficacité métabolique de la photosynthèse. La petite taille peut, en outre, être une conséquence de la réduction de croissance dans les milieux froids. La croissance et la reproduction réduites, ainsi qu'une plus forte longévité sont souvent considérées comme étant la conséquence d'une limitation dans la durée de la saison de croissance, ainsi que dans la fertilité des sols alpins, plus que la conséquence de la température elle-même (KÖRNER, 2003).

L'adaptation locale observée sur ces traits pourrait donc être le résultat d'une *assimilation génétique* (WADDINGTON, 1959) de ces limitations dans les populations d'altitude. Lorsqu'une contrainte environnementale impose, sur plusieurs générations, une modification plastique du phénotype des individus, il est en effet possible que les individus dont le fond génétique leur permet de mieux se plier à ces contraintes (c'est à dire, souvent, à moindre coût) soient sélectionnés. Il en résulte qu'après une période d'évolution par sélection de ces individus, il existe une réponse phénotypique même en l'absence de contrainte phénotypique. On parle alors d'assimilation génétique (WADDINGTON, 1953 ; WADDINGTON, 1959 ; PIGLIUCCI et al., 2006), puisque tout se passe comme si un comportement plastique avait été "assimilé" de manière génétique<sup>1</sup>.

Des adaptations le long d'un gradient d'altitude et correspondant à ce syndrome phénotypique ont été trouvées chez au moins deux autres espèces alpines de Poacées : *Festuca eskia* (survie et reproduction, GONZALO-TURPIN & HAZARD, 2009) et *Poa alpina* (taille et reproduction, HAUTIER et al., 2009). Ces résultats convergents, dans une famille taxonomique différente, permettent de suggérer une certaine généralité de ce syndrome phénotypique. Notons que notre étude se distingue tout de même de ces deux autres de par l'amplitude du gradient altitudinal testé, ainsi que l'altitude maximale impliquée, représentant effectivement l'étage alpin, plutôt que subalpin.

---

<sup>1</sup>Ce qui ne constitue qu'une image, le mécanisme étant largement plus indirect.

### 1.3 Détection de gènes candidats

En utilisant des marqueurs issus d'un protocole de génotypage haut-débit, nous avons pu isoler 40 locus associés à des traits adaptatifs et montrant des signes de sélection, dont 13 étaient dans des régions géniques. L'homologie de ces 13 gènes avec le génome d'*Arabidopsis thaliana* laisse entrevoir la fonction de certains d'entre eux chez *A. alpina*.

À l'aune de ces informations, certains candidats semblent plus solides que d'autres. Étant donné que la production de fruits peut être fortement consommatrice de réserves chez la plante, le lien entre la longueur totale de fruits, l'altitude et le gène 62173, homologue du gène SS1 (impliqué dans la synthèse de l'amidon chez *A. thaliana*), est probablement le plus prometteur.

De manière intéressante, l'homologue du gène 1688 (associé à la hauteur et à l'amplitude de température) a été étudié par *knock-out* chez *A. thaliana*, et a été suggéré comme étant impliqué dans la sensibilité à la température chez cette plante (HSIEH, 2009).

Un autre gène présente un possible intérêt biologique important est le gène 12168, associé à hauteur reproductive et dont l'homologue chez *A. thaliana* a été montré, dans l'application d'un brevet, comme étant impliqué dans l'efficacité de l'utilisation de l'azote, en rapport avec la production de biomasse (« Use of CAD genes to increase nitrogen use efficiency and low nitrogen tolerance to a plant » 2015).

Deux autres gènes sont à mentionner. Premièrement, le gène 15932, le seul gène possiblement sélectionné et associé à la survie, dont l'homologue chez *A. thaliana* semble être impliqué dans le fonctionnement des trichomes. Cette liaison est tenue, non seulement parce que le rôle adaptatif des trichomes en conditions d'altitude n'est pas très claire (leur rôle classique étant plutôt de la protection contre l'herbivorie, LEVIN, 1973). Ensuite parce que le rôle de cet homologue, BCHA2, dans le fonctionnement du trichome est inconnu (WIENKOOP et al., 2004).

Deuxièmement, le gène 1190 semble appartenir à la famille des protéines à répétitions de motifs pentatricopeptide. Cette famille, nouvellement découverte chez *A. thaliana* (LURIN et al., 2004) semble principalement avoir un rôle dans la régulation post-transcriptionnelle dans chloroplaste et la mitochondrie (SCHMITZ-LINNEWEBER & SMALL, 2008). Malgré, encore une fois, des indications ténues, il est tout à fait envisageable que ce gène présente en effet une variabilité adaptative entre populations.

Le gène 17383, dont l'homologue est impliqué dans la résistance aux infections chez *A. thaliana* peut aussi constituer un candidat intéressant, mais pour des raisons différentes. Il est associé à la taille reproductive, dont on sait qu'elle a été influencée par la présence de la rouille blanche sur les plantes infectées. Or, la rouille a préférentiellement attaqué les populations de la région du Lautaret (LAU, GAL et PIC). Il se peut donc que ce gène soit lié à une certaine co-évolution entre les plantes de cette région et les pathogènes de cette région.

Il faut encore une fois rappeler ici que les locus qui ne sont pas à l'intérieur d'un gène ne sont pas nécessairement des faux positifs. Premièrement, il se peut qu'ils soient en réalité placés à l'intérieur d'un gène qui n'est pas encore annoté. Deuxièmement, il se peut qu'ils soient en déséquilibre de liaison avec un gène (annoté ou non) à proximité. Ou même, troisièmement, il se peut qu'en étant isolé de toute région génique, il possède une fonction régulatrice (région amplificatrice par exemple).

Il n'en reste pas moins qu'il faut garder à l'esprit que le taux de faux positifs de ces approches peut être assez important et que, compte-tenu des taux de fausses découvertes typiquement obtenus lors des études du CHAPITRE 1, il n'est pas impossible que cette liste soit composée, pour une bonne partie, de faux positifs et ce, malgré nos efforts pour limiter ce problème.

Bien sûr, compte-tenu de la puissance attendue de cette étude et de la densité des marqueurs utilisés, le risque de faux négatif est, lui aussi, extrêmement élevé.

## 2 Les obstacles méthodologiques liés à l'étude de l'adaptation locale

### 2.1 L'évolution : un phénomène stochastique

À travers tous les développements méthodologiques présents dans cette thèse, on voit qu'il existe une problématique commune : prendre en compte la nature stochastique des phénomènes évolutifs. Il s'agit là d'un problème fondamental, qui est, en quelque sorte, constitutif de toute réflexion évolutive. Comme l'ensemble des travaux présentés ici l'illustre, la biologie évolutive moderne traite le problème à l'aide du raisonnement statistique (SOBER, 2008). Pour être précis, le raisonnement basique consiste à considérer un modèle statistique, basé sur une hypothèse d'évolution neutre. Il s'agit donc d'abord de rejeter ce modèle avant d'accepter, par exemple, un modèle adaptatif.

Toute la difficulté de cette démarche est qu'il n'existe aucun modèle d'évolution neutre universel. L'évolution neutre est en effet influencée par un grand nombre de facteurs démographiques, historiques et écologiques. Les modèles "neutres" doivent donc être suffisamment flexibles pour prendre en compte ces facteurs. Cependant, s'il est possible (bien que parfois long et fastidieux) d'étudier les facteurs biologiques actuels des espèces d'intérêt qui peuvent influencer cette évolution neutre, il est beaucoup plus difficile d'obtenir des informations sur les événements historiques passés, en particulier aux échelles fines.

Devant la difficulté de réussir à élaborer et calibrer finement un modèle d'évolution neutre pour chaque espèce étudiée, nous avons régulièrement recouru à des modèles supposés "robustes" à une grande variété de situations. Bien que dans certains cas, la robustesse de ces modèles soit effectivement relativement bonne (par exemple, le modèle brownien utilisé pour les analyses comparatives phylogénétiques, GARAMSZEGI, 2014), il reste relativement difficile d'obtenir de tels modèles pour des échelles micro-évolutives dans le cadre de la génomique des populations. C'est ce que nous avons montré dans le CHAPITRE 1 et, dans une certaine mesure, dans le CHAPITRE 2.

### 2.2 Échelle et échantillonnage

Pour définir le concept d'adaptation locale dans l'Introduction générale, nous avons eu recours au concept de méta-population, que nous avons défini comme un ensemble de populations en contact par un flux de gène. Ce concept s'applique parfaitement bien aux données simulées dans les CHAPITRES 1 et 2. Dans nos simulations, en effet, l'ensemble des populations sont en contact (directement ou à travers d'autres populations), mais aucune n'est en contact avec une population extérieure : le système est parfaitement isolé. Dans les études empiriques, en revanche, échantillonner un système de populations totalement isolées est la plupart du temps virtuellement impossible.

Dans les modèles statistiques, la plupart du temps, l'hypothèse est faite que l'ensemble des populations échantillonnées constitue la vraie méta-population. Si cette hypothèse est nécessaire pour l'élaboration mathématique de ces modèles, elle peut entraîner des difficultés dans leur application. Par exemple, lorsque nous avons voulu tester BayeScEnv à l'aide des données génétiques humaines du HGDP, nous nous étions d'abord focalisés sur les 8 populations du Pakistan (avec donc, un fort gradient altitudinal). Or, les populations d'altitude, du Nord-Est du Pakistan sont en réalité en contact

avec d'autres populations d'Inde ou de Chine. Par conséquent, un grand nombre de marqueurs donnés comme positivement liés à l'altitude par notre méthode étaient en réalité tout à fait courant dans les plaines de Chine, par exemple. D'où notre décision d'élargir l'échantillonnage à toute l'Asie.

Outre ces problèmes de systèmes non isolés, l'échantillonnage des populations peut être optimisé pour la détection d'adaptation locale. Lorsque l'on échantillon le long d'un gradient environnemental, par exemple, il peut être intéressant de répliquer ce gradient (BUEHLER et al., 2014 ; RELSTAB et al., 2015). C'est ce que nous avons tenté de faire sur *A. alpina* en employant deux gradients, l'un dans le Vercors, l'autre autour du col du Lautaret. Malheureusement, cette configuration pose le problème que ces deux gradients ne se recouvrent pas, ce qui limite la comparaison entre les deux.

Une autre possibilité est d'échantillonner les populations par paires : deux populations proches géographiquement, donc avec un certain flux de gènes, mais dans des environnements contrastés (LOTTERHOS & WHITLOCK, 2015 ; RELSTAB et al., 2015). De cette façon, le flux de gène important entre chaque paire de populations est suffisamment important pour limiter les faux positifs, tout en permettant la détection d'une adaptation locale systématique sur les différentes paires.

Enfin, il est important de décider l'échelle à laquelle on veut détecter de l'adaptation "locale". Le patron d'adaptation que nous avons décrit chez *A. alpina* dans les massifs du Vercors et du Lautaret pourrait être bien différent à l'échelle entière des Alpes, ou même de l'Europe, par l'action d'autres facteurs environnementaux (par exemple les variations entre climat méditerranéen, continental, océanique ou arctique). Les patrons adaptatifs sont en effet très caractéristiques de l'échelle à laquelle ils sont étudiés (BRACHI et al., 2013).

### 2.3 Qu'est-ce qu'un trait phénotypique ?

La question de la mesure du phénotype peut paraître ancienne, résolue et même naïve, elle ne l'est pourtant pas. Bien qu'un exercice extrêmement courant en biologie, la mesure d'un trait phénotypique est toujours une action relativement complexe, qui demande une réflexion préalable. La mesure scientifique correspond en réalité, comme l'a montré LATOUR (1996, p.171–225), en un *alignement de références* (conceptuelles) consécutives, dont le but est de permettre à la fois la conservation de certaines propriétés des objets mesurés, sans en être une copie exhaustive. Il s'agit donc d'une part de réduire la complexité de l'organisme vivant à une mesure (ou un ensemble restreint de mesures), et d'autre part d'en extraire une information biologiquement pertinente. Cette information, chiffrée, mathématisée, pourra ensuite permettre la comparaison entre individus, entre populations ou entre espèces.

L'aller-retour le long de cet alignement de références permet finalement au biologiste d'étudier les propriétés des êtres vivants en analysant des nombres, des cartes ou des modèles qui seront commensurables entre différentes études. En suivant ce modèle épistémologique, la question de la mesure d'un trait phénotypique devient celle de la conservation de cette information biologique, de cet "alignement des références".

Ces réflexions concernent évidemment divers champs de la biologie, mais elles possèdent un impact fort sur une grande partie des analyses présentes dans cette thèse. Prenons l'exemple de la phénologie de la floraison analysée dans l'ARTICLE 5. Pour étudier ce trait, nous sommes montés dans chacun des jardins toutes les semaines pour enregistrer les nouvelles floraisons. La mesure enregistrée est donc la date à laquelle nous sommes venus et avons constaté l'ouverture de la première fleur sur chaque plante. Nous avons cependant décidé que le vrai trait phénotypique, celui que nous voulions étudier, n'était pas la *date* de floraison, mais le *temps* écoulé entre la fonte des neiges et

cette première fleur. Il se trouve que ce temps floraison n'était pas plastique entre les deux jardins. Or, la fonte des neiges au Lautaret étant 2 mois plus tardive qu'au Vercors, la date, elle, est nécessairement plastique. Se pose donc la question suivante : quelle mesure reflète le mieux la phénologie de la floraison, la date ou le temps ? Par conséquent, cette phénologie est-elle plastique ou non ?

En réalité, il n'existe pas réellement de réponse aux questions ci-dessus. Tout dépend de ce que l'on veut étudier et du processus biologique à l'œuvre en réalité. Ici, le temps de floraison n'est pas plastique justement *parce que* la date de floraison est plastique. D'une manière ou d'une autre (par exemple à travers la température ambiante), *A. alpina* s'ajuste à la date de la fonte des neiges dans nos jardins expérimentaux. Il est fort probable que ce signal soit lié à une accumulation de degrés-jours plutôt qu'à la fonte des neiges. Sur le terrain, en effet, les dates de floraisons dans les populations GAL et PIC sont vaguement comparables<sup>1</sup>, alors que la fonte des neiges est largement plus tardive à GAL qu'à PIC.

On voit, à travers cet exemple, que la définition et la mesure d'un trait phénotypique influent largement sur ce que ce trait peut nous dire d'un point de vue biologique.

Un autre exemple de mesure phénotypique intéressante est celle de l'aire. En réalité, les mesures réellement effectuées sont celles des deux diamètres qui servent ensuite au calcul de l'aire. Pour autant, ces deux diamètres constituent déjà une modélisation de la plante : il s'agit de deux diamètres orthogonaux dont l'un tente de mesurer l'élongation maximale de la partie végétative de la plante<sup>2</sup>. En quelque sorte, ce trait constitue un compromis entre la diversité presque incommensurable de tailles et de formes des individus et une possible comparaison d'une seule de leurs caractéristiques : la surface au sol qu'ils occupent. Un argument pratique est en jeu ici aussi : il serait possible, par exemple, de photographier chaque individu par dessus, puis d'analyser, de manière informatique, l'aire (projetée au sol) qu'ils recouvrent. Si cette manière de faire pourrait paraître plus "réaliste", elle n'en demande pas moins, paradoxalement, un équipement technologique largement plus complexe d'une règle. En outre, on peut se poser la question du gain réel en objectivité : à quel moment de la saison prendre la photographie ? Quel algorithme utiliser pour détecter l'aire ? Doit-on calibrer cet algorithme pour cibler la partie "végétative" de la plante ?

Un tout dernier exemple illustre les problèmes qui résident dans l'analyse statistique du trait en question plutôt que dans sa mesure elle-même. Prenons la croissance : nous avons souvent analysé le logarithme de la croissance, plutôt que la croissance elle-même directement. Cette décision était motivée par le fait que la distribution du logarithme permettait d'obtenir des résidus normalement distribués dans un modèle linéaire. Il s'agit donc là d'un impératif pratique, mais qui peut avoir de fortes répercussions sur l'interprétation des analyses. Que dire si, à la place d'une transformation logarithmique, nous avons choisi, comme cela se fait parfois, une transformation "racine-carré" et que l'effet ne soit soudainement plus significatif ? Quel serait alors la "bonne" transformation ?

L'ensemble de ces interrogations reflètent un fait très important en biologie : mesurer un phénotype, c'est déjà commencer à analyser le système et donc commencer à faire des choix (conceptuels, pratiques et méthodologiques). Cependant, ces considérations ne doivent pas mener à une conception relativiste ou nihiliste de la science. Le fait est que si les variations biologiques sont suffisamment fortes ou claires, les détails dans la mesure et l'analyse des phénotypes n'aura qu'un impact limité dans la réussite et la pertinence de l'analyse. Par contre, cela peut s'avérer être un problème lorsque

---

<sup>1</sup>Ceci n'a pas été quantifié rigoureusement et ne vient que d'une impression recueillie sur le terrain.

<sup>2</sup>Le nombre de concepts mathématiques et biologiques implicites dans cette phrase illustre parfaitement le propos : mesurer un trait, c'est déjà le modéliser.



l'on manque de puissance, qu'un effet est très marginal ou qu'un effet est fortement non linéaire. Dans tous ces cas, la mesure et l'analyse du trait phénotypique devront être rigoureusement réfléchies pour éviter de rompre, par inadvertance, ce fameux "alignement des références" qui permet de parler de biologie en manipulant des nombres.

## 2.4 Le lien phénotype-génotype

Outre le défi de définir, mesurer et analyser un trait phénotypique possédant un sens biologique, la biologie évolutive est confrontée à un autre problème permanent : à part pour une toute petite fraction de traits phénotypiques à déterminisme simple, on ne sait quasiment rien de la forme du lien entre phénotype et génotype. Or la forme de ce lien joue un rôle extrêmement important dans les mécanismes évolutifs. Comme expliqué dans la conclusion de l'ARTICLE 3, les phénomènes évolutifs comme l'adaptation locale font intervenir trois types de facteurs : l'environnement, le phénotype et le génotype. En gros, l'environnement (au sens large) exerce des pressions sélectives sur les individus, entraînant une reproduction différentielle entre individus, en partie déterminée par leur propre phénotype. Cette reproduction différentielle entraîne une évolution des fréquences alléliques de certains locus impliqués dans la variabilité des traits phénotypiques sélectionnés.

L'environnement exerce donc un effet sur l'évolution du génotype des individus par le biais de leur phénotype. On voit ici que le lien entre phénotype et génotype est un élément essentiel à la compréhension des phénomènes micro-évolutifs. Pourtant, à cause de l'immense complexité des interactions entre les produits de différents gènes, non seulement à l'échelle d'une cellule, mais encore plus à l'échelle d'un organisme multicellulaire, il est virtuellement impossible d'appréhender concrètement ce lien.

Bien que des travaux sur ce sujet existent, par exemple utilisant la théorie métabolique (FIÉVET et al., 2009) ou des modèles de développement non linéaires (MORRISSEY, 2015) la biologie évolutive moderne repose essentiellement sur des modèles supposant un lien linéaire entre le génotype et le phénotype (dont le modèle infinitésimal de FISHER, 1918). Or ce lien linéaire n'est pas toujours une hypothèse plausible, particulièrement lorsque les traits phénotypiques observés possèdent une distribution bornée (voir ARTICLE 6), ce qui est le cas de la majeure partie des traits étudiés ici.

En outre, si ce lien peut être linéaire pour un trait mesuré d'une certaine manière, n'importe quelle transformation non linéaire romprait cette linéarité. Revenons par exemple sur la croissance : le lien génotype-phénotype est-il linéaire avec la croissance mesurée comme un rapport entre deux tailles, comme le logarithme de ce rapport ou comme une différence entre deux tailles ? Clairement, le lien ne peut pas être linéaire avec les trois mesures différentes, bien que toutes soient légitimes.

## 3 Perspectives méthodologiques

### 3.1 Distinguer entre adaptation locale et plasticité phénotypique *in situ*

Grâce à l'expérience de jardin commun, il est possible de lisser l'effet de la plasticité phénotypique sur les populations étudiées, afin d'analyser la composante génétique de la différenciation phénotypique entre populations. Par définition, cette composante génétique doit être constante quel que soit l'environnement<sup>1</sup>. La plasticité phénotypique, au contraire, est extrêmement sensible aux conditions

---

<sup>1</sup>Une des raisons pour lesquelles il est intéressant de répliquer une expérience de jardin commun, outre l'estimation de la plasticité phénotypique et du G×E, est d'assurer que la composante génétique mesurée n'est pas spécifique au seul environnement mesuré.

environnementales dans lesquelles elle est mesurée.

Par conséquent, la plasticité phénotypique mesurée en jardin commun ne peut se rapporter qu'aux conditions expérimentales. Ceci est donc peu informatif sur l'influence de la plasticité phénotypique en conditions naturelles. En effet, les conditions expérimentales sont toujours beaucoup plus "douces" (meilleure disponibilité en ressources, moindre compétition, etc...) que les conditions naturelles. Si la plasticité phénotypique possède un coût (DEWITT et al., 1998), il est par exemple attendu qu'elle possède un effet plus fort dans des conditions expérimentales plutôt que dans des conditions naturelles.

Comment alors estimer la contribution relative de la génétique (adaptation locale et dérive) et de la plasticité phénotypique à l'adéquation phénotypique des individus d'une population aux conditions naturelles locales ? La solution vient du fait exposé ci-dessus : la composante génétique, due à l'adaptation locale ne devrait pas, par définition, varier avec l'environnement. En théorie, il devrait donc être possible en combinant des analyses *in situ* à une expérience de jardin commun (répliquée dans différents environnements pour prendre en compte l'interaction  $G \times E$ ) d'estimer ces contributions relatives.

Il faudrait estimer la composante génétique de la variabilité phénotypique entre populations à l'aide des expériences de jardin commun, puis retrancher cette composante à la variance phénotypique observée entre populations naturelles. Pour des raisons évidentes de commensurabilité, il faudrait que les traits phénotypiques soient mesurés de manière identique entre les populations naturelles et les individus du jardin commun. Il faudrait, en outre, des outils statistiques appropriés pour combiner les deux analyses en prenant en compte les incertitudes sur l'estimation, par exemple, de la composante génétique. Enfin, séparer les contributions relatives entre adaptation locale et dérive pourrait être possible, mais demanderait quelques développements théoriques et méthodologiques.

Dans le cadre de cette thèse, il n'est pas tout à fait sûr qu'on puisse mesurer ces contributions relatives en milieu naturel, étant donné que bien que des efforts aient été fait en ce sens, tous les phénotypes ne sont pas tout à fait comparables entre les deux études (conditions naturelles et jardin expérimental). En effet, les phénotypes mesurés dans les deux études ne sont pas tout à fait comparables (bien qu'ils renvoient des signaux similaires, FIG. 9). Par exemple, le nombre total de tige est absent dans les expériences de jardin commun et l'aire est manquante en milieu naturel. Pour finir, seulement deux réplicats de jardin commun ne permettent pas tout à fait d'estimer correctement l'interaction  $G \times E$ .

### 3.2 Étude de l'architecture des traits phénotypiques adaptatifs

Dans la section 2.4, nous avons abordé le problème du lien phénotype-génotype. Une des incertitudes concernant ce lien porte sur le nombre de gènes impliqués dans la variabilité des caractères phénotypiques, ainsi que sur la distribution de leurs effets (ROCKMAN, 2012). L'applicabilité du modèle infinitésimal de FISHER (1918) à une vaste majorité de traits phénotypiques, nous renseigne assez mal sur le nombre de gènes effectivement impliqués dans la variabilité de ceux-ci, étant donné qu'il s'agit d'une bonne approximation même quand le nombre de gènes impliqués est modeste (ROFF, 2007).

En outre, il existe actuellement une controverse sur la distribution des effets de ces gènes. D'un côté, une distribution "en L" des effets des gènes (peu de gènes à forts effets, de nombreux gènes à faibles effets) est régulièrement observée (T. F. C. MACKAY, 2001). De l'autre côté, ce résultat est soupçonné d'être un artefact des méthodes utilisées (BEAVIS, 1994 ; ROCKMAN, 2012).

En réalité, il est fort probable que l'architecture génétique des traits phénotypiques évolue elle-même sous l'effet de la sélection. La caractéristique finale de l'architecture des traits phénotypiques adaptatifs peut ainsi dépendre de la force de la sélection (RAJON & PLOTKIN, 2013) ou de la force du flux de gènes (YEAMAN & WHITLOCK, 2011).

Cependant, à cause des limites des méthodes de détection de locus impliqués dans la variabilité d'un phénotype, il est difficile de fournir des éléments de preuve empiriques à ce débat. Les méthodes d'association décrites dans cette thèse ne permettent pas, en toute rigueur, d'inférer la distribution des effets des gènes (seulement de détecter ceux dont l'effet est suffisamment fort pour être significatif). Cependant, certaines des dernières méthodes disponibles (par exemple BayesR, MOSER et al., 2015) permettent de s'approcher de cette inférence. Un travail méthodologique utilisant les données phénotypiques et génomiques typiques des études d'association, mais axée sur la distribution des effets des locus, plutôt que sur leur détection reste encore à développer.

## 4 Conclusion générale

Les travaux présentés dans cette thèse s'adressent principalement à l'étude des phénomènes adaptatifs aux échelles micro-évolutives, avec un accent très particulier sur l'adaptation locale. Ces travaux portent d'un côté sur des réflexions méthodologiques autour de l'étude de l'adaptation locale et d'un autre côté sur l'étude empirique de l'adaptation d'*Arabis alpina* aux conditions d'altitude.

Les trois chapitres de la première partie nous ont permis d'explorer les enjeux issus de l'arrivée de la génomique et du génotype haut-débit au sein des études de micro-évolution.

Dans le CHAPITRE 1, nous avons vu que les méthodes de scan génomiques pour détecter la sélection nécessitent des précautions quant à leur utilisation. Tout d'abord, l'ampleur des tests multiples en jeu dans ce type d'analyse nécessite d'utiliser un contrôleur de l'erreur approprié : c'est à dire contrôler le taux de fausse découverte plutôt que le taux de faux positif classique. Ensuite, il est nécessaire de prendre en compte que les modèles "neutres" des nombreuses méthodes disponibles possèdent leurs forces et leurs faiblesses. Aucun modèle universel d'évolution neutre ne permettra jamais de prendre en compte toutes les situations. Enfin, il est probable que, face à des jeux de données compliqués (sélection faible, démographies complexes), comme ceux que nous avons simulés, ces méthodes renvoient, dans l'ensemble, de nombreux faux positifs.

Dans le CHAPITRE 2, nous avons présenté le développement d'une méthode, BayeScEnv, basée sur la notion de  $F_{ST}$ , mais utilisant une information environnementale. En la développant, nous avons souhaité mettre l'accent sur la réduction du taux de faux positif de notre nouvelle méthode. Des études par simulation montrent que notre méthode permet de détecter des signaux d'adaptation locale, tout en réduisant de manière importante le taux de fausse découverte. Par ailleurs, notre méthode permet de détecter des patrons adaptatifs qui ne seront pas détectés par d'autres méthodes.

Dans le CHAPITRE 3 enfin, nous avons montré les synergies possibles entre les expériences de jardin commun, le génotypage haut-débit et certains développements statistiques modernes. Nous avons montré que ces synergies peuvent permettre d'ouvrir la porte à de nouvelles approches combinant génétique, phénotype et environnement.

Les deux chapitres de la deuxième partie portent plus particulièrement sur l'étude de la plante alpine *Arabis alpina* et de son adaptation aux conditions d'altitude. Les études de cette partie portent sur six populations réparties entre 900m et 3000m d'altitude, et entre deux massifs : le Vercors d'une part, la région autour du col du Lautaret, d'autre part.

Dans le CHAPITRE 4, nous avons étudié certaines caractéristiques phénotypiques et démographiques de ces populations. Nous avons pu montrer qu'*A. alpina* était une espèce à espérance de vie assez courte, et dont la survie et la reproduction était extrêmement variable entre individus et entre populations. En outre, nous avons pu montrer que la croissance et la survie étaient, entre autre, influencées par la température moyenne. Tandis que la croissance augmente avec la température, la survie diminue. La population la plus froide possède, de plus, une reproduction significativement diminuée.

Dans le CHAPITRE 5, la composante génétique de ces variations a été démontrée à l'aide d'une expérience de jardin commun. La croissance et la reproduction, mais aussi la taille et le port des plantes, varient génétiquement avec la température moyenne. De manière adaptative, donc, les plantes des milieux froids croissent moins, se reproduisent moins et sont plus petites et compactes. À l'aide de données de génotypage haut-débit, nous avons proposé, en outre, une série de gènes candidats impliqués dans ces adaptations.

Ces résultats sur *A. alpina* montrent l'existence d'un syndrome phénotypique variant avec l'altitude. Ce syndrome est adaptatif et est vraisemblablement lié, en ce qui concerne la croissance et la reproduction au moins, à un phénomène d'assimilation génétique lié aux contraintes posées par la faible disponibilité en temps et en ressources dans les milieux alpins. Il est très probable que cette adaptation locale soit, entre autres choses, une caractéristique clé de cette espèce lui permettant de vivre dans une gamme altitudinale remarquablement large.

Une caractéristique forte des travaux présentés ici est donc cette complémentarité entre réflexions méthodologiques, empiriques et biologiques. La nécessité de la réflexion méthodologique résulte de la complexité des phénomènes évolutifs décrits tout au long de cette thèse, et plus particulièrement en Introduction et Discussion. La maîtrise à la fois des concepts évolutifs, biologiques, écologiques, statistico-mathématiques et moléculaires nécessaire à l'étude, par exemple, de l'adaptation locale, font de la biologie évolutive une discipline fondamentalement ouverte, où la collaboration doit être forte entre des compétences complémentaires pour appréhender les phénomènes dans toute leur complexité. Ces collaborations sont d'autant plus fluides qu'il existe des interfaces, c'est à dire des espaces où les différentes disciplines arrivent à se parler, à s'articuler. Cette thèse et l'ensemble des travaux qui la constituent est, me semble-t-il, l'un de ces espaces.

Pour conclure, il est clair, d'après les travaux présenté dans cette thèse et leur discussion, que malgré l'âge relativement avancé de la biologie évolutive (nous avons récemment fêté les 150 ans de la publication de *L'Origine des Espèces*), il existe toujours une dynamique forte d'avancées théoriques, méthodologiques et empiriques. Les influences modernes les plus fortes, constatées au moins dans mes travaux, viennent des avancées du séquençage haut-débit d'une part et des avancées statistiques d'autre part<sup>1</sup>. Ces deux avancées techniques permettent une étude de plus en plus fine, ou à tout le moins renouvelée, des phénomènes adaptatifs, notamment dans la relation complexe entre le phénotype, le génotype et l'environnement. Les progrès, notamment en termes de génotypage, sont tels qu'il est probable que dans un avenir proche, la limite soit posée davantage sur le phénotypage d'un grand nombre d'individus, plutôt que sur leur génotypage.

---

<sup>1</sup>Notons que ces deux domaines doivent beaucoup à la révolution technologique induite par l'informatique et la croissance constante de la puissance de calcul des ordinateurs.

# **ANNEXES**



# ANNEXE S1

## Supplementary Information (ARTICLE 1)

---

### S1.1 Population structure scenarios and migration

Below (Fig. S1.1) are the schematic representation of the demographic scenarios used for the simulation study, showing the fission events and the migration between populations. Note that only some illustrative migration combinations are represented for the sake of simplicity. Also recall that the total duration of simulation for HsIMM is 500 generations, whereas it is of 400 generations for the IMM and SS scenarios.

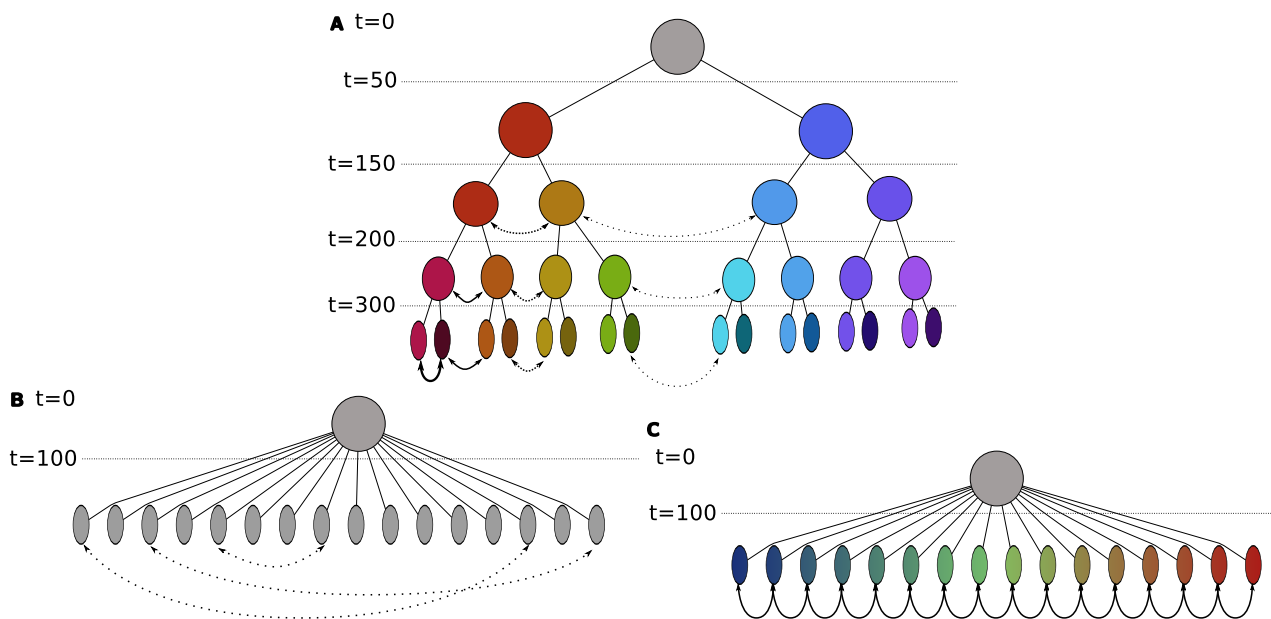
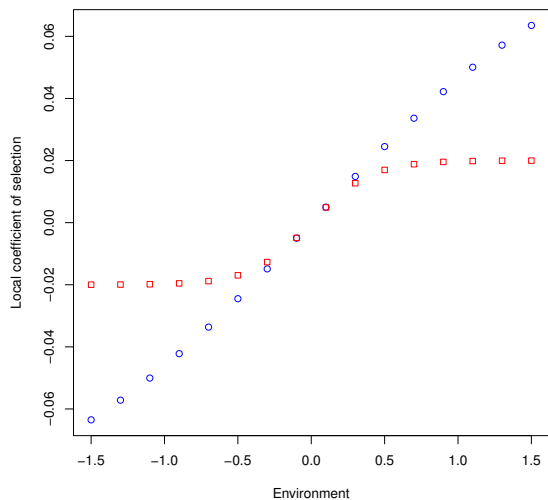


FIG. S1.1 : Schematic representation of the different demographic scenarios. A : Hierarchically structured isolation with migration (HsIMM); B : Isolation with migration (IMM); C : Stepping-stone (SS). “Historical” relationship between populations are represented using connectors, and migration with double-headed curves.





**FIG. S1.2 :** Local coefficient of selection against local environment. Blue circles : for monogenic scenarios; red squares : for polygenic scenarios.

## S1.2 Selective gradient

The local coefficient of selection for each local value of environment is shown below in Fig. S1.2. Note that although the coefficient seem to be smaller for the polygenic case, it is actually multiplied along 50 loci. Indeed, dividing the coefficient of selection by the number of loci made the selective signal totally undetectable.

## S1.3 Bayescan and BayEnv MCMC parameters

Bayescan was run for 50,000 iterations with a thinning interval of 10. The burn-in was of 5,000 iterations. We ran 20 pilot studies each of length 5,000 iterations. BayEnv was run for 15,000 iterations during its first MCMC step, and 30,000 in its second MCMC step. Note that we did not used a defined value for the burn-in and thinning interval, because the binary used does not allow for setting up these values.

## S1.4 False Discovery Rate and its relation to other test characteristics

The False Discovery Rate (FDR) is defined as the proportion of false positive among the significant loci (i.e. true and false positives) :

$$FDR = \frac{FP}{FP + TP} \tag{S1.4}$$

whereas the False Positive Rate (FPR) is defined as the proportion of false positive among the non selected loci :

$$FPR = \frac{FP}{FP + TN} \tag{S1.5}$$

The power (here POW) is defined as the proportion of true positives among the truly selected loci :

$$POW = \frac{TP}{TP + FN} \tag{S1.6}$$

If we denote  $r$  the ratio between the number of truly selected loci  $S$  and the truly non selected loci  $NS$  ( $r = S/NS$ ), then we have the following relationship between the three test characteristics :

$$FDR = \frac{FPR}{FPR + rPOW} \tag{S1.7}$$



Note that the FDR depends on both the false positive rate and the power : a near zero power will mechanically make the FDR close to one, which is sensible. Also, a  $r$  close to zero (i.e. very few truly selected loci among all the tested loci) will automatically produce a FDR near one. This reflect the fact that if selected loci are very few, or if they are hard to detect (or even both), then most of the significant loci must be seen as suspect, since they are very likely to be false positives.

Thus, it is important to distinguish between the false positive rate (FPR) which measures the error due to long-run sampling of tests (i.e. multiple testing on non selected loci) and the false discovery rate (FDR) which measures the error due to too much confidence in saying that a significant result is indeed a true one. We advocate, along with STOREY & TIBSHIRANI (2003), that the FDR is indeed the error rate we want to control for while performing genome-wide tests.

## S1.5 Transformation of Bayes Factor (BF) into $q$ -values

In order to transform the BF yielded by BayEnv into  $q$ -values, we used the same algorithm as the one implemented in Bayescan (MULLER et al., 2006). However this algorithm need posterior probabilities  $P(H_1|Data)$  instead of BF.

The two quantities are related like :

$$\frac{P(H_1|Data)}{P(H_0|Data)} = \text{BF} \frac{\pi_1}{\pi_0} \quad (\text{S1.8})$$

where  $\pi_1$  and  $\pi_0$  are the prior probabilities for  $H_1$  and  $H_0$  respectively. Recall that  $H_0 = \{\beta = 0\}$  and  $H_1 = \{\beta \neq 0\}$ , which means that, in the universe of the tested models, the two hypothesis are complementary. We thus deduce that :

$$\pi_0 = 1 - \pi_1 \quad (\text{S1.9})$$

and :

$$P(H_0|Data) = 1 - P(H_1|Data) \quad (\text{S1.10})$$

Reinjecting these results into EQ. S1.8 leads to :

$$P(H_1|Data) = \frac{1}{1 + \frac{1 - \pi_1}{\text{BF} \pi_1}} \quad (\text{S1.11})$$

Obviously, one needs to define the prior probability of  $H_1$  beforehand. In our analysis, we used the same value for BayEnv as the one we used for Bayescan :  $\pi_1 = 0.01$ . Now, the  $q$ -values are estimated using the following algorithm MULLER et al., 2006, PP stands for posterior probability :

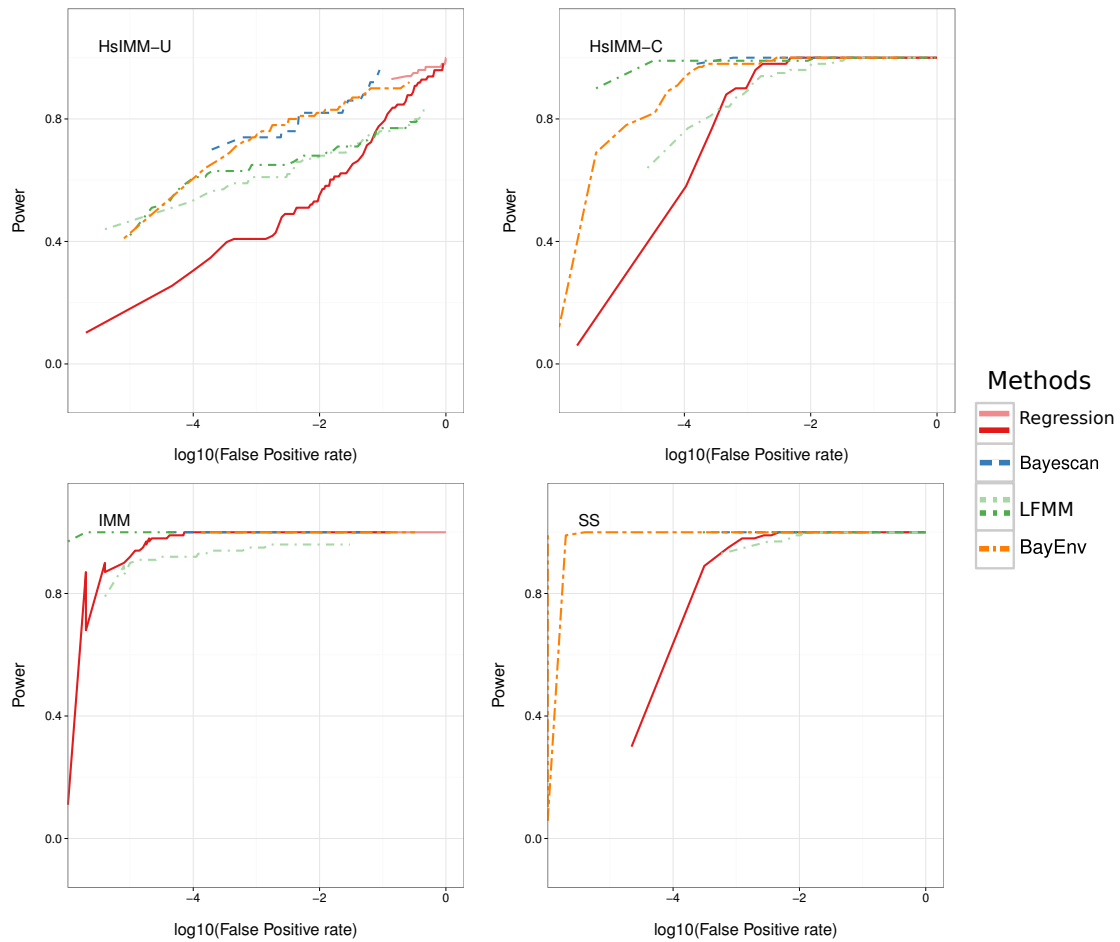
```

For each SNP i :
Do
  Set qval<-0
  Set number_of_significant<-0
  For each SNP j :
  Do
    If PP(j) is greater than PP(i)
      qval <- qval + (1-PP(j))
      number_of_significant++
  Done
  qval<-qval/number_of_significant
Done

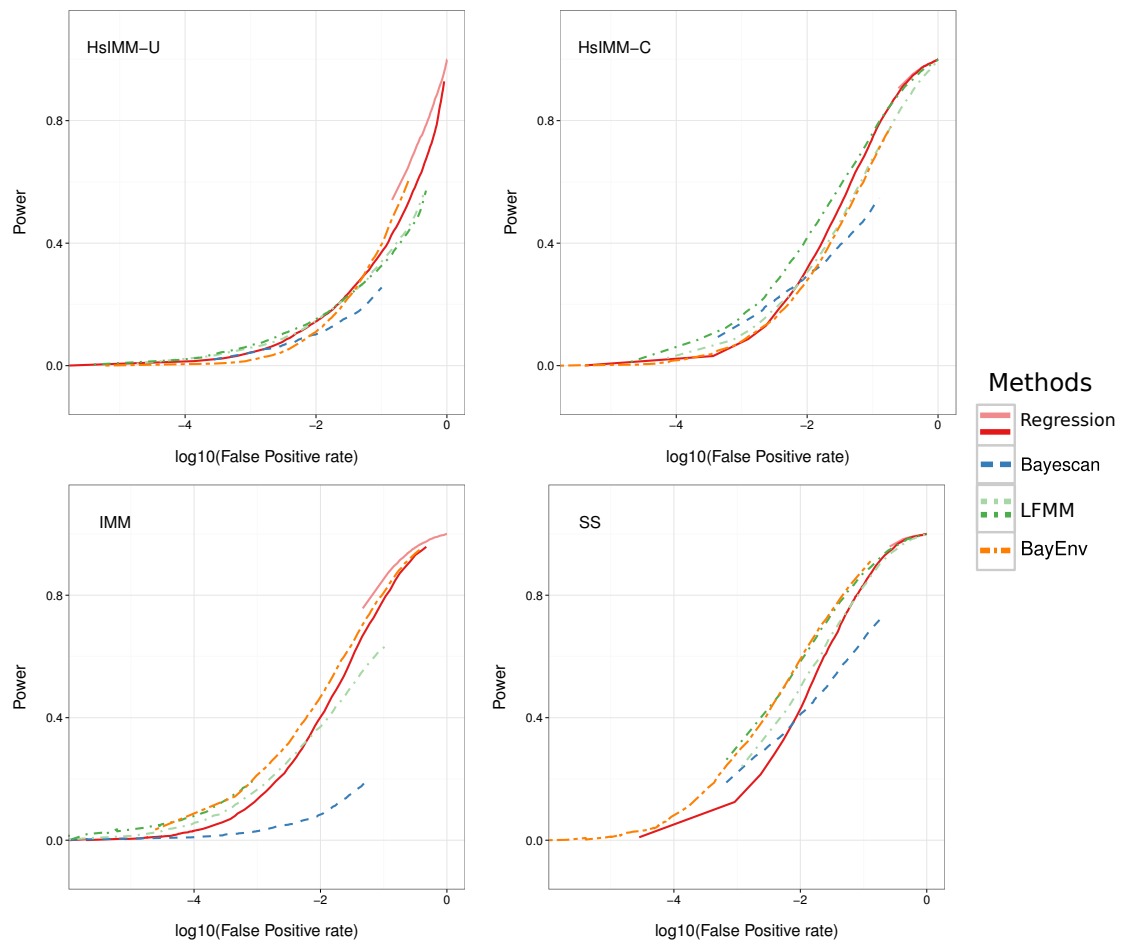
```

## S1.6 Power against False Positive Rate (ROC curves)

Below are given the power against false positive rates curves (a.k.a. ROC curves) for the different scenarios, in the monogenic (Fig. S1.3) and in the polygenic case (Fig. S1.4). Note that, whereas the methods can be quite easily distinguished for the former (Fig. S1.3, the regression in particular has a poor compromise between power and false positives), all methods are quite similar in the latter case (Fig. S1.4, with the notable exception of Bayescan in the IMM scenario).



**FIG. S1.3 :** Statistical power against false positive rate (ROC curve) for monogenic selection. Lines are for linear regression (plain red line), Latent Factor Mixed Model (LFMM, green dot-dashed line), Bayescan (blue dashed line) and BayEnv (orange two-dashed line). Light lines are for individual genotype data specification for the linear regression (light red) and LFMM (light green). Note that some lines (e.g. light red) can be hidden by the others, especially in the top-right corner.



**FIG. S1.4 :** Statistical power against false positive rate (ROC curve) for polygenic selection. Lines are for linear regression (plain red line), Latent Factor Mixed Model (LFMM, green dot-dashed line), Bayescan (blue dashed line) and BayEnv (orange two-dashed line). Light lines are for individual genotype data specification for the linear regression (light red) and LFMM (light green). Note that some lines (e.g. light red) can be hidden by the others, especially in the top-right corner.

## S1.7 False Positive Rate against ( $q$ -value) significant threshold

Below (Fig. S1.5) given the false positive rate (FPR) against the significance threshold ( $\alpha$  : In terms of  $q$ -values. We thus have no expectation regarding the false positive rate and the significance threshold, as we can have when using  $p$ -values). Note the strong similarity between this graph, and the Fig. A1.4 of the main text. Although the FDR is not solely explained by an inflation of false positives, it is clearly the main component of this inflation. The graphs are only provided for the polygenic selection case, since the FPR is not much influenced by the selection structure.

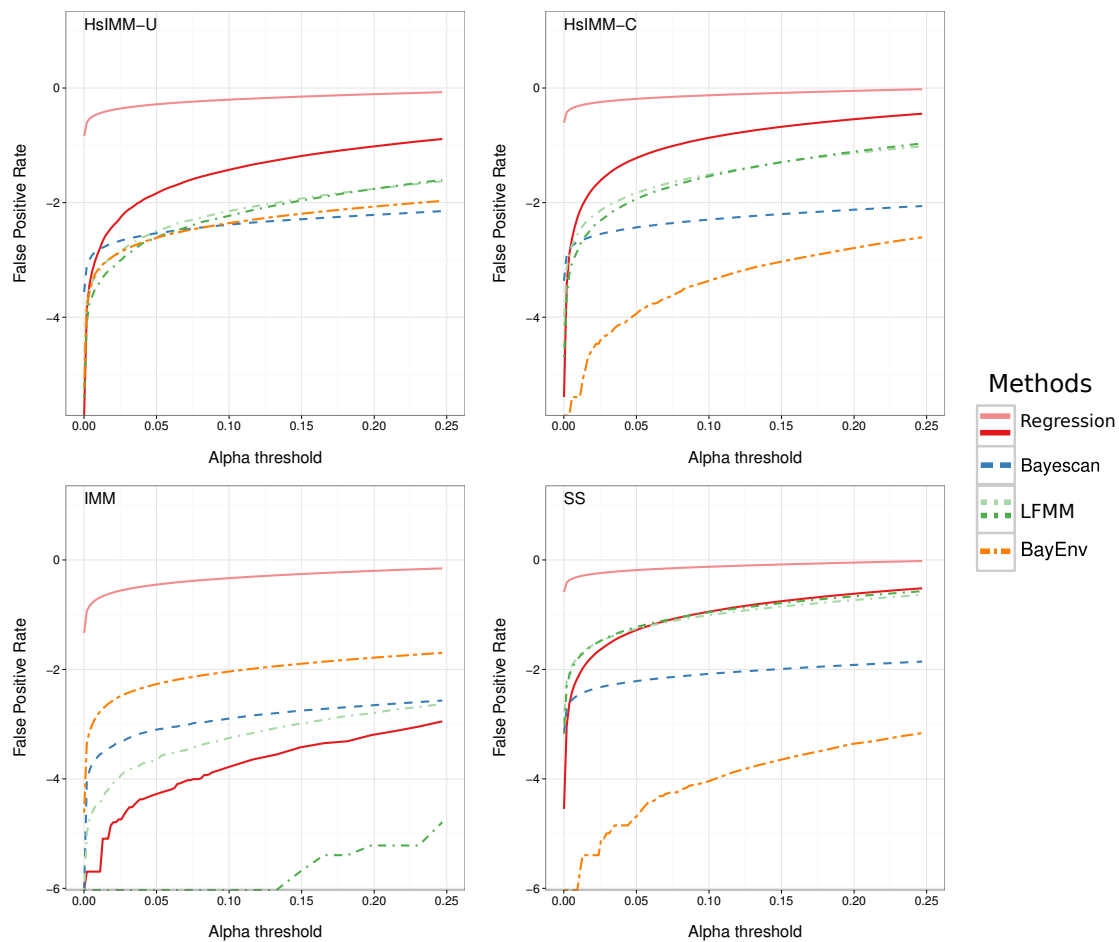


FIG. S1.5 : False positive rate (in log scale) against significance threshold for polygenic scenarios. Lines are linear regression (plain red line), Latent Factor Mixed Model (LFMM, green dot-dashed line), Bayescan (blue dashed line) and BayEnv (orange two-dashed line). Light lines are for individual genotype data specification for the linear regression (light red) and LFMM (light green).

## S1.8 Power against False Discovery Rate (FDR)

The Fig. S1.6 (monogenic case) and S1.7 (polygenic case) show the power yielded by the different methods of estimation, according to their actual FDR. In the monogenic selection case (Fig. S1.6) we can see that, while all methods perform quite as good for the scenario HsIMM-C, the “recent” methods (Bayescan, BayEnv and LFMM) perform better than the regression for the scenario HsIMM-U. In the polygenic case (Fig. S1.7), and as explained in the main text, the power yielded by LFMM (using its allele frequencies data specification) for a fixed FDR is better than the other methods in all cases. Bayescan yields quite powerful results for low FDR in HsIMM-C (Fig. S1.7), although BayEnv becomes the second best for more classical scenarios (Fig. S1.7, IMM and SS). For the hierarchical models (Fig. S1.7, HsIMM-U and HsIMM-C), the regression performances are below LFMM, but it is still a strong competitor against BayEnv and Bayescan. Note that a better power for a fixed FDR only means that more SNPs were considered as positive.

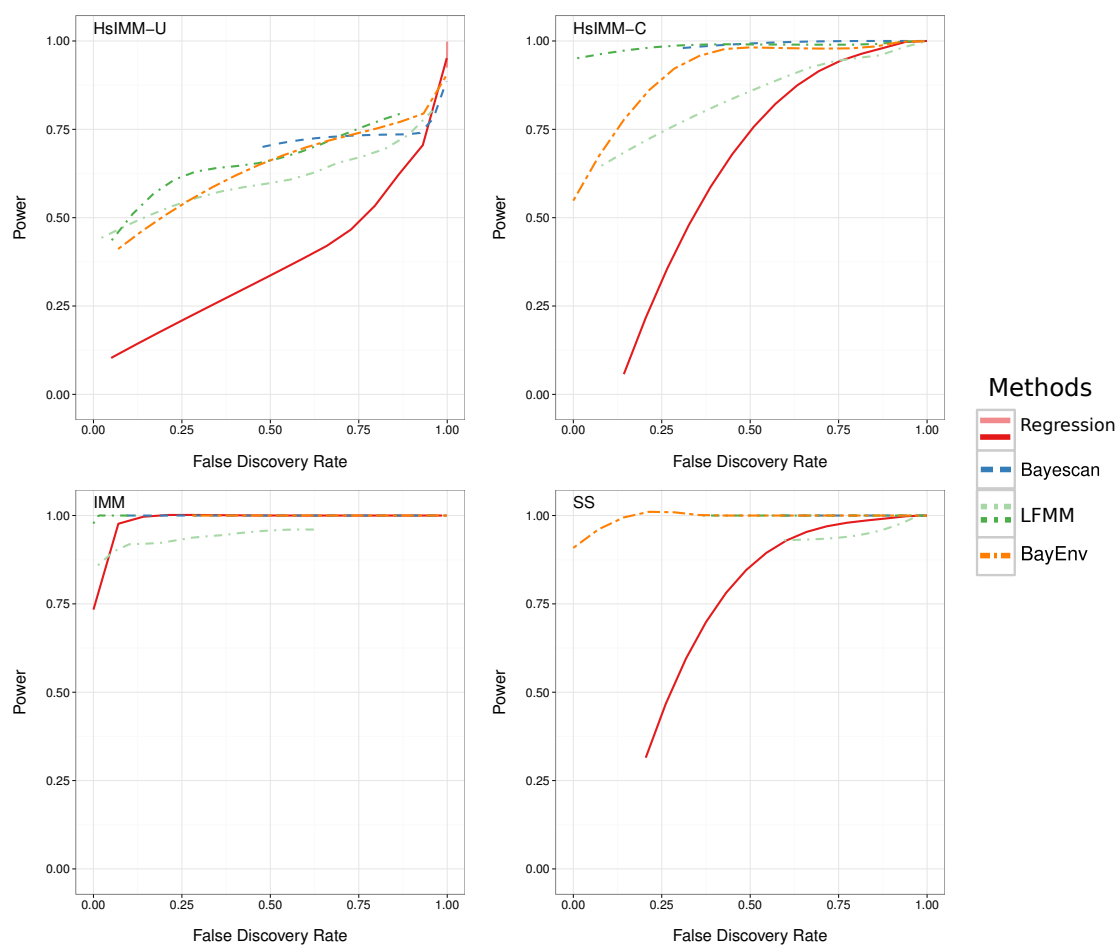
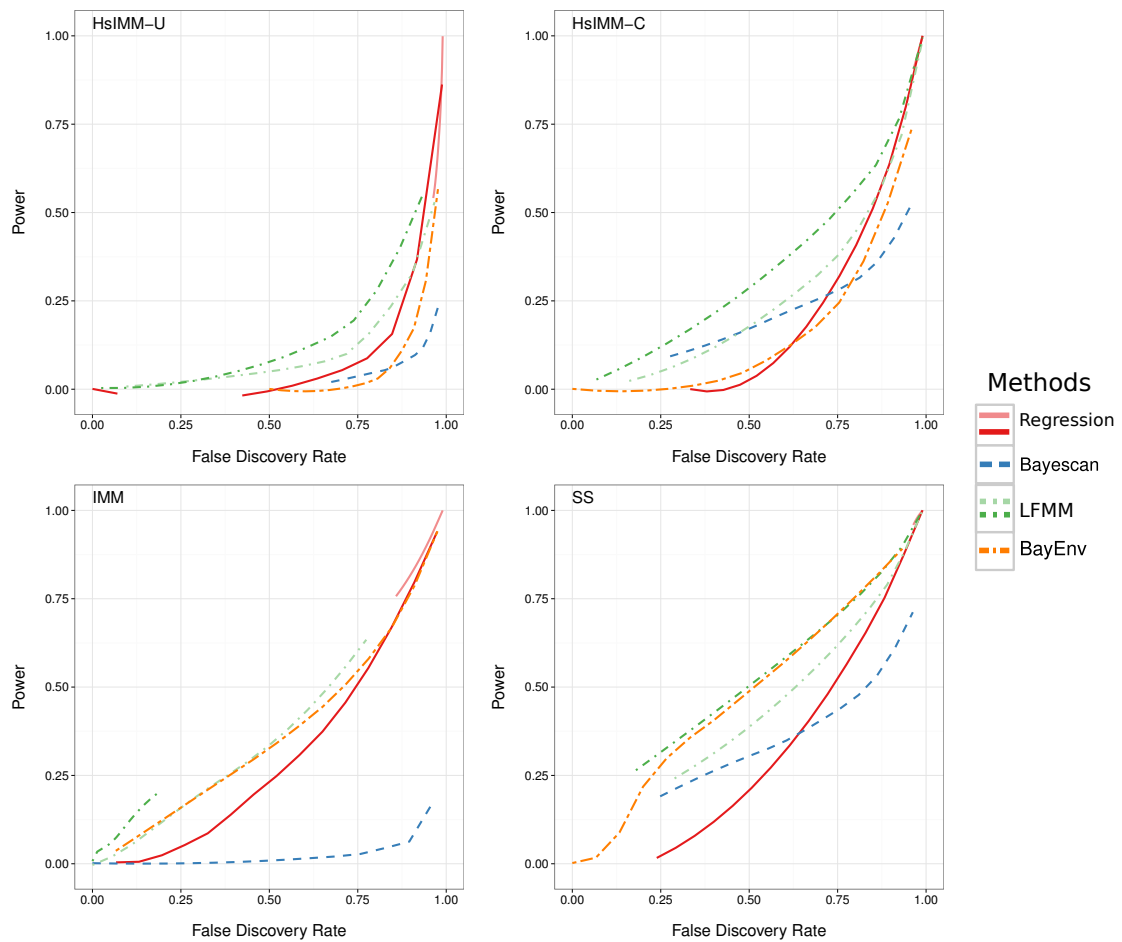


FIG. S1.6 : Statistical power against actual false discovery rate for monogenic selection. Lines are LOESS smooth for linear regression (plain red line), Latent Factor Mixed Model (LFMM, green dot-dashed line), Bayescan (blue dashed line) and BayEnv (orange two-dashed line). Light lines are for individual genotype data specification for the linear regression (light red) and LFMM (light green). Note that some lines (e.g. light red) can be hidden by the others, especially in the top-right corner.



**FIG. S1.7 :** Statistical power against actual false discovery rate for polygenic selection. Lines are for linear regression (plain red line), Latent Factor Mixed Model (LFMM, green dot-dashed line), Bayescan (blue dashed line) and BayEnv (orange two-dashed line). Light lines are for individual genotype data specification for the linear regression (light red) and LFMM (light green). Note that some lines (e.g. light red) can be hidden by the others, especially in the top-right corner.

## S1.9 Spurious power against false positive rate (FPR)

Fig. S1.8 shows the spurious power (i.e. probability of selecting a selected locus using a spurious environmental variable) against the false positive rate for the other loci. Because we expect no association between selected loci and spurious variables, they should be equal (black line). However, LFMM tend to assign selected loci to the spurious environmental variable more often than it does for neutral loci in scenarios IMM and SS (Fig. S1.8).

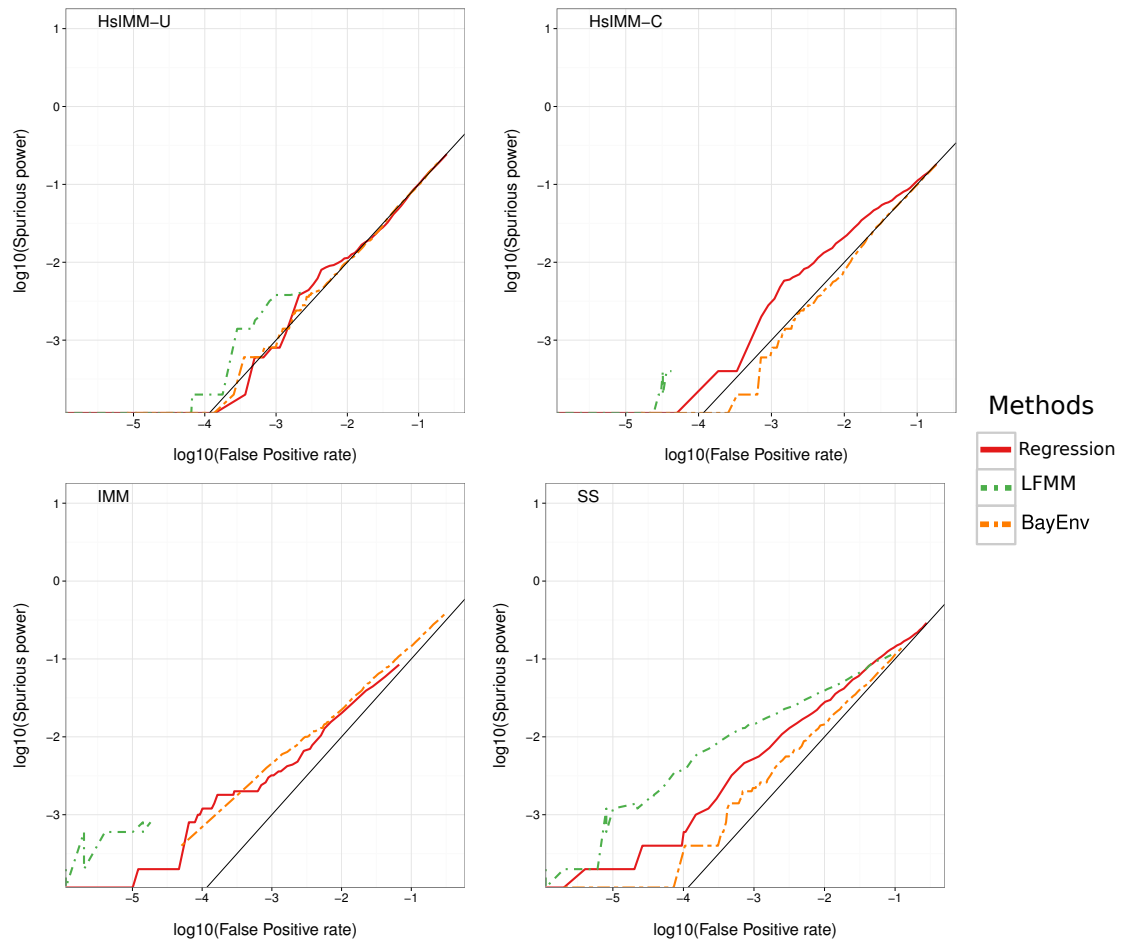


FIG. S1.8 : Spurious “power” (i.e. power to detect selected loci using an independent spurious variable) against false positive rate for the null environmental variable. Scale is log-log. Black line : Expected relation (spurious “power” should be equal to false positive rate). Line are for linear regression (plain red line), Latent Factor Mixed Model (LFMM, green dot-dashed line) and BayEnv (orange two-dashed line).

# ANNEXE S2

## Supplementary Information (ARTICLE 2)

---

### S2.1 Definition of the prior probabilities of jump between models

Recall the three models of which we want to infer posterior probabilities :

**M1** Neutral model :  $\log(\frac{1}{\theta_{ij}}) = \beta_j$ ,

**M2** Local adaptation model with environmental differentiation  $\log(\frac{1}{\theta_{ij}}) = \beta_j + g_i E_j$ ,

**M3** Locus-specific model :  $\log(\frac{1}{\theta_{ij}}) = \alpha_i + \beta_j$ .

Let  $\Pi_2$  be the prior probability of model M2 and  $\Pi_3$  the prior probability of model M3. We assume that the probability of going from M1 or M2 to the model M2 is equal to  $\Pi_2$  (the same reasoning applies for M3 and  $\Pi_3$ ), which leads to the following transition matrix :

$$\begin{pmatrix} 1 - \Pi_2 - \Pi_3 & \Pi_2 & \Pi_3 \\ (1 - \Pi_2)(1 - \Pi_3) & \Pi_2 & \Pi_3(1 - \Pi_2) \\ (1 - \Pi_2)(1 - \Pi_3) & \Pi_2(1 - \Pi_3) & \Pi_3 \end{pmatrix}. \quad (\text{S2.1})$$

If we consider  $\Pi_2 = \pi(1 - p)$  and  $\Pi_3 = \pi p$  where  $\pi$  is the probability of jumping away from model M1 and  $p$  the “preference” for model M3 (i.e. the probability of choosing the model M3 instead of model M2, when jumping away from model M1), then we can write :

$$\begin{pmatrix} 1 - \pi & \pi(1 - p) & \pi p \\ 1 - \pi + \pi^2 p(1 - p) & \pi(1 - p) & \pi p - \pi^2 p(1 - p) \\ 1 - \pi + \pi^2 p(1 - p) & \pi(1 - p) - \pi^2 p(1 - p) & \pi p \end{pmatrix}. \quad (\text{S2.2})$$

Thus, when  $\pi$  is small (in practice  $\pi < 0.5$ ), the transition between models depends only very slightly on the current state of the model, and the prior probabilities of each model reduce approximately to (ignoring second order terms)

$$\begin{aligned} P(\mathbf{M1}) &= 1 - \pi, \\ P(\mathbf{M2}) &= \pi(1 - p), \\ P(\mathbf{M3}) &= \pi p. \end{aligned} \quad (\text{S2.3})$$



## S2.2 Reversible jumps between the models

According to S. BROOKS (1998) and GELMAN et al. (2004), the jump from model  $l$  to model  $k$  should be accepted with a probability  $\min(r, 1)$ , with

$$r = \frac{L(Y|\theta_k, M_k)P(\theta_k|M_k)P(M_k)}{L(Y|\theta_l, M_l)P(\theta_l|M_l)P(M_l)} \frac{J_{k \rightarrow l}J(u_k|\theta_k, k, l)}{J_{l \rightarrow k}J(u_l|\theta_l, l, k)} \left| \frac{\nabla g_{l,k}(\theta_l, u)}{\nabla(\theta_l, u)} \right|, \quad (\text{S2.4})$$

where  $\theta_\bullet$  is the parameter vector for model  $\bullet$ ,  $L$  stands for the likelihood of parameters  $\theta_\bullet$  assuming the model  $M_\bullet$ , and  $J$  is the proposal kernel for the (potentially) new parameter  $u_\bullet$ . Note also the presence of the proposal of new model  $J_{\circ \rightarrow \bullet}$ .

Because a jump toward one of the two alternative models is proposed at each iterations,  $\frac{J_{k \rightarrow l}}{J_{l \rightarrow k}}$  simplifies to one. Likewise, since the transformations from  $\theta_l$  to  $\theta_k$  only consists in setting some parameters to 0, the Jacobian determinant  $\left| \frac{\nabla g_{l,k}(\theta_l, u)}{\nabla(\theta_l, u)} \right|$  also simplifies to one.

The most efficient way to propose a value for  $u_\bullet$  is to draw from its own posterior S. P. BROOKS et al., 2003. Thus, pilot runs are carried out before the actual reversible jump MCMC in order to approximate the posteriors of  $\alpha_i$  and  $g_i$  for each locus  $i$ . During these runs, the parameters  $\beta_j$  are inferred alone (M1 model), whereas the parameters  $\alpha_i$  and  $g_i$  are inferred using models containing  $\beta_j$  (M2 and M3 models). The posterior mean and variance obtained for these parameters from the pilot runs are used to parametrise Normal distributions. These distributions are used to propose a new value of a parameter when a jump to a model including it is proposed. Note that, since  $g_i$  cannot be negative, a truncated Normal is used, and the  $J$  kernel is modified accordingly in Eq. S2.4.

## S2.3 Statistical tests

Using the posterior probability of model M2 for the locus  $i$ ,  $P_i(\mathbf{M2}|\mathbf{a}_i, E)$ , we can calculate the posterior error probability PEP, KÄLL et al., 2008 for locus  $i$  as

$$\text{PEP}_i = 1 - P_i(\mathbf{M2}|\mathbf{a}_i, E). \quad (\text{S2.5})$$

In order to calculate the  $q$ -value STOREY, 2003 ; MULLER et al., 2006, we rank the  $\text{PEP}_i$  from from the lowest to the highest value, and define the  $q$ -value for locus  $i$  as the average PEP for all loci having a PEP lower, or equal to,  $\text{PEP}_i$  :

$$q_i = \frac{1}{i} \sum_{k, k \leq i} \text{PEP}_k. \quad (\text{S2.6})$$

Note that, because we calculate the average using only PEPs that are lower than  $\text{PEP}_i$ , we have  $q_i \leq \text{PEP}_i$  for all  $i$ . The equality only holds for the minimal PEP(s). KÄLL et al. (2008) advocate the use of the  $q$ -value because it is optimal in the sense of Bayesian classification theory see also STOREY, 2003. Our code outputs both PEPs and  $q$ -values.

Both of these test statistics are strongly related to the control of False Discovery Rate FDR, STOREY, 2002 during multiple testing. Contrary to the commonly used False Positive Rate (FPR), which is the probability of declaring a locus as positive given that it is actually neutral, the FDR is the proportion of the positive results that are in fact false positives. Note that the PEP is a “locally” (i.e. regarding only the focal locus) inferred measure of the FDR see KÄLL et al., 2008, whereas the  $q$ -value is based on inferring what the FDR would be when stating that the focal locus, and all the loci with a lower score should be considered as positives.

In the following, we will focus on the  $q$ -value of the local adaptation model (M2). Since we have a strong uncertainty regarding the biological origin of the locus-specific effect  $\alpha_i$ , we can consider it as a “nuisance” parameter in this particular inference framework.

## S2.4 Hierarchically Structured (HS) scenario

Fig. S2.1 gives a schematic representation of the demographic scenario called HS in the main text, showing the fission events and the migration between populations. Note that only some illustrative migration combinations are represented for the sake of simplicity. An important feature of this model is that the probability of migration between two populations decreases as their relatedness (measured by the number of fission events separating them) decreases. A full model description can be found in de VILLEMEREUIL et al. (2014).

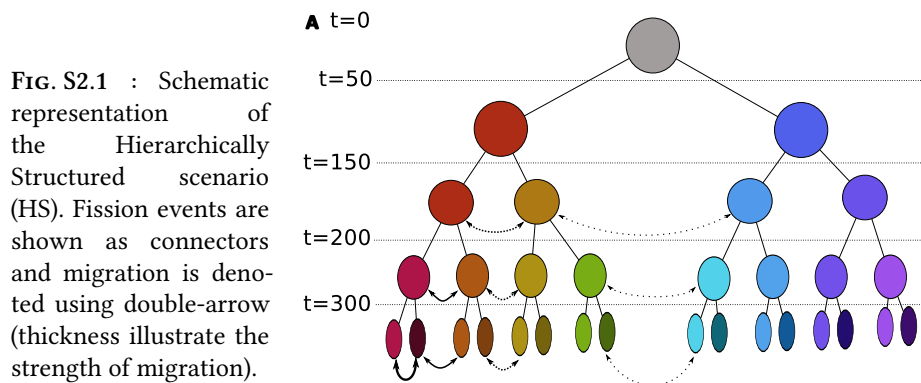


FIG. S2.1 : Schematic representation of the Hierarchically Structured scenario (HS). Fission events are shown as connectors and migration is denoted using double-arrow (thickness illustrate the strength of migration).

## S2.5 Simulation scripts

Four Python scripts are provided as Supplementary Files :

- IM\_mono.py Island model with monogenic selection
- IM\_poly.py Island model with polygenic selection
- SS\_poly.py Stepping-stone model with polygenic selection
- HS\_poly.py Hierarchically structured model with polygenic selection

These scripts require Python 2.7 and SimuPOP 1.1. Note that the monogenic version is provided only for the island model, as the modification are identical for the two other models.

## S2.6 Environmental gradient

The environmental gradient was designed to be independent from (i.e. not be confounded with) the population structure for the scenarios SS and HS (see Fig. S2.2).

This environmental variable was already standardised for its use in BayeScEnv. A transformation of this variable was used to define selective pressure on an appropriate scale :

$$s_j = s_0 \frac{1 - e^{-E_j}}{1 + e^{-E_j}}, \tag{S2.7}$$

where  $s_0$ , the strength of the selection was chosen to be 0.1 for the monogenic case and 0.02 for the polygenic case. The individual fitness was calculated in a multiplicative fashion :

$$W = (1 + s_j)^{n_{11}} (1 - s_j)^{n_{00}}, \tag{S2.8}$$

where  $n_{11}$  and  $n_{00}$  are the number of loci at which the individual is homozygous for the advantageous and disadvantageous allele, respectively. Note that this fitness function assumes co-dominance, with a heterozygous fitness of 1.

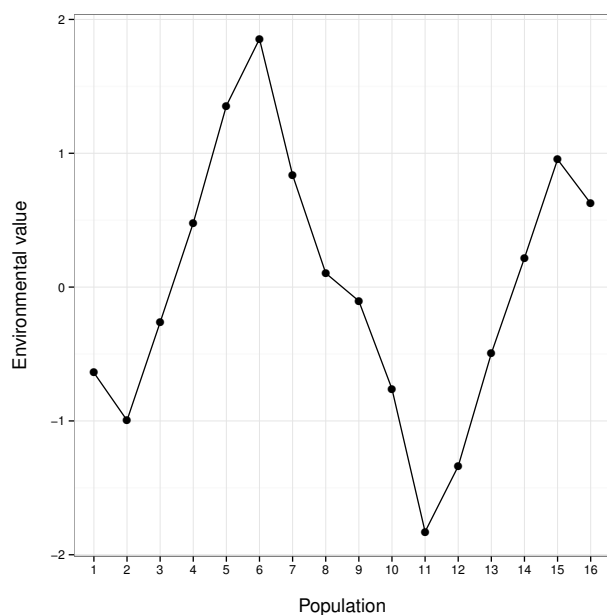


FIG. S2.2 : Environmental value for all 16 populations in the scenarios SS and HS.

The recombination rate was set to 0.002 (one recombination between two adjacent loci, *per* population and *per* generation). The mutation rate was set to  $10^{-7}$  *per* generation at every locus. The allele frequencies were initialised using a beta-binomial distribution truncated between 0.1 and 0.9 to avoid too many monomorphic loci.

## S2.7 Simulation results for the False Positive Rate

The results regarding the False Positive Rate (FPR, Fig. S2.3) were qualitatively comparable to the results regarding the False Discovery Rate (FDR, Fig. A2.2, main text). Indeed, we again found that the most error-prone method was BayeScan, where BayeScEnv yielded fewer false positives. For the latter, the parametrisation  $\pi = 0.1$ ,  $p = 0.5$  was, as expected, the most conservative, whereas  $\pi = 0.1$ ,  $p = 0$  was the most laxist.

## S2.8 List of candidate genes associated with significant GO terms

Below is a list of the genes that fulfill the following two criteria. First, there is at least one significant SNPs in their neighbourhood, indicating them as potential candidates for local adaptation. Second, at least one of their associated GO terms were found to be significantly enriched for candidates compared to the rest of the genome.

- **For the altitude analysis** : SCARB1, SLC12A1, MUCL1, DNM2, MLANA, ATP6V1C2, CLDN12, FBN1, OTUD7A, SLC24A5, NOS1AP, SLC12A8
- **For the temperature analysis** : SYNE2, SPTB, ANKRD46, HAO1, HCK, FOXP1, ONECUT2, CDH15, ATP8A2, FADS2, ESR2, ATP6V1C2, FADS1, NRG1, APBB2, CMYA5, SERPINA6, SLC8A1, PRKG1, LAMA2, SERPINA1

Note that the majority of the significant GO terms were represented by only one gene for the altitude and temperature analysis. The list of genes fulfilling the two criteria is not shown for the precipitation analysis and BayeScan, as there are too many of them.

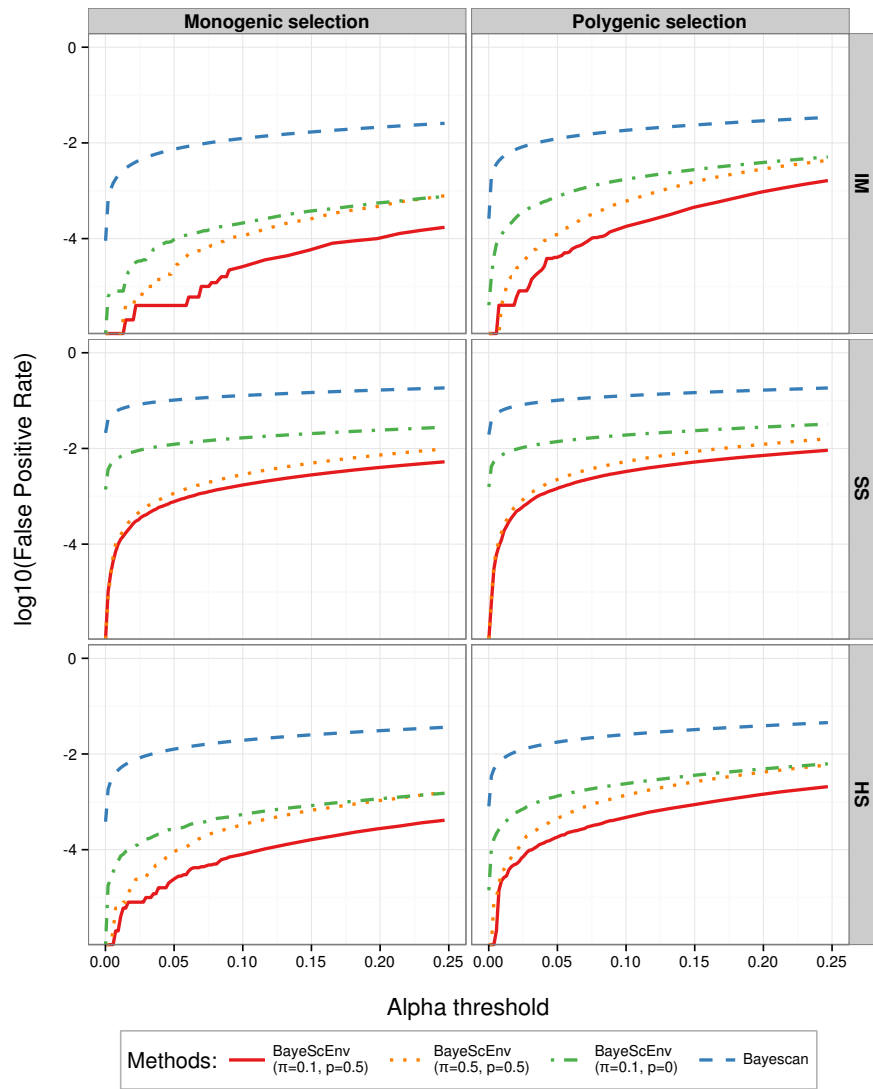


FIG. S2.3 : False Positive Rate (FPR) against significance threshold  $\alpha$  for three scenarios (IM : Island model, SS : Stepping-Stone model and HS : Hierarchically Structured model) and monogenic/polygenic selection. The models tested are BayeScan (blue dashed), and BayeScEnv (orange dotted, green dot-dashed and solid red) with different probabilities  $\pi$  of jumping away from the neutral model and different preferences  $p$  for the locus-specific model. Note that  $p = 0$  means the environmental model is tested against the neutral one only.

## S2.9 Analysis of the simulation scenarios from de VILLEMEREUIL et al. (2014)

**Scenarios** Given the computationally expensive nature of the simulations necessary to generate synthetic data, we compare BayeScEnv with methods other than BayeScan using the simulated datasets from de VILLEMEREUIL et al. (2014). For more information regarding these scenarios, please refer to the article. Briefly, four polygenic scenarios were tested :

**HsIMM-C** Hierarchical scenario with a clinal environment following population structure

**HsIMM-U** Hierarchical scenario with a random environment strongly correlated with population structure

**IMM** Isolation with Migration Model

**SS** Stepping-Stone model with a clinal environment, following the clinal population structure

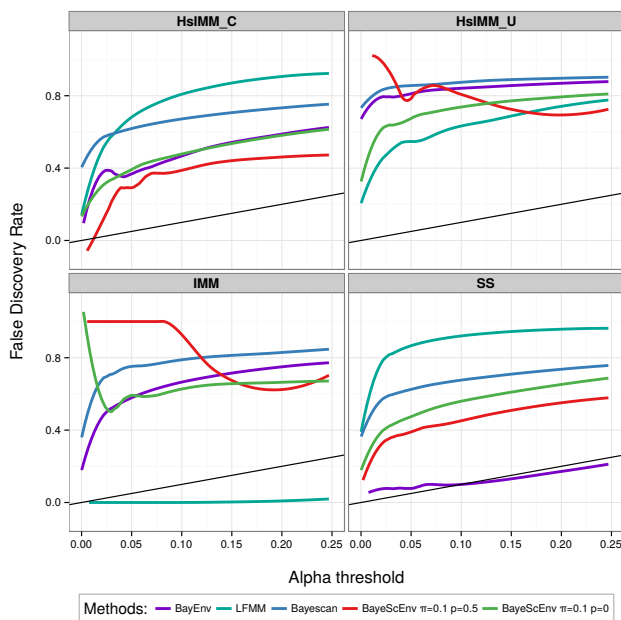
**Foreword** These scenarios were tested against BayeScan FOLL & GAGGIOTTI, 2008, Bayenv COOP et al., 2010 and LFMM FRICHOT et al., 2013. Note that these scenarios are very difficult for all methods, thus we do not expect the FDR to be well calibrated. Also, in contrast with the study in the main

text, de VILLEMEREUIL et al. used a prior odds of 100 instead of 10 for BayeScan.

When interpreting the results, it should be remembered that the FDR depends on both the FPR and power. All things being equal, the FDR will be higher if the FPR is higher, and lower if the power is higher.

**Results** The results are presented in Figs. S2.4–S2.6 (below). They show that, even under very difficult conditions, BayeScEnv inferences are fairly reliable. As expected, both the FPR and the power of BayeScEnv are lower than that of BayeScan, resulting in a more conservative method overall. However, in scenarios with low power for all methods, BayeScEnv’s lack of power can drastically inflate its FDR (e.g. Fig. S2.4, red line, IMM model). Interestingly, BayeScEnv with  $p = 0$  is more robust in this regard, since its power is generally much higher.

When compared to the other association methods (Bayenv and LFMM), BayeScEnv performed very well in “clinal” scenarios (HsIMM-C and SS), but more poorly in the other scenarios. However when considering the canonical  $\alpha = 0.05$  threshold, BayeScEnv’s FDR was always lower than at least one of the association methods, except in the IMM scenario.



**False Discovery Rate** When comparing the FDR to other methods, BayeScEnv performs relatively well. Especially, for  $p = 0.5$  (red line), its FDR can be the lowest (HsIMM-C) or second best (SS), but it can reach very high values under some scenarios (HsIMM-U and IMM). Surprisingly, the parametrisation  $p = 0$  (green) is more stable across scenarios, whereas Bayenv (purple) and LFMM (turquoise) constantly “switch” between best-or-so and poorest-or-so. Overall, BayeScEnv’s FDRs with  $p = 0$  are lower than those of BayeScan’s, at least for the canonical threshold  $\alpha = 0.05$ . Large values of FDR for small  $\alpha$ ’s are due to a lack of power (see below), not to a high False Positive Rate.

FIG. S2.4 : False Discovery Rate (FDR) against significance threshold  $\alpha$  for de VILLEMEREUIL et al. (2014) polygenic scenarios.

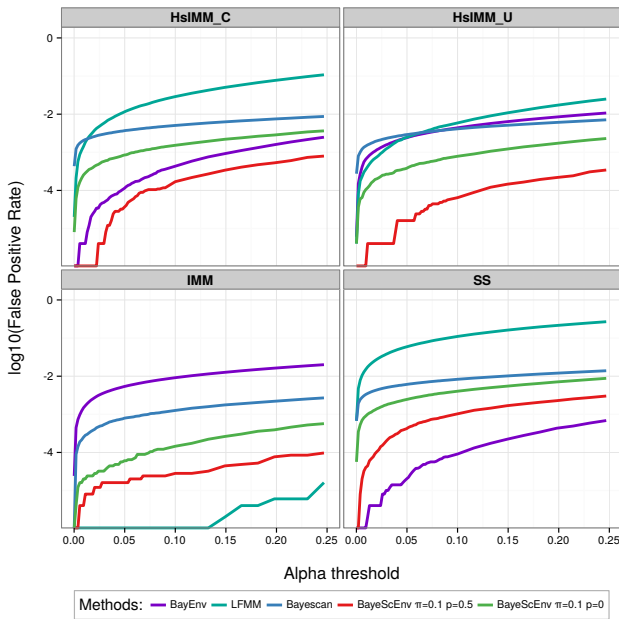


FIG. S2.5 : False Positive Rate (FPR) against significance threshold  $\alpha$  for de VILLEMEREUIL et al. (2014) polygenic scenarios.

**False Positive Rate** FPRs are more predictable than FDRs regarding the  $F$  model family : BayeScan (blue) is the most error-prone method, followed by BayeScEnv with  $p = 0$  (green) while BayeScEnv with  $p = 0.5$  (red) is one of the most conservative methods. Interestingly, the FPRs of the  $F$  model family are more stable than the FPRs of Bayenv (purple) and LFMM (turquoise), which vary greatly across scenarios.

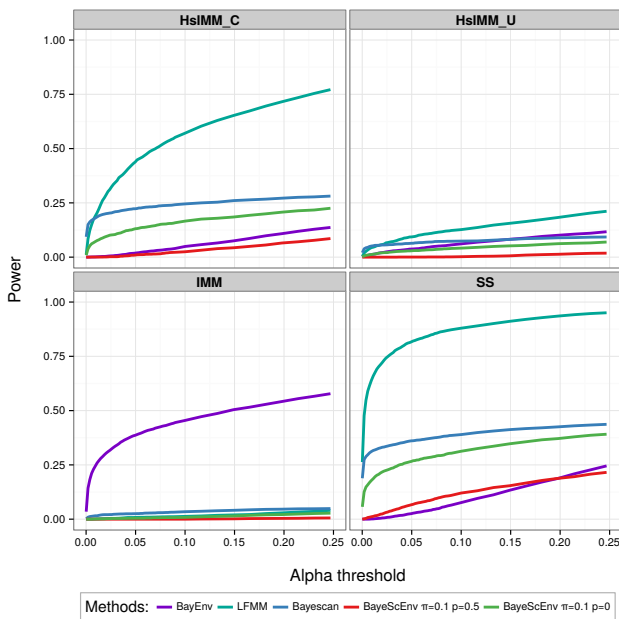


FIG. S2.6 : Power against significance threshold  $\alpha$  for de VILLEMEREUIL et al. (2014) polygenic scenarios.

**Power** Overall power varies greatly across scenarios, HsIMM-U and IMM being the most difficult ones. As expected, the power of BayeScEnv (red and green) is always lower than the power of BayeScan (blue). However, the power of BayeScEnv with  $p = 0$  (green), is always comparable to that of BayeScan's. BayeScEnv with  $p = 0.5$  (red) is always among the less powerful method. For all scenarios, at least on of the environmental association methods (Bayenv (purple) and LFMM (turquoise)), has greater power than BayeScan and BayeScEnv.

## ANNEXE S3

# Méthodes de génétique quantitative pour les modèles linéaires mixtes généralisés (ARTICLE 6)

---

### Présentation de l'article

#### Contributions

Le cadre conceptuel de cette étude provient essentiellement de Michael Morrissey et moi-même, avec l'aide de Holger Schielzeth et Shinichi Nakagawa. J'ai effectué les analyses par simulation, et Micheal Morrissey a effectué les analyses empiriques. J'ai écrit cet article avec de fortes contributions de Michael Morrissey, ainsi que de Holger Schielzeth et Shinichi Nakagawa.

#### Résumé de l'article

Les méthodes pour l'inférence et l'interprétation des paramètres de génétique quantitative évolutive, ainsi que pour la prédiction des réponses évolutives à la sélection naturelle, sont plutôt développés pour des traits possédant une distribution Gaussienne. Pourtant, beaucoup de traits d'intérêt évolutif, comme la plupart des traits liés à l'histoire de vie ou au comportement, possèdent des distributions intrinsèquement non Gaussiennes.

Il est donc de plus en plus courant de recourir aux modèles linéaires mixtes généralisés pour estimer des paramètres de génétique quantitative, comme la variance génétique additive, pour ces traits non Gaussiens. Ces méthodes statistiques permettent de faire des inférences sur l'échelle d'un trait sous-jacent statistiquement convenient, mais pas sur l'échelle du phénotype observé.

Nous fournissons une approche générale pour calculer les paramètres de génétique quantitative sur l'échelle du phénotype observée pour n'importe quel type de modèle généralisé. Notre approche ne demande aucune hypothèse supplémentaires que celles déjà supposées par les modèles généralisés. Nous montrons que les formules existant pour les modèles binomiaux (notamment le modèle à seuil) et de Poisson sont des cas spéciaux de notre cadre théorique.

À l'aide de simulations, nous démontrons l'exactitude de notre méthode pour prédire la réponse évolutive à la sélection, and illustrons notre approche par une analyse de données empiriques sur une population naturelle avec informations sur le pédigrée. Notre méthode est implémentée dans un paquet R, QGGLMM (<https://github.com/devillemereuil/qgglmm>).

**Journal:** *Evolution*, Soumis

Les *Supplementary Information* de cet article sont disponibles dans l'Annexe S6

## ARTICLE 6

# General methods for evolutionary quantitative genetic inference from generalised mixed models

Pierre de Villemereuil, Holger Schielzeth, Shinichi Nakagawa & Michael Morrissey

### Abstract

Methods for inference and interpretation of evolutionary quantitative genetic parameters, and for prediction of evolution in response to natural selection, are best developed for traits with normal distributions, yet many traits of evolutionary interest, such as most life history and behavioural traits, have inherently non-normal distributions. It is therefore becoming increasingly common to estimate quantitative genetic parameters, such as additive genetic variances, for non-normal traits using generalised linear mixed model (GLMM) analyses. These statistical methods provide inferences on a statistically-convenient latent scale, but not on the scale upon which traits are expressed. We provide a general approach for calculating quantitative genetic parameters on the observed scale for arbitrary GLMMs. Our approach requires no additional assumptions beyond those that made when adopting GLMM-based quantitative genetic analysis. We show that existing formulae for Binomial and Poisson distributions, in particular the threshold models for binary traits, are special cases in our framework. We use simulations to demonstrate the accuracy of our method for predicting the evolutionary response to selection and further illustrate our approach by applying it to data from a natural study population with pedigree information. Our framework is implemented in an R package, QGGLMM (<https://github.com/devillemereuil/qgglm>).

### Introduction

Additive genetic variances and covariances of phenotypic traits determine the response to selection, and so are key determinants of the processes of adaptation in response to natural selection and of genetic improvement in response to artificial selection (FISHER, 1918; FALCONER, 1960; LYNCH & WALSH, 1998; WALSH & LYNCH, p.d.). While the concept of additive genetic variance (FISHER, 1918; FALCONER, 1960) is very general, being applicable to any type of character with any arbitrary distribution, including, for example, fitness (FISHER, 1930), techniques for estimating additive genetic variances and covariances are best developed for Gaussian traits (i.e., traits with a normal distribution; HENDERSON 1950; LYNCH & WALSH 1998). Furthermore, quantitative genetic theory for predicting responses to selection are also best developed and established for Gaussian characters (WALSH & LYNCH, p.d.), but see MORRISSEY (2015).

Thus, many characters of potential evolutionary interest, are not Gaussian (e.g. survival or number of offspring), and so are not well-handled with existing theory and methods. Comprehensive systems for estimating genetic parameters and predicting evolutionary trajectories of non-Gaussian traits will hence be very useful for quantitative genetic studies of adaptation.

For Gaussian traits, a linear mixed model allows various analyses of factors that contribute to the mean and variance of phenotype. In particular, a formulation of a linear mixed model called the 'animal model' (HENDERSON, 1973; KRUK, 2004; WILSON et al., 2010) provides a very general method for estimating additive genetic variances and covariances, given arbitrary pedigree data, and while potentially accounting for a range of different types of confounding variables, such as environmental effects, measurement error, or maternal effects.



A general statement of an animal model analysis decomposing variation in a trait,  $z$ , into additive genetic and other components would be

$$\mathbf{z} = \mu + \mathbf{X}\mathbf{b} + \mathbf{Z}_a\mathbf{a} + \mathbf{Z}_1\mathbf{u}_1 + \dots + \mathbf{Z}_k\mathbf{u}_k + \mathbf{e}, \quad (\text{A6.1})$$

where  $\mu$  is the model intercept, and  $\mathbf{b}$  is a vector of fixed effects such as sex and age, potentially relating both continuous and categorical effects to observations via the fixed effects design matrix  $\mathbf{X}$ , just as in an ordinary linear model.

An arbitrary number of random effects can be modelled, with design matrices  $\mathbf{Z}$ , where effects ( $\mathbf{a}, \mathbf{u}_1 \dots \mathbf{u}_k$ ) are assumed to be drawn from normal distributions with variances to be estimated. The key feature of the animal model is that it includes individual additive genetic effects, or breeding values (conventionally denoted  $\mathbf{a}$ ). These additive genetic effects, and critically, their variance, are estimable given arbitrary pedigree data, which defines the relatedness of all individuals in an analysis. The covariances of breeding values among individuals can be modelled according to

$$\mathbf{a} \sim \mathcal{N}(\mathbf{0}, \mathbf{A}V_A), \quad (\text{A6.2})$$

where  $\mathbf{A}$  is the additive genetic relatedness matrix derived from the pedigree and  $V_A$  is the genetic additive variance.

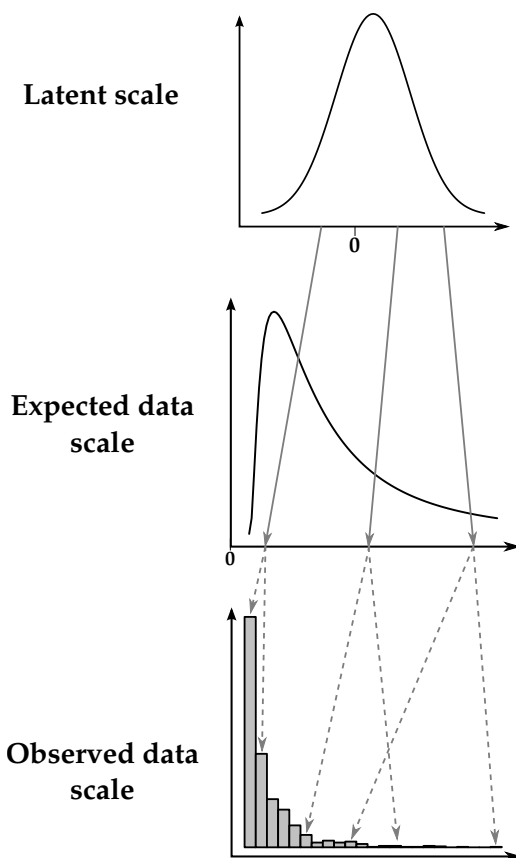


FIG. A6.1: Example of the relationships between the three scales of the GLMM using a Poisson distribution and a logarithm link function. Deterministic relationships are denoted using grey plain arrows, whereas stochastic relationships are denoted using grey dashed arrows.

Many non-Gaussian traits, however, just cannot be strictly additive on the scale on which they are expressed. Consider, for example, survival probability. In a population where the average probability of survival is 0.7, imagine there are two loci with major effects. If, at each locus separately, substituting one allele for the other would increase survival by 0.2, the genetic effects cannot be additive, as their combination must generate a survival probability that does not exceed one. In such a scenario, it may be reasonable to assume that there exists an underlying scale, related to survival probability, upon which genetic and other effects are additive.

Generalised linear mixed model (GLMM) analysis can be used for inference of quantitative genetic parameters by applying precisely this kind of reasoning. A latent scale is assumed (FIG. A6.1), on which effects on the propensity for expression of some trait are assumed to be additive. A function, called a 'link function' is applied that links expected values for a trait to the latent scale.

For example, a trait that is expressed in counts, say, number of behaviours expressed in a unit time, is a strictly non-negative quantity. As depicted in FIG. A6.1, a strictly positive distribution of expected values may related to latent values ranging from  $-\infty$  to  $+\infty$  by a function such as the log link. Finally, a distribution function is required to model the "noise" of observed values around expected value (FIG. A6.1). Different distributions make sense for different traits, but for the example of a count trait such as that depicted in FIG. A6.1, observed values may be modelled as drawn from a Poisson distribution, with expectations related to the latent scale via the log link function.

The GLMM framework thus involves three scales on which we can think of variation in a trait occurring. More formally, GLMM can be written:

$$\mathbf{l} = \mu + \mathbf{X}\mathbf{b} + \mathbf{Z}_a\mathbf{a} + \mathbf{Z}_1\mathbf{u}_1 + \dots + \mathbf{Z}_k\mathbf{u}_k + \mathbf{o} \quad (\text{A6.3a})$$

$$\boldsymbol{\eta} = g^{-1}(\mathbf{l}) \quad (\text{A6.3b})$$

$$\mathbf{z} \sim \mathcal{D}(\boldsymbol{\eta}, \boldsymbol{\theta}) \quad (\text{A6.3c})$$

where Eq. A6.3a is just as for a LMM (Eq. A6.1), except that it describes variation in the latent trait  $\mathbf{l}$ , rather than the response directly. Note that we now refer to the “residual” (noted  $\mathbf{e}$  in Eq. A6.1) as “overdispersion” (denoted  $\mathbf{o}$ , with a variance denoted  $V_O$ ), as residuals (variation around expected values) are defined by the distribution function,  $\mathcal{D}$ , in this model. Eq. A6.3b formalises the idea of the link function.

Any link function has an associated inverse link function,  $g^{-1}$ , which is often useful for converting specific latent values to expected values. For example, where the log link function translates expected values to the latent scale, its inverse, the exponential function, translates latent values to expected values.

Finally, Eq. A6.3c specifies the distribution by which observations occur around the expected values defined in Eq. A6.3b according to some distribution function, that may involve parameters other than the expectation, denoted  $\boldsymbol{\theta}$ . Just as for the LMM (Eq. A6.1), all random effects are assumed to follow a normal distribution with variances to be estimated, but on the latent scale. Particularly, the variance of additive genetic effects  $\mathbf{a}$  are assumed to follow Eq. A6.2 on the latent scale.

Linear mixed model-based inferences of genetic parameters, using the ‘animal model’, have become routine (KRUK, 2004; WILSON et al., 2010). The use of generalised linear mixed animal model analysis is now also growing (e.g. MILOT et al., 2011; WILSON et al., 2011; MORRISSEY et al., 2012; de VILLEMEREUIL et al., 2013; AYERS et al., 2013). However, whereas Gaussian animal model analysis directly estimates additive genetic parameters on the scale on which traits are expressed and selected, and upon which we may most naturally consider their evolution, this is not so for generalised analyses. Genetic variance components estimated in a generalised animal model are obtained on the latent scale. Hence, the “conventional” formula to compute heritability:

$$h_{\text{lat}}^2 = \frac{V_A}{V_A + V_{\text{RE}} + V_O} \quad (\text{A6.4})$$

where  $V_{\text{RE}}$  is the summed variance of all random effects apart from the additive genetic variance, and  $V_O$  is the overdispersion variance, is the heritability on the latent scale, not on the observed data scale.

Although it might sometimes be sensible to measure the heritability of a trait on the latent scale (for example, in animal breeding, where selection might be based on latent breeding values), it is natural to seek inferences on the scale upon which the trait is expressed, and on which we may think of selection as acting. Some expressions exist by which various parameters can be obtained or approximated on the observed data scale. For example, various expressions for the intra-class correlation coefficients on the data scale exist (reviewed in NAKAGAWA & SCHIELZETH, 2010), but these do not provide inferences of additive genetic variance on the data scale.

Exact analytical expressions exist for the additive genetic variance and heritability on the observed data scale for two specific and important families of GLMMs (i.e. combinations of link functions and distribution functions): for a binomial model with a probit link function (i.e., the “threshold model,” DEMPSTER & LERNER, 1950) and for a Poisson model with a logarithm link function (FOULLEY & IM, 1993). A general system for calculating genetic parameters on the expected and observed data scales for arbitrary GLMMs is currently lacking.

In addition to handling the relationship between observed data and the latent trait via the link and distribution functions, any system for expected and observed scale quantitative genetic inference with GLMMs will have to account for complex ways in which fixed effects can influence quantitative genetic parameters. It is currently appreciated that fixed effects in LMMs explain variance, and

that variance associated with fixed effects can have a large influence on summary statistics such as repeatability (NAKAGAWA & SCHIELZETH, 2010) and heritability (WILSON, 2008).

This principle holds for GLMMs as well; but fixed effects result in additional, and very important, complications for interpreting GLMMs. While random and fixed effects are independent in a GLMM on the latent scale, the non-linearity of the link function renders them inter-related on the expected and observed scales. Consider, for example, a GLMM with a log link function. Because the exponential is an upwardly curved function, the influence of fixed and random effects will create more variance on the expected and observed data scales for larger values than for smaller values.

While it will undoubtedly be desirable to develop a comprehensive method for making data-scale inferences of quantitative genetic parameters with GLMMs, such an endeavour will not yield a system for predicting evolution in response to natural or artificial selection, even if a particular empirical system is very well served by the assumptions of a GLMM.

This is because systems for evolutionary prediction, specifically the Breeder's equation (LUSH, 1937; FISHER, 1924) and the Lande equation (LANDE, 1979; LANDE & ARNOLD, 1983), assume that breeding values (and in most applications, phenotypes) are multivariate normal (WALSH & LYNCH, p.d.). Even if it is possible to estimate additive genetic variances of traits on the expected and observed data scales, we will show that these quantities will not strictly be usable for evolutionary prediction. However, the latent scale in a GLMM does, by definition, satisfy the assumptions of the Breeder's and Lande equations.

Thus, for the purpose of predicting evolution, it may be useful to be able to express selection of non-Gaussian traits, not on the observed scale, but rather on the scale of the latent. Such an approach could yield a system for evolutionary prediction of characters that have been modelled with a GLMM, requiring no more assumptions than those that are already made in applying the statistical model.

Here, we propose a system for making inferences of quantitative genetic parameters on the expected and observed scales, for arbitrary GLMMs. We show how to estimate broad-sense and additive genetic variances and covariances, accounting for fixed effects as necessary. We lay out the formal theory underlying the system, apply it to an empirical dataset, and provide software for implementation. The relationships between existing analytical formulae and our general framework are also highlighted. Next, we outline a system of evolutionary prediction for non-Gaussian traits that capitalises on the fact that the latent scale in a GLMM satisfies the assumptions of available equations for the prediction of evolution. We show in a simulation study that (i) additive genetic variances on the observed data scale do not predict evolution, and can, in fact, give substantial errors, and (ii) that making inferences via the latent scale provides unbiased predictions. We also provide software for making evolutionary predictions using the latent scale. Although all examples and most equations in this article are presented in a univariate form, all our results are applicable to multivariate analysis, which is implemented in our software. Together, these approaches provide a comprehensive treatment of the evolutionary quantitative genetics of traits that may be modelled with GLMMs.

## Quantitative genetic parameters in GLMMs

Throughout this section we will refer to the additive genetic variance as  $V_A$ , the summed variance of other random effects as  $V_{RE}$  and the overdispersion variance (i.e. the variance of  $\mathbf{o}$ ) as  $V_O$ .

### Mean and phenotypic variance

**Observed-scale population mean** Mean phenotype on the expected and observed data scales is

$$\bar{z} = \int g^{-1}(l)f(l)dl, \quad (\text{A6.5})$$

where  $f(l)$  is the probability density of  $l$ . Typically, and especially in the absence of fixed effects, the distribution of  $l$  will be a normal with a mean  $\mu$  and variance  $V_A + V_{RE} + V_O$ .

In the presence of fixed effects, it is necessary to average over the components of the predictive values marginalised over the random effects (i.e.  $\mathbf{X}\hat{\mathbf{b}}$ , where  $\hat{\mathbf{b}}$  are the fixed effects estimates) as well as integrating over the random parts of  $l$

$$\bar{z} = \frac{1}{N} \sum_{i=1}^N \int g^{-1}(l) f_{\mathcal{N}}(l, \mu + \hat{\mathbf{l}}_i, V_A + V_{RE} + V_O) dl, \quad (\text{A6.6})$$

where  $N$  is the number of predicted latent values in  $\hat{\mathbf{l}} = \mathbf{X}\hat{\mathbf{b}}$ . Typically,  $\mathbf{X}$  will be the fixed effects design matrix used when fitting the generalised animal model (equations A6.1, A6.2, and A6.3), and  $N$  will be the number of data observations. However,  $\mathbf{X}$  could profitably be modified to a general prediction matrix in some scenarios. For example, if a model included a fixed effect for sex, and if the population in question had an equal sex ratio but the data did not, an  $\mathbf{X}$  matrix might be used that represented both sexes equally. Throughout the rest of this section, and for the sake of clarity, we will assume the simple case of no fixed effects, but all equations can easily be transformed as for Eq. A6.6. We will only specify versions of a few fundamental equations that account for fixed effects.

**Expected-scale phenotypic variance** Phenotypic variance on the expected data scale can be obtained analogously to the observed scale population mean. Having obtained  $\bar{z}$ , the phenotypic variance is

$$V_{P,\text{exp}} = \int (g^{-1}(l) - \bar{z})^2 f(l) dl. \quad (\text{A6.7})$$

**Observed-scale phenotypic variance** Phenotypic variance of observed values is the sum of the variance in expected values and variance arising from the distribution function. Since these variances are independent by construction in a GLMM, they can be added. This distribution variance is influenced by the latent trait value, but might also depend on additional distribution parameters included in  $\theta$  (e.g. for Negative Binomial, see Eq. A6.3c). Given a distribution-specific variance function  $v$ :

$$V_{P,\text{obs}} = V_{P,\text{exp}} + \int v(l, \theta) f(l) dl. \quad (\text{A6.8})$$

## Broad-sense genetic parameters

The expected phenotype of an individual with a given latent genetic value  $a$ , i.e., its genotypic value  $E[z|a]$ , is given by

$$E[z|a] = \int g^{-1}(l) f(l|a) dl \quad (\text{A6.9})$$

where  $f(l|a)$  is the density of the latent trait for a given value of  $a$ . For example, in absence of fixed effects,  $f(l|a)$  would be  $f_{\mathcal{N}}(l, \mu + a, V_{RE} + V_O)$ .

The broad-sense genetic variance on both data the expected and observed data scales is then

$$V(E[z|a]) = \int (E[z|a] - \bar{z})^2 f_{\mathcal{N}}(a, 0, V_A) da. \quad (\text{A6.10})$$

This is the total genetic variance on the observed scale, arising from strictly additive genetic variance on the latent scale. Some additive genetic variance on the latent scale is manifested as non-additive variance on the expected and observed scales because of the non-linearity of the link function. If non-additive genetic effects are modelled on the latent scale, they would be included in the expectations and integrals in Eq. A6.9 and A6.10.

## Narrow-sense genetic parameters

The narrow-sense additive genetic variance is limited to the part of the genetic variance that is additive on the observed scales. In other words, it is the variance of the breeding values computed

on the observed scales. Following Robertson (1950; see also FISHER 1918), breeding values on the expected and observed scales  $a_{\text{obs}}$  are the part of the observed phenotype  $z$  that depends linearly on the latent breeding values. The breeding values on the expected and observed data scales can then be defined as the predictions of a least-squares regression of the observed data on the latent breeding values,

$$a_{\text{obs}} = \hat{z}|a = m + ba, \quad (\text{A6.11})$$

where  $\hat{z}$  is the value of  $z$  predicted by the regression. Thus, we have  $V_{A,\text{obs}} = b^2 V_A$  and, from standard regression theory:

$$b = \frac{\text{cov}(z, a)}{V_A}. \quad (\text{A6.12})$$

Because of the independence between the expected values of  $z$  (i.e. the expected data scale  $g^{-1}(l)$ ) and the distribution “noise” (see EQ. A6.8), we can obtain the result that  $\text{cov}(z, a) = \text{cov}(g^{-1}(l), a)$ , hence:

$$b = \frac{\text{cov}(g^{-1}(l), a)}{V_A}. \quad (\text{A6.13})$$

Stein’s (1973) lemma proves that if  $X$  and  $Y$  are bivariate normally distributed random variables, then the covariance of  $Y$  and some function of  $X$ ,  $f(X)$ , is equal to the expected value of  $f'(X)$  times the covariance between  $X$  and  $Y$ , so,

$$\text{cov}(g^{-1}(l), a) = E \left[ \frac{dg^{-1}(l)}{dl} \right] \text{cov}(l, a) = E \left[ \frac{dg^{-1}(l)}{dl} \right] V_A, \quad (\text{A6.14})$$

noting that the covariance of latent breeding values and latent values is the variance of breeding values. Finally, combining EQ. A6.13 with EQ. A6.14, we obtain:

$$b = E \left[ \frac{dg^{-1}(l)}{dl} \right]. \quad (\text{A6.15})$$

To avoid confusion with various uses of  $b$  as other forms of regression coefficients, and for consistency with MORRISSEY (2015), we denote the average derivative of expected value with respect to latent value as  $\Psi$ . In the absence of fixed effects in the model,  $\Psi$  is

$$\Psi = \int \frac{dg^{-1}(l)}{dl} f_{\mathcal{N}}(l, \mu, V_A + V_{\text{RE}} + V_{\text{O}}) dl. \quad (\text{A6.16})$$

If fixed effects (other than the intercept  $\mu$ ) are included in the model, the equation above should be modified accordingly:

$$\Psi = \frac{1}{N} \sum_{i=1}^N \int \frac{dg^{-1}(l)}{dl} f_{\mathcal{N}}(l, \mu + \hat{l}_i, V_A + V_{\text{RE}} + V_{\text{O}}) dl. \quad (\text{A6.17})$$

The additive genetic variance on the phenotypic scales is then

$$V_{A,\text{obs}} = V_{A,\text{exp}} = \Psi^2 V_A. \quad (\text{A6.18})$$

An alternative derivation of EQ. A6.17, and its associated definition of  $\Psi$ , for the general calculation of the additive genetic variances following a non-linear transformation, is given in MORRISSEY (2015).

## Summary statistics and multivariate extensions

Equations A6.5 through A6.18 give the values of different parameters that are useful for deriving other evolutionary quantitative genetic parameters on the observed data scale. Hence, from them, other parameters can be computed. Heritability on the observed data scale can be written as

$$h_{\text{obs}}^2 = \frac{V_{A,\text{obs}}}{V_{P,\text{obs}}}. \quad (\text{A6.19})$$

Replacing  $V_{P,obs}$  by  $V_{P,exp}$  will lead to the heritability on the expected data scale  $h_{exp}^2$ .

Parameters such as additive genetic coefficient of variance and evolvability (HOULE, 1992) can be just as easily derived. The coefficient of variation on the observed phenotypic scale will be computed as

$$CV_{A,obs} = 100 \frac{\sqrt{V_{A,obs}}}{\bar{z}}, \quad (\text{A6.20})$$

and the evolvability on the observed scale will be

$$I_{A,obs} = \frac{V_{A,obs}}{\bar{z}^2}. \quad (\text{A6.21})$$

The genetic basis of multivariate phenotype, especially as summarised by the  $\mathbf{G}$  matrix is often of interest. For simplicity, all expressions considered to this point have been presented in univariate form. However, every expression has a fairly simple multivariate extension.

For example, the mean vector of phenotype in a multivariate analysis is obtained by defining the link function to be a vector-valued function, returning a vector of expected values from a vector of values on the latent scale. Mean phenotype and phenotypic variance is then obtained by integrating the vector-valued link function times the multivariate normal distribution total variance on the latent scale, as in Eq. A6.5 and Eq. A6.8. Integration over fixed effects for calculation of the multivariate mean is directly analogous to the extension of Eq. A6.5 given in Eq. A6.6.

Calculation of other parameters, such as multivariate genotypic values, broad-sense covariance matrices, and phenotypic covariance matrices, have directly equivalent multivariate versions as well. The additive genetic variance-covariance matrix (the  $\mathbf{G}$  matrix) on the observed scale is simply the multivariate extension of Eq. A6.18, i.e.,  $\mathbf{G}_{obs} = \mathbf{\Psi} \mathbf{G}_l \mathbf{\Psi}^T$ . Here,  $\mathbf{G}_l$  is the latent  $\mathbf{G}$  matrix and  $\mathbf{\Psi}$  is the average gradient matrix of the vector-valued link function, which is a diagonal matrix of  $\Psi$  values for each trait (simultaneously computed from a multivariate version of Eq. A6.16).

## Relationships with existing analytical formulae

### ► Binomial distribution and the threshold model

Heritabilities of binary traits have a long history of analysis with a threshold model (S. WRIGHT, 1934 ; DEMPSTER & LERNER, 1950), whereby an alternate trait category is expressed when a trait on a latent “liability scale” exceeds a threshold. It can be shown (see Supplementary Information, section A) that a GLMM with a binomial distribution and a probit link function is exactly equivalent to such a model.

Heritability can then be computed on this “liability” scale (different from the expected data scale!) by using adding the “link variance”  $V_L$  to the denominator (NAKAGAWA & SCHIELZETH, 2010 ; de VILLEMEREUIL et al., 2013) :

$$h_{thres}^2 = \frac{V_A}{V_A + V_{RE} + V_O + V_L}. \quad (\text{A6.22})$$

Because the probit link function is the inverse of the cumulative standard normal distribution function, the “link variance”  $V_L$  is equal to one in this case.

When the heritability is computed using the threshold model, DEMPSTER & LERNER (1950) and ROBERTSON (1950) derived an exact analytical formula relating this estimate to the observed data scale:

$$h_{obs}^2 = \frac{t^2}{p(1-p)} h_{thres}^2, \quad (\text{A6.23})$$

where  $p$  is the probability of occurrence of the minor phenotype and  $t$  is the density of a standard normal distribution at the  $p$ th quantile (see also ROFF, 1997). It can be shown (see SI, section A) that this formula is an exact analytical solution to Eq. A6.5 to A6.19 in the case of a GLMM with binomial distribution and a probit link.

When fixed effects are included in the model, it is still possible to use these formulae by integration over the marginalised predictions (see SI, section A). Note that this expression applies only

to analyses conducted with a probit link, it does not apply to a binomial model with a logit link function.

### ► Poisson distribution with a logarithm link

For a log link function and a Poisson distribution, both the derivative of the inverse link function, and the variance of the distribution, are equal to the expected value. Consequently, analytical results are obtainable for a log/Poisson model for quantities such as broad-sense and narrow-sense heritabilities. FOULLEY & IM (1993) derived an analytical formula to compute heritability on the observed scale:

$$h_{\text{obs}}^2 = \frac{\lambda^2 V_A}{\lambda^2 [\exp(V_A + V_{\text{RE}} + V_O) - 1] + \lambda} = \frac{\lambda V_A}{\lambda [\exp(V_A + V_{\text{RE}} + V_O) - 1] + 1}, \quad (\text{A6.24})$$

where  $\lambda$  is the observed phenotypic mean, computed analytically as:

$$\lambda = \exp\left(\mu + \frac{V_A + V_{\text{RE}} + V_O}{2}\right). \quad (\text{A6.25})$$

Again, it can be shown (see SI, section B) that these formulae are exact solutions to Eq. A6.5 to A6.19 when assuming a Poisson distribution with a log link. The inclusion of fixed effects in the model make the expression slightly more complicated (see SI, section B).

These results can also be extended to the Negative-Binomial distribution with log link with slight modifications of the analytical expressions (see SI, section B).

The broad-sense heritability estimate can be computed as any intra-class correlation coefficient (i.e. repeatability) for this kind of model (FOULLEY & IM, 1993; NAKAGAWA & SCHIELZETH, 2010) :

$$H_{\text{obs}}^2 = \frac{V(E[z|a])}{V_{\text{P,obs}}} = \frac{\lambda(\exp(V_A) - 1)}{\lambda [\exp(V_A + V_{\text{RE}} + V_O) - 1] + 1}. \quad (\text{A6.26})$$

Note that the broad-sense (Eq. A6.26) and narrow-sense (Eq. A6.24) heritabilities converge for small values of  $V_A$ .

## Example analysis: quantitative genetic parameters of a non-normal character

We modelled the first year survival of Soay sheep (*Ovis aries*) lambs on St Kilda, Outer Hebrides, Scotland. The data are comprised of 3814 individuals born between 1985 and 2011, and that are known to either have died in their first year, defined operationally as having died before the first of April in the year following their birth, or were known to have survived beyond their first year. Months of mortality for sheep of all ages are generally known from direct observation, and day of mortality is typically known. Furthermore, every lamb included in this analysis had a known sex and twin status (whether or not it had a twin), and a mother of a known age.

Pedigree information is available for the St Kilda Soay sheep study population. Maternal links are known from direct observation, with occasional inconsistencies corrected with genetic data. Paternal links are known from molecular data. Most paternity assignments are made with very high confidence, using a panel of 384 SNP markers, each with high minor allele frequencies, and spread evenly throughout the genome. Details of marker data and pedigree reconstruction are given in BÉRÉNOIS et al. (2014). The pedigree information was pruned to include only phenotyped individuals and their ancestors. The pedigree used in our analyses thus included 4687 individuals with 4165 maternal links and 4054 paternal links.

We fitted a generalised linear mixed model of survival, with a *logit* link function and a binomial distribution function. We included fixed effects of individual's sex and twin status, and linear, quadratic, and cubic effects of maternal age (*matAge<sub>i</sub>*). Maternal age was mean-centred by subtracting

**TAB. A6.1:** Parameters from the GLMM-based quantitative genetic analysis of Soay sheep (*Ovis aries*) lamb first-year survival. All estimates are reported as posterior modes with 95% credible intervals. The intercept in this model is arbitrarily defined for female lambs without twins, born to average age (4.8 years) mothers.

Parameter	Posterior mode with 95% CI
<b>(a) Fixed effects</b>	
Intercept	2.686 (1.631 – 3.299)
Sex (male vs. female)	-1.185 (-1.436 – -0.932)
Twin (twin vs. singleton)	-2.383 (-3.315 – -1.760)
Maternal age, linear term	0.238 (0.092 – 0.384)
Maternal age, quadratic term	-0.169 (-0.196 – -0.148)
Maternal age, cubic term	0.014 (0.010 – 0.019)
Sex-twin interaction	0.497 (0.016 – 1.016)
Sex-maternal age interaction	-0.020 (-0.103 – 0.070)
<b>(b) Random effects</b>	
$V_A$	0.882 (0.256 – 1.542)
$V_{\text{mother}}$	0.467 (0.213 – 0.876)
$V_{\text{year}}$	3.062 (1.814 – 5.635)

the overall mean. We also included an interaction of sex and twin status, and an interaction of twin status with maternal age. We included random effects of breeding value (as for Eq. A6.2), maternal identity, and birth year. Because the overdispersion variance  $V_O$  in a binomial GLMM is unobservable for binary data, we set its variance to one. The model was fitted in MCMCGLMM (HADFIELD, 2010), with diffuse independent normal priors on all fixed effects, and parameter-expanded priors for the variances of all estimated random effects.

We identified important effects on individual survival probability, i.e., several fixed effects were substantial, and also, each of the additive genetic, maternal, and among-year random effects explained appreciable variances (Table A6.1). The model intercept corresponds to the expected value on the latent scale of a female singleton (i.e. not a twin) lamb with an average age (4.8 years) mother. Males have lower survival than females, and twins have lower survival than singletons.

There were also substantial effects of maternal age, corresponding to a rapid increase in lamb survival with maternal age among relatively young mothers, and a negative curvature, such that the maximum survival probabilities occur among offspring of mothers aged 6 or 7 years. The trajectory of maternal age effects in the cubic model are similar to those obtained when maternal age is fitted as a multi-level effect.

To illustrate the consequences of accounting for different fixed effects in expected and observed data scale inferences, we calculated several parameters under a series of different treatments of the latent scale parameters of the GLMM. We calculated the phenotypic mean, the additive genetic variance, the total variance of expected values, the total variance of observed values, and the heritability of survival on the expected and observed scales.

First, we calculated parameters using only the model intercept ( $\mu$  in Eq. A6.1 and A6.3a). In general, linear modelling software will essentially arbitrarily define a model's intercept. In the current case, due to the details of how the data were coded, the intercept is the latent scale prediction for female singletons with average aged (4.8 years) mothers. In an average year, singleton females with average aged mothers have a probability of survival of about 80%. The additive genetic variance  $V_{A,\text{obs}}$ , calculated with Eq. A6.18 is about 0.005, and corresponds to heritabilities on the expected and observed scales of 0.115 and 0.042 (Table A6.2).



**TAB. A6.2:** Estimates of expected and observed scale phenotypic mean and variances, and additive genetic variance, for three different treatments of the fixed effects, as modelled on the linear scale with a GLMM, and reported in TAB. A6.1.

Quantity	Arbitrary intercept (singleton female)	Arbitrary intercept (twin male)	Averaging over all fixed effects
$\bar{z}$	0.838 (0.721 – 0.887)	0.352 (0.220 – 0.474)	0.446 (0.334 – 0.511)
$V_{A, \text{obs}}$	0.006 (0.002 – 0.015)	0.011 (0.005 – 0.023)	0.013 (0.005 – 0.021)
$V_{P, \text{exp}}$	0.060 (0.034 – 0.095)	0.098 (0.072 – 0.124)	0.123 (0.106 – 0.137)
$V_{P, \text{obs}}$	0.136 (0.108 – 0.206)	0.250 (0.183 – 0.250)	0.250 (0.227 – 0.250)
$h_{\text{exp}}^2$	0.109 (0.043 – 0.201)	0.122 (0.054 – 0.227)	0.102 (0.039 – 0.166)
$h_{\text{obs}}^2$	0.047 (0.017 – 0.085)	0.051 (0.022 – 0.101)	0.043 (0.020 – 0.087)

In contrast, if we wanted to calculate parameters using a different (but potentially equally arbitrary) intercept, corresponding to male twin males, we would obtain a mean survival rate of 0.32, an additive genetic variance that is twice as large, but similar heritabilities (Table A6.1). Note that we have not modelled any systematic differences in genetic parameters between females and males, or between singletons and twins. These differences in parameter estimates arise from the exact same estimated variance components on the latent scale, as a result of different fixed effects.

This first comparison has illustrated a major way in which the fixed effects in a GLMM influence inferences on the expected and observed value scales. For linear mixed models, it has been noted that variance in the response is explained by the fixed predictors, and that this may inappropriately reduce the phenotypic variance and inflate heritability estimates for some purposes (WILSON, 2008). However, in the example so far, we have simply considered two different intercepts (i.e. no difference in explained variance): female singletons vs male twins, in both cases, assuming focal groups of individuals are all born to average aged mothers. Again these differences in phenotypic variances and heritabilities arise from differences in intercepts, not any differences in variance explained by fixed effects.

All parameters on the expected and observed value scales are dependent on the intercept, including the mean, the additive genetic variance and the total variance generated from random effects. Heritability is modestly affected by the intercept, because additive genetic and total variances are similarly, but not identically, influenced by the model intercept.

Additive genetic effects are those arising from the average effect of alleles on phenotype, integrated over all background genetic and environmental circumstances in which alternate alleles might occur. Fixed effects are part of this background. Following, for example, Eq. A6.6 and A6.17, we can integrate our calculation of  $\Psi$  and ultimately  $V_{A, \text{obs}}$  over all fixed effects.

Considering all fixed and random effects, quantitative genetic parameters on the expected and observed scales are given in TAB. A6.2, 3rd column. Note that additive genetic variance is not intermediate between the two extremes (concerning sex and twin status), that we previously considered. The calculation of  $V_{A, \text{obs}}$  now includes an average slope calculated over a wide range of the steep part of the inverse-link function (near 0 on the latent scale, and near 0.5 on the expected value scale), and so is relatively high. The observed total phenotypic variance  $V_{P, \text{obs}}$  is also quite high.

The increase in  $V_{P, \text{obs}}$  has two causes. First the survival mean is closer to 0.5, so the random effects variance is now manifested as greater total variance on the expected and observed scales. Second, there is now variance arising from fixed effects that is included in the total variance.

## Evolutionary prediction

Systems for predicting adaptive evolution in response to phenotypic selection assume that the distribution of breeding values is multivariate normal, and in most applications, that the joint distri-

bution of phenotypes and breeding values is multivariate normal (LANDE, 1979 ; LANDE & ARNOLD, 1983 ; MORRISSEY, 2014 ; WALSH & LYNCH, p.d.).

Breeding values on the expected and observed scales will not be normal in GLMMs. Breeding values are normal by construction on the latent scale, and the non-linear (inverse) link functions cause the corresponding distributions on the expected and observed scales to be non-normal. Consequently, even with quantities such as additive genetic variances, heritabilities, etc., calculated on the expected and observed data scales, evolutionary predictions using statistical genetic machinery developed assuming normality will not hold. The Breeder's and Lande equations may hold approximately, and may perhaps be useful.

However, having taken up the non-trivial task of pursuing GLMM-based quantitative genetic analysis, the investigator has at their disposal inferences on the latent scale. On this scale, the assumptions required to predict the evolution of quantitative traits hold. In this section we will first demonstrate by simulation how application of the Breeder's equation will generate biased predictions of evolution.

We then proceed to an exposition of some statistical machinery that can be used to predict evolution on the latent scale (from which evolution on the expected and observed scale can subsequently be calculated, using Eq. A6.5), given inference of the function relating traits to fitness.

### Direct application of the Breeder's and Lande equations on the data scale

In order to explore the predictions of the Breeder's equation applied at the level of observed phenotype, we conducted a simulation in which phenotypes were generated according to a Poisson GLMM (Eq. A6.3a to A6.3c, with a Poisson distribution function and a log link function), and then selected the largest observed count values (positive selection) with a range of proportions of selected individuals (from 5% to 95%, creating a range of selection differentials), a range of latent-scale heritabilities (0.1, 0.3, 0.5 and 0.8, with a latent phenotypic variance fixed to 0.1), and a range of latent means  $\mu$  (from 0 to 3). We simulated 10,000 replicates of each scenario, each composed of a different array of 10,000 individuals.

For each simulation, we simulated 10,000 offspring. For each offspring, a breeding value was simulated according to  $a_{l,i} \sim \mathcal{N}((a_{l,d} + a_{l,s})/2, V_A/2)$ , where  $a_{l,i}$  is the focal individual's breeding value,  $a_{l,d}$  and  $a_{l,s}$  are the breeding values of simulated dams and sires and  $V_A/2$  represents the segregational variance assuming parents are not inbred. Dams and sires were individuals chosen at random from among the pool of simulated selected individuals. For each scenario, we calculated the realised selection differential arising from the simulated truncation selection,  $S_{\text{obs}}$ , and the average evolutionary response across simulations,  $R_{\text{obs}}$ . For each scenario, we calculated the heritability on the observed scale using Eq. A6.19. If the Breeder's equation was strictly valid for a Poisson GLMM on the observed scale, the realised heritability  $R_{\text{obs}}/S_{\text{obs}}$  would be equal to the observed-scale heritability  $h_{\text{obs}}^2$ .

The correspondence between  $R_{\text{obs}}/S_{\text{obs}}$  and  $h_{\text{obs}}^2$  is only approximate (FIG. A6.2), and strongly depends on the selection differential (controlled here by the proportion of selected individuals). In particular, evolutionary predictions are poorest in absolute terms for large  $\mu$  and large (latent) heritabilities. However, because we were analysing simulation data, we could track the selection differential of latent value (by calculating the difference in its mean between simulated survivors and the mean simulated before selection). We can also calculate the mean latent breeding value after selection. Across all simulation scenarios, the ratio of the change in breeding value after selection, to the change in breeding value before selection was equal to the latent heritability (see FIG. A6.2), showing that evolutionary changes could be accurately predicted on the latent scale.

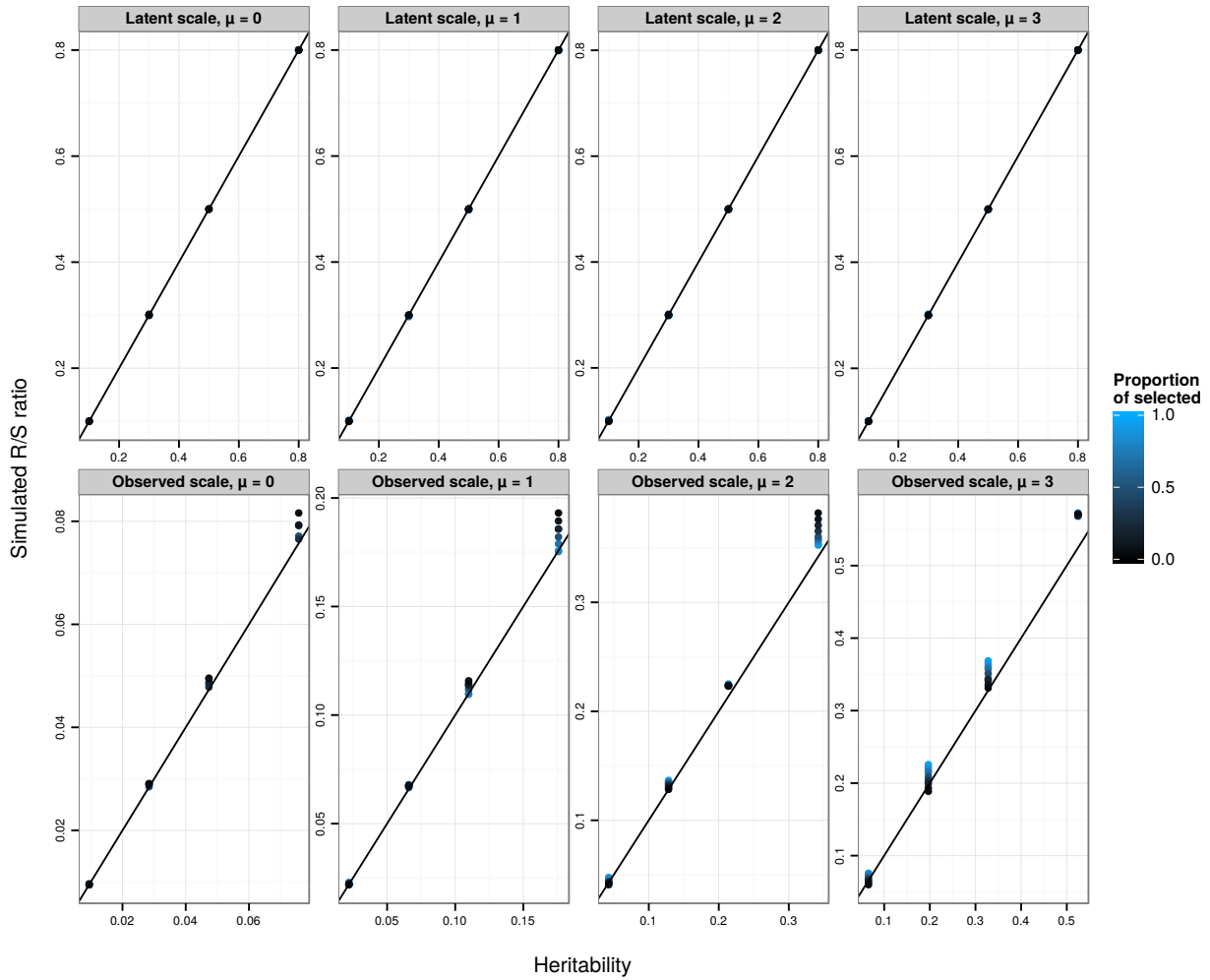


FIG. A6.2: Simulated  $R/S$  (evolutionary response over selection differential, or the realised heritability) on the latent (upper panels) or observed data (lower panels) scales against the corresponding-scale heritabilities. Each data point is the average over 10,000 replicates of 10,000 individuals for various latent heritabilities  $h_{\text{lat}}^2$  (0.1, 0.3, 0.5, 0.8), latent population mean ( $\mu$  from 0 to 3, from left to right) and proportion of selected individuals (5%, 10%, 20%, 30%, 50%, 70%, 80%, 90%, 95%, varying from black to blue). The 1:1 line is plotted in black. The breeder's equation is predictive on the latent scale (upper panels), but approximate on the observed data scale (lower panels), because phenotypes and breeding values are not jointly multivariate normal on that scale.

## Evolutionary change on the latent scale, and associated change on the expected and observed scales

In an analysis of real data, latent breeding values are, of course, not measured. However, given an estimate of the effect of traits on fitness, say via regression analysis, we can derive the parameters necessary to predict evolution on the latent scale. The idea is thus to relate measured fitness on the observed data scale to the latent scale, compute the evolutionary response on the latent scale, then to compute the evolutionary response on the observed data scale.

To relate the measured fitness on the observed scale to the latent scale, we need to compute the expected fitness  $W_{\text{exp}}$  given latent trait value  $l$ , which is

$$W_{\text{exp}}(l) = \sum_k W_P(k)P(Z = k|l), \quad (\text{A6.27})$$

where  $W_p(k)$  is the measure of fitness for the  $k$ th data scale category (assuming the observed data scale is discrete). Population mean fitness, can then be calculated in an analogous way to Eq. A6.5:

$$\bar{W} = \int W_{\text{exp}}(l) f_{\mathcal{N}}(l, \mu, V_A + V_{\text{RE}} + V_O) dl. \quad (\text{A6.28})$$

These expressions comprise the basic functions necessary to predict evolution. Given a fitted GLMM, and a given estimate of the fitness function  $W_p(k)$ , each of several approaches could give identical results. For simplicity, we proceed via application of the breeder's equation at the level of the latent scale.

The change in the mean genetic value of any character due to selection is equal to the covariance of breeding value with relative fitness (ROBERTSON, 1966 ; ROBERTSON, 1968). Using Stein's (1973) lemma again, this covariance can be obtained as the product of the additive genetic variance of latent values and the average derivative of expected fitness with respect to latent value, i.e.,  $E \left[ \frac{dW_{\text{exp}}}{dl} \right]$ . Evolution on the latent scale can therefore be predicted by

$$\Delta\mu = V_A E \left[ \frac{dW_{\text{exp}}}{dl} \right] \frac{1}{\bar{W}}. \quad (\text{A6.29})$$

In the case of a multivariate analysis, note that the derivative above should be a vector of partial derivatives (first order partial derivative for each trait).

If fixed effects need to be considered, the approach can be modified in the same way as integration over fixed effects is accomplished for calculating other quantities, i.e. the expression

$$\bar{W} = \frac{1}{N} \sum_{i=1}^N \int W_{\text{exp}}(l) f_{\mathcal{N}}(l, \mu + \hat{\mathbf{l}}_i, V_A + V_{\text{RE}} + V_O) dl, \quad (\text{A6.30})$$

would be used in calculations of mean fitness and the average derivative of expected fitness with respect to latent value.

Evolution on the observed data scale is calculated as

$$\Delta\bar{z} = \int g^{-1}(l) f_{\mathcal{N}}(l, \mu + \Delta\mu, V_A + V_{\text{RE}} + V_O) dl - \bar{z}. \quad (\text{A6.31})$$

Or, if fixed effects are included in the model:

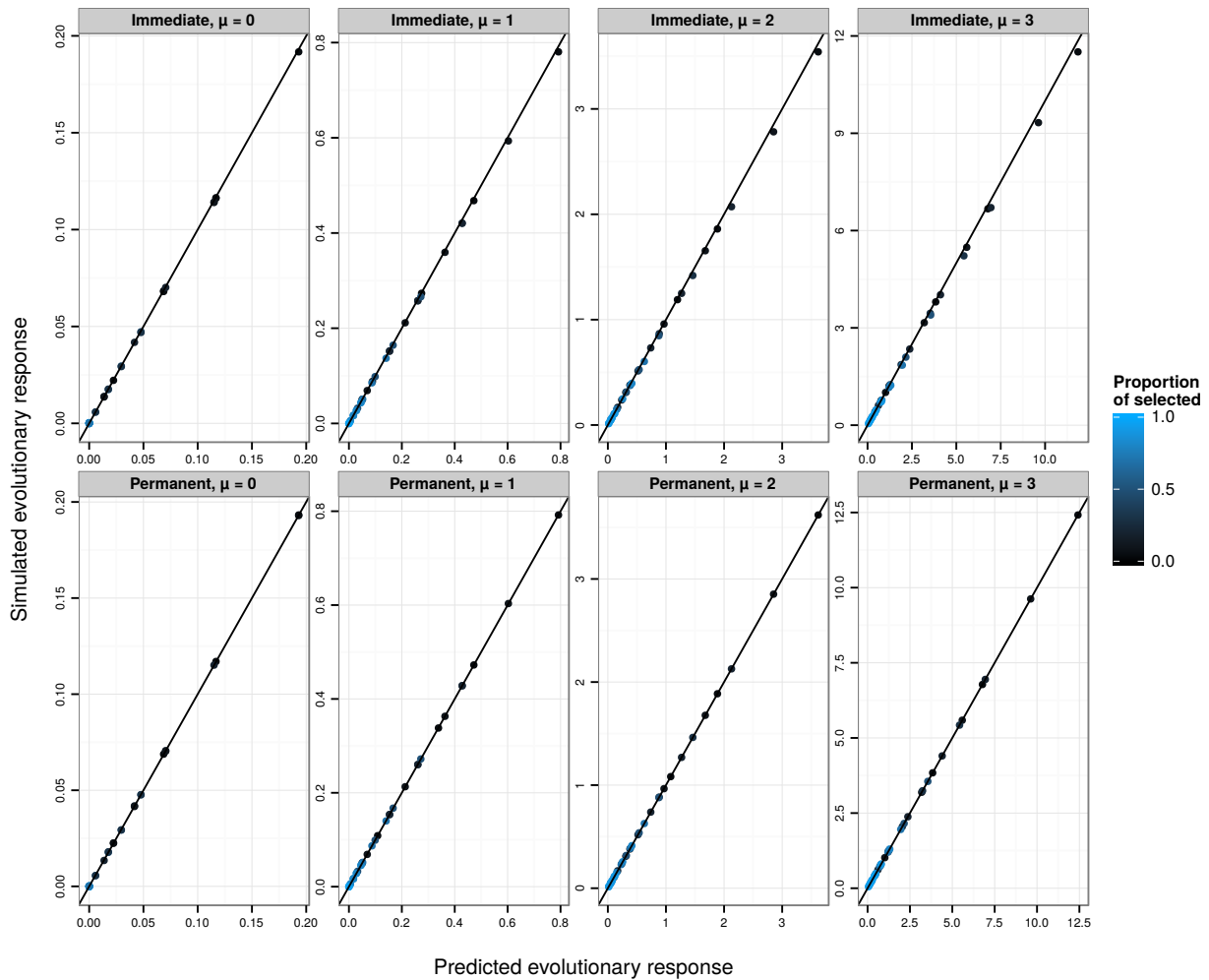
$$\Delta\bar{z} = \frac{1}{N} \sum_{i=1}^N \int g^{-1}(l) f_{\mathcal{N}}(l, \mu + \hat{\mathbf{l}}_i + \Delta\mu, V_A + V_{\text{RE}} + V_O) dl - \bar{z}. \quad (\text{A6.32})$$

Note that, in this second equation,  $\bar{z}$  must be computed as in Eq. A6.6 and that this equation assumes that the distribution of fixed effects for the offspring generation is the same as for the parental one.

Another derivation of the expected evolutionary response using the Price-Robertson identity (ROBERTSON, 1966 ; PRICE, 1970) is given in the Supplementary Information (section C).

## The simulation study revisited

Using the same replicates as in the simulation study above (FIG. A6.2, top row), we used Eq. A6.27 to A6.32 to predict phenotypic evolution. This procedure provides greatly improved predictions of evolutionary change on the observed scale (FIG. A6.3). However, somewhat less response to selection is observed than is predicted. This behaviour occurs because, in addition to producing a permanent evolutionary response in the mean value on the latent scale, directional selection creates a transient reduction of additive genetic variance due to the linkage disequilibrium.



**FIG. A6.3:** Predicted  $R_{\text{obs}}$  (phenotypic evolutionary response on the observed scale, see Eq. A6.32) against the simulated  $R_{\text{obs}}$ , via evolutionary predictions applied on the latent scale. Each data point is the average over 10,000 replicates of 10,000 individuals for various latent heritabilities  $h_{\text{lat}}^2$  (0.1, 0.3, 0.5, 0.8), latent population mean ( $\mu$  from 0 to 3) and proportion of selected individuals (5%, 10%, 20%, 30%, 50%, 70%, 80%, 90%, 95%, varying from black to blue). The 1:1 line is plotted in black. The upper panels (“Immediate”) show simulations for the response after a single generation, which include both a permanent and transient response to selection arising from linkage disequilibrium. The bottom panels (“permanent”) show simulation results modified to reflect only the permanent response to selection.

Because the link function is non-linear, this transient change in the variance on the latent scale generates a transient change in the mean on the expected and observed scales. Following several generations of random mating, the evolutionary change on the observed scale would converge on the predicted values.

We simulated such a generation at equilibrium by simply drawing breeding values for the post-selection sample from a distribution with the same variance as in the parental generation. This procedure necessarily generated a strong match between predicted and simulated evolution (FIG. A6.3, bottom row). Additionally, the effects of transient reduction in genetic variance on the latent scale could be directly modelled, for example, using Bulmer’s (1971) approximations for the transient dynamics of the genetic variance in response to selection.

## Discussion

The expressions given here for quantitative genetic parameters on the expected and observed data scales are exact, given the GLMM model assumptions, in two senses. First, they are not approximations, such as might be obtained by linear approximations (BROWNE et al., 2005). Second, they are expressions for the parameters of direct interest, rather than convenient substitutes. For example, the common calculation of variance partition coefficients (intraclass correlations) on an underlying scale with a logistic distribution, (as also suggested by BROWNE et al. 2005) provides a value of the broad-sense heritability when applied to genetic parameters estimated in a logistic GLMM. The expressions given here can provide quantitative genetic inferences of the additive genetic parameters, and on the scale on which the traits are observed. The whole framework developed here (including univariate and multivariate parameters computation and evolutionary predictions on the observed data scale) is implemented in the R package QGGLMM available at <https://github.com/devillemereuil/qgglmm>.

While the calculations we provide will often (i.e. when no analytical formula exists) be more computationally demanding than available alternatives based on approximations, intra-class correlations (broad-sense heritabilities) or calculations on different scales, they will be direct ascertainment of specific parameters of interest, since the scale of evolutionary interest is likely to be the observed data scale, rather than the latent scale (unless some artificial selection is applied to predicted latent breeding values as in modern animal breeding). Most applications should not be onerous. Computations of means and (additive genetic) variances took less than a second on a 1.7 GHz processor when using our R functions on the Soay sheep data set. Summation over fixed effects, and integration over 1000 posterior samples of the fitted model took several minutes. When analytical expressions are available (e.g. for Poisson/log, Binomial/probit and Negative-Binomial/log, see the supplementary information and R package documentation), these computations are considerably accelerated.

We have highlighted additional and important ways in which fixed effects influence quantitative genetic inferences with GLMMs, and developed an approach for handling these complexities. In LMMs, the main consideration pertaining to fixed effects is that they explain variance, and some or all of this variance might be inappropriate to exclude from an assessment of  $V_P$  when calculating heritabilities (WILSON, 2008). This aspect of fixed effects is relevant to GLMMs, but furthermore, all parameters on the expected and observed scales, not just means, are influenced by fixed effects in GLMMs; this includes additive genetic and phenotypic variances. This fact necessitates particular care in interpreting GLMMs. A more practical guide to navigating the complexities involving fixed effects in LMM- and GLMM-based quantitative genetic inferences is presented in (manuscript in prep).

In our example analysis in Soay sheep,  $V_A$  and  $V_P$  changed substantially depending on different treatments of the fixed effects (especially, arbitrary different definitions of the model's latent intercept). We do not intend to suggest that any of these treatments of the fixed effects is correct or wrong. Rather, any of the analyses we presented (and many other conceivable variations) may be appropriate to any particular purpose. For this particular case, different treatments of fixed effects changed  $V_A$  and  $V_P$  in roughly, but not exactly, similar proportions. Consequently, heritabilities on the data scales were not greatly different among treatments. This need not necessarily be the case in all quantitative genetic analyses using GLMMs, although it is likely to be common in binomial response models.

Currently, with the increasing applicability of GLMMs, investigators seem eager to convert to the observed data scale. It seems clear that conversions between scales are generally useful. However, it is of note that assumptions required of quantitative genetic methods for evolutionary prediction hold on the latent scale under the assumptions of the GLMM. Therefore, given an estimate of a fitness function, no further assumptions are necessary to predict evolution via the latent scale (as with equations A6.27, A6.29, and A6.31), over and above those that are made in the first place upon deciding to pursue GLMM-based quantitative genetic analysis. The approach we suggest treats the relationships between the levels of a GLMM as a simple developmental system, and the approach

described here is essentially the general theory laid out in MORRISSEY (2015), with specific extensions to handle distribution functions and integration over fixed effects.

## Acknowledgements

PdV was supported by a doctoral studentship from the French *Ministère de la Recherche et de l'Enseignement Supérieur*. HS was supported by an Emmy Noether fellowship from the German Research Foundation (DFG; SCHI 1188/1-1/2). SN is supported by an Future Fellowship, Australia (FT130100268). MBM is supported by a University Research Fellowship from the Royal Society (London). The Soay sheep data were provided by Josephine Pemberton and Loeske Kruuk, and were collected primarily by Jill Pilkington and Andrew MacColl with the help of many volunteers. The collection of the Soay sheep data is supported by the National Trust for Scotland and QinetQ, with funding from NERC, the Royal Society, and the Leverhulme Trust. We thank Kerry Johnson, Paul Johnson, Alastair Wilson, Loeske Kruuk and Josephine Pemberton for valuable discussions and comments on this manuscript.

# ANNEXE S4

## Supplementary Information (ARTICLE 4)

TABLE S4.1 : Location of the weather stations used to impute temperature data.

Station ID	Latitude	Longitude	Mountain range	Distance to the closest site
207	45.12973N	5.58719E	Vercors	3 km
209	45.02010N	5.61786E	Vercors	4 km
218	45.02015N	6.63496E	Lautaret-Galibier	19 km
29-217	44.83946N	6.49312E	Lautaret-Galibier	22 km

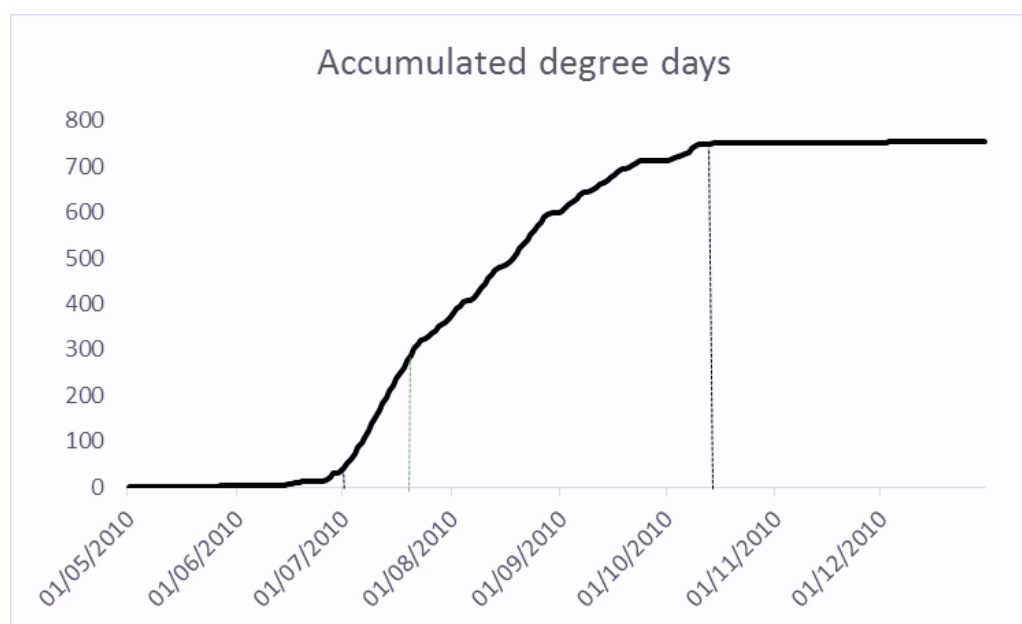


FIG. S4.1 : Example of an accumulated degree days curve used to define the limits of the vegetative season (vertical black lines). The date of census is also showed (green vertical line).

TABLE S4.2 : Dates of beginning and end of the vegetative season and dates of census for the Vercors population.

Site	Quadrat	Year	Beginning of vegetative season	Census date	End of vegetative season
BRU	1	2009	31 March 2009	18 June 2009	01 December 2009
BRU	1	2010	21 March 2010	07 June 2010	15 November 2010
BRU	1	2011	11 March 2011	24 May 2011	NA
BRU	1	2012	11 March 2012	30 May 2012	01 December 2012
BRU	1	2013	01 April 2013	11 July 2013	10 November 2013



BRU	1	2014	01 April 2014	19 June 2014	10 November 2014
BRU	2	2009	29 March 2009	18 June 2009	05 December 2009
BRU	2	2010	20 March 2010	07 June 2010	15 November 2010
BRU	2	2011	12 March 2011	24 May 2011	NA
BRU	2	2012	12 March 2012	30 May 2012	01 December 2012
BRU	2	2013	01 April 2013	11 July 2013	10 November 2013
BRU	2	2014	01 April 2014	19 June 2014	10 November 2014
CHA	1	2009	28 March 2009	15 July 2009	25 November 2009
CHA	1	2010	19 March 2010	09 June 2010	20 November 2010
CHA	1	2011	20 March 2011	28 June 2011	NA
CHA	1	2012	20 March 2012	26 June 2012	20 October 2012
CHA	1	2013	12 April 2013	18 July 2013	NA
CHA	2	2009	28 March 2009	15 July 2009	25 November 2009
CHA	2	2010	15 March 2010	09 June 2010	20 November 2010
CHA	2	2011	16 March 2011	28 June 2011	NA
CHA	2	2012	16 March 2012	26 June 2012	20 October 2012
CHA	2	2013	12 April 2013	18 July 2013	NA
CHA	3	2009	07 May 2009	16 July 2009	25 November 2009
CHA	3	2010	24 April 2010	09 June 2010	20 November 2010
CHA	3	2011	31 March 2011	28 June 2011	NA
CHA	3	2012	31 March 2012	26 June 2012	20 October 2012
CHA	3	2013	03 May 2013	18 July 2013	NA
VIL	1	2009	03 May 2009	30 July 2009	01 December 2009
VIL	1	2010	28 April 2010	13 July 2010	01 November 2010
VIL	1	2011	06 April 2011	15 July 2011	01 December 2011
VIL	1	2012	26 March 2012	09 July 2012	01 December 2012
VIL	1	2013	23 April 2013	22 July 2013	30 October 2013
VIL	1	2014	15 April 2014	18 July 2014	30 October 2014
VIL	2	2009	09 May 2009	29 July 2009	01 December 2009
VIL	2	2010	28 April 2010	13 July 2010	15 October 2010
VIL	2	2011	09 April 2011	15 July 2011	01 December 2011
VIL	2	2012	24 March 2012	09 July 2012	01 December 2012
VIL	2	2013	23 April 2013	22 July 2013	30 October 2013
VIL	2	2014	15 April 2014	18 July 2014	30 October 2014
VIL	3	2009	06 May 2009	29 July 2009	01 December 2009
VIL	3	2010	27 April 2010	13 July 2010	15 November 2010
VIL	3	2011	09 April 2011	15 July 2011	01 December 2011
VIL	3	2012	25 March 2012	09 July 2012	01 December 2012
VIL	3	2013	23 April 2013	22 July 2013	30 October 2013
VIL	3	2014	15 April 2014	18 July 2014	30 October 2014
LAU	1	2009	10 June 2009	05 August 2009	15 October 2009
LAU	1	2010	09 June 2010	18 July 2010	15 October 2010
LAU	1	2011	13 June 2011	10 August 2011	17 October 2011
LAU	1	2012	01 June 2012	23 July 2012	NA
LAU	1	2013	01 June 2013	15 August 2013	NA
LAU	1	2014	20 May 2014	06 August 2014	NA
LAU	2	2009	11 June 2009	05 August 2009	15 October 2009
LAU	2	2010	09 June 2010	18 July 2010	15 October 2010
LAU	2	2011	14 June 2011	10 August 2011	17 October 2011
LAU	2	2012	31 May 2012	23 July 2012	NA
LAU	2	2013	31 May 2013	15 August 2013	NA
LAU	2	2014	20 May 2014	06 August 2014	NA
LAU	3	2009	31 May 2009	05 August 2009	15 October 2009
LAU	3	2010	11 June 2010	18 July 2010	15 October 2010
LAU	3	2011	31 May 2011	19 July 2011	17 October 2011
LAU	3	2012	01 June 2012	23 July 2012	NA
LAU	3	2013	01 June 2013	15 August 2013	NA
LAU	3	2014	20 May 2014	06 August 2014	NA
GAL	1	2009	30 May 2009	06 August 2009	09 October 2009
GAL	1	2010	13 June 2010	19 July 2010	12 October 2010
GAL	1	2011	05 June 2011	18 July 2011	20 October 2011
GAL	1	2012	15 June 2012	24 July 2012	01 November 2012
GAL	1	2013	28 June 2013	15 August 2013	09 October 2013
GAL	1	2014	19 June 2014	06 August 2014	09 October 2014
GAL	2	2009	17 June 2009	06 August 2009	09 October 2009
GAL	2	2010	02 July 2010	19 July 2010	12 October 2010
GAL	2	2011	16 June 2011	18 July 2011	20 October 2011
GAL	2	2012	28 June 2012	24 July 2012	01 November 2012

GAL	2	2013	06 July 2013	15 August 2013	09 October 2013
GAL	2	2014	19 June 2014	06 August 2014	09 October 2014
GAL	3	2009	14 June 2009	06 August 2009	09 October 2009
GAL	3	2010	13 June 2010	19 July 2010	12 October 2010
GAL	3	2011	14 June 2011	18 July 2011	20 October 2011
GAL	3	2012	22 June 2012	24 July 2012	01 November 2012
GAL	3	2013	06 July 2013	15 August 2013	09 October 2013
GAL	3	2014	19 June 2014	06 August 2014	09 October 2014
PIC	1	2009	16 May 2009	11 August 2009	01 November 2009
PIC	1	2010	23 April 2010	03 August 2010	20 October 2010
PIC	1	2011	04 April 2011	10 August 2011	20 October 2011
PIC	1	2012	09 May 2012	24 July 2012	25 October 2012
PIC	1	2013	12 June 2013	16 August 2013	10 October 2013
PIC	1	2014	07 June 2014	07 August 2014	10 October 2014
PIC	2	2009	18 June 2009	11 August 2009	01 November 2009
PIC	2	2010	11 June 2010	03 August 2010	20 October 2010
PIC	2	2011	23 May 2011	10 August 2011	20 October 2011
PIC	2	2012	17 June 2012	24 July 2012	25 October 2012
PIC	2	2013	12 June 2013	16 August 2013	10 October 2013
PIC	2	2014	21 June 2014	07 August 2014	10 October 2014
PIC	3	2009	05 May 2009	11 August 2009	01 November 2009
PIC	3	2010	23 May 2010	03 August 2010	20 October 2010
PIC	3	2011	01 April 2011	10 August 2011	20 October 2011
PIC	3	2012	09 May 2012	24 July 2012	25 October 2012
PIC	3	2013	12 June 2013	16 August 2013	10 October 2013
PIC	3	2014	07 June 2014	07 August 2014	10 October 2014
PIC	4	2009	15 May 2009	11 August 2009	01 November 2009
PIC	4	2010	23 May 2010	03 August 2010	20 October 2010
PIC	4	2011	04 April 2011	10 August 2011	20 October 2011
PIC	4	2012	10 May 2012	24 July 2012	25 October 2012
PIC	4	2013	12 June 2013	16 August 2013	10 October 2013
PIC	4	2014	07 June 2014	07 August 2014	10 October 2014

TABLE S4.3 : Pearson's linear correlation coefficients between early season and late season thermic variables. Correlations higher than 0.7 are in bold.

	T.min.1	T.mean.1	T.max.1	DD.1	T.range.1	Frost.1	Frost.e.1
T.min.1	1.00	0.67	-0.02	0.09	-0.49	-0.66	-0.67
T.mean.1	0.67	1.00	<b>0.71</b>	0.45	0.30	-0.37	-0.36
T.max.1	-0.02	<b>0.71</b>	1.00	0.59	<b>0.88</b>	0.08	0.10
DD.1	0.09	0.45	0.59	1.00	0.47	0.10	0.10
T.range.1	-0.49	0.30	<b>0.88</b>	0.47	1.00	0.39	0.41
Frost.1	-0.66	-0.37	0.08	0.10	0.39	1.00	<b>0.99</b>
Frost.e.1	-0.67	-0.36	0.10	0.10	0.41	<b>0.99</b>	1.00
	T.min.2	T.mean.2	T.max.2	DD.2	T.range.2	Frost.2	Frost.e.2
T.min.2	1.00	<b>0.79</b>	0.12	<b>0.77</b>	-0.47	-0.55	-0.61
T.mean.2	<b>0.79</b>	1.00	0.69	<b>0.89</b>	0.16	-0.27	-0.31
T.max.2	0.12	0.69	1.00	0.52	<b>0.82</b>	0.14	0.15
DD.2	<b>0.77</b>	<b>0.89</b>	0.52	1.00	0.02	-0.23	-0.29
T.range.2	-0.47	0.16	<b>0.82</b>	0.02	1.00	0.44	0.49
Frost.2	-0.55	-0.27	0.14	-0.23	0.44	1.00	<b>0.98</b>
Frost.e.2	-0.61	-0.31	0.15	-0.29	0.49	<b>0.98</b>	1.00

# ANNEXE S5

## Supplementary Information (ARTICLE 5)

---

### S5.1 *Arabis alpina* morphotypes

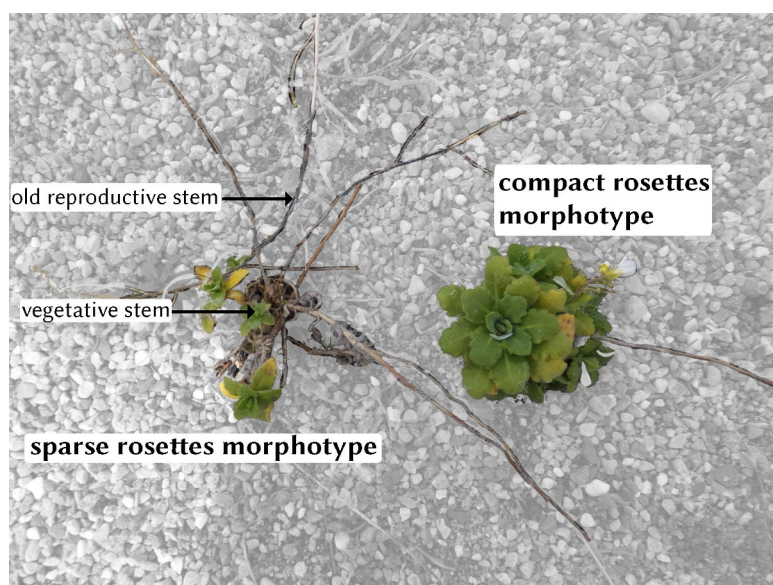


FIG. S5.1 : Picture showing two extreme morphotypes : the “sparse rosettes” morphotype (left) and the “compact rosettes” morphotype (right). This picture was shot during the end of the growing season, when the reproductive stems already died out. As a result, the green parts are the vegetative stems (though a flowering reproductive stem is visible on the right plant).

To analyse the morphotype, we classified the plants into four categories :

**Sparse rosettes** Very little number of vegetative stems in the plants.

**Intermediate** Intermediate phenotype between “sparse” and “numerous” rosettes

**Numerous rosettes** Numerous rosettes at the centre of the plant, but not spheric organisation.

**Compact rosettes** Numerous rosettes at the centre of the plant with a compact, spheric organisation.

Differences between these morphotypes are illustrated in Fig S5.1 and S5.2 (all pictures were taken late in the growing season, during biomass measurement). Fig. S5.1 shows two extreme morphotypes (“Sparse” and “Compact”) growing alongside in the Lautaret experimental garden. The “sparse” morphotype (Fig. S5.1, left) is composed of only 3 vegetative stems growing on the bare centre of the plant. Numerous dying-out reproductive stems are visible on the ground (background has been faded out to highlight them). In the “compact” morphotype (Fig. S5.1, right), vegetative stems are too

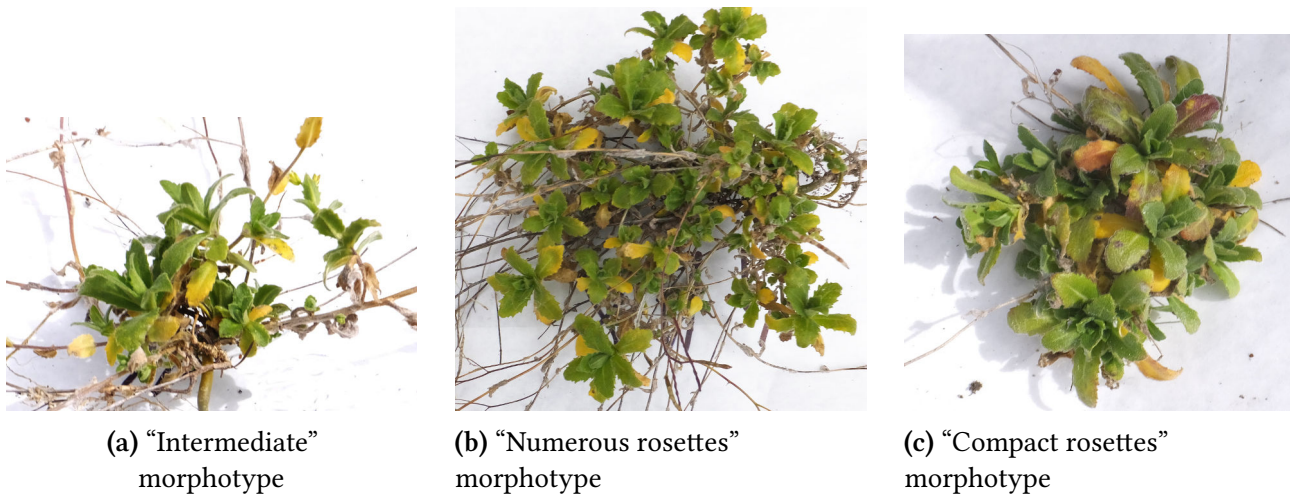


FIG. S5.2 : Different pictures of three different morphotypes : Intermediate (a), Numerous rosettes (b) and Compact rosettes (c).

numerous and too compact to be counted (although one upper rosette is mainly visible from above, this spheric area is composed of many more lateral ones).

Fig. S5.2 shows different morphotypes (“Intermediate”, “Numerous” and “Compact”) taken during biomass measurement. We see that the “Intermediate” (Fig. S5.2, a) morphotype has typically more vegetative stems than the sparse one, but still has a reduced number of them. The “Numerous” (Fig. S5.2, b) has many more vegetative stems, more or less compact, but not organised in a roughly spherical structure. The picture of the “compact” morphotype (Fig. S5.2, c) clearly shows this roughly spherical structure, with vegetative stems growing in a centrifugal way from the centre of the plant.

## S5.2 sNMF clustering

We performed a clustering using the sNMF software (FRICHOT et al., 2014). The clustering was first performed on the whole dataset to remove individuals with aberrant clustering (i.e. between-massif hybrids and individuals with no clustering signal). Then we performed the clustering again.

This latter run of sNMF yielded a likely number of cluster  $K = 6$  (Fig. S5.3), which corresponds to the number of studied populations. The ancestry coefficients showed a quite strict clustering between populations, with very little gene flow between populations (Fig. S5.4, only two individuals display signs of hybridisation).

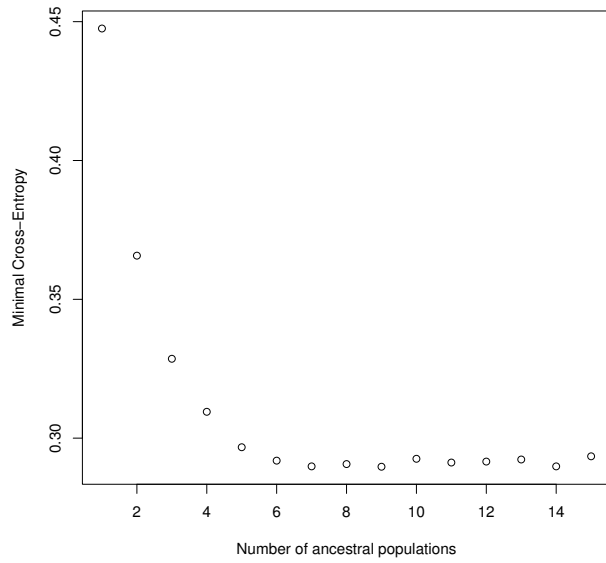


FIG. S5.3 : Cross-Entropy criterion for  $K$  varying between 1 and 15. The criterion plateaus at  $K = 6$ .

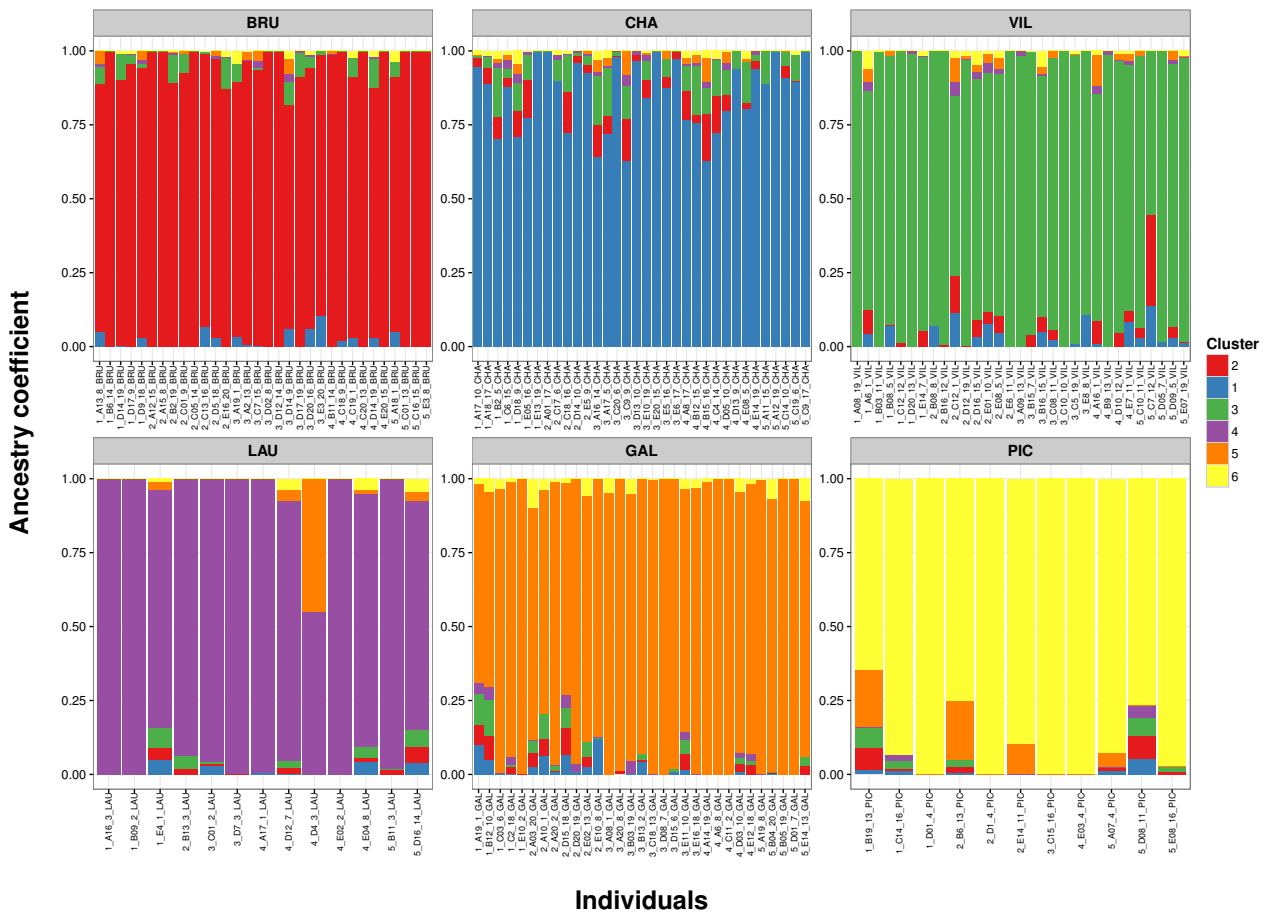


FIG. S5.4 : Output of sNMF for  $K = 6$  showing the ancestry coefficients for individuals of the six populations (BRU, CHA, VIL, LAU, GAL, PIC). Each cluster has a specific colour, here ordered according to the populations.

# ANNEXE S6

## Supplementary Information (ARTICLE 6)

---

Throughout this document, when the mean and variance of the Normal distribution probability and cumulative density functions are not mentioned, they are assumed equal to 0 and 1, respectively for the mean and variance. Hence, by default,  $f_{\mathcal{N}}(l)$  and  $\Phi_{\mathcal{N}}(l)$  correspond to a standard Normal distribution.

### S6.1 Proof for analytical solutions for binary traits

**Equivalence between a Binomial/Probit GLMM and the threshold model** We consider the probability of the observed data  $z$  to be 1, knowing that the latent trait value is  $l$  (hence from the GLMM perspective) :

$$P(z = 1) = \Phi_{\mathcal{N}}(l) = P(X < l) = P(0 < l - X), \quad (\text{S6.1})$$

where  $X$  is a random variable following a standard Normal distribution. Let  $\epsilon$  be a variable such that  $\epsilon = -X$ , then  $\epsilon$  also follows a standard Normal distribution. This allows us to rewrite the above equation as :

$$P(z = 1) = P(0 < l + \epsilon), \quad (\text{S6.2})$$

which is the probability defined from the threshold model perspective. The addition of the  $\epsilon$  term, which has a variance of 1, explains the appearance of the so-called “link variance”. Note that the exact same reasoning apply to the logit link function, which inverse function is the cumulative distribution function of a logistic distribution of location 0 and scale 1. Thus, the “link variance” associated with a logit link is  $\pi^3/2$ .

**Observed phenotypic mean** Let  $p$  be the average phenotype (as well as the probability of 1, which is an equivalent definition). Note that because of the equivalence between the GLMM and the threshold model :

$$p = \int \Phi_{\mathcal{N}}(l) f_{\mathcal{N}}(l, \mu, V_{\text{P}}) dl = 1 - \Phi_{\mathcal{N}}(0, \mu, V_{\text{P}} + 1), \quad (\text{S6.3})$$

which provides a quick analytical way to compute  $p$ . If fixed effects are included in the model,  $p$  is simply this calculation averaged over the elements of  $\mathbf{X}\hat{\mathbf{b}}$  :

$$p = 1 - \frac{1}{N} \sum_i \Phi_{\mathcal{N}}(0, \mu + (\mathbf{X}\hat{\mathbf{b}})_i, V_{\text{P}} + 1). \quad (\text{S6.4})$$

**Observed phenotypic variance** The observed variance of the phenotype only depends on the mean  $p$  and is  $p(1 - p)$ . This is a property of a binomial distribution with only one trial and is consistent

with EQ. A6.8 :

$$\begin{aligned}
 V_{P,obs} &= \int (\Phi_{\mathcal{N}}(l) - p)^2 f_{\mathcal{N}}(l, \mu, V_P) dl + \int v(l) f_{\mathcal{N}}(l, \mu, V_P) dl, \\
 &= \int \Phi_{\mathcal{N}}(l)^2 f_{\mathcal{N}}(l, \mu, V_P) dl - p^2 + \int \Phi_{\mathcal{N}}(l) f_{\mathcal{N}}(l, \mu, V_P) dl - \int \Phi_{\mathcal{N}}(l)^2 f_{\mathcal{N}}(l, \mu, V_P) dl, \\
 &= p - p^2, \\
 &= p(1 - p).
 \end{aligned} \tag{S6.5}$$

This very simple analytical solution allows to easily compute  $V_{P,obs}$ . Of course, when fixed effects are included,  $p$  should be computed using EQ. S6.4.

**Consistency with Dempster & Lerner equation** Using DEMPSTER & LERNER (1950) equation and the threshold model framework, one would compute the observed-scale heritability as :

$$h_{obs, DL}^2 = \frac{t^2}{p(1-p)} \frac{V_A}{V_P + 1} \tag{S6.6}$$

Note that the  $V_P + 1$  arise because of the addition of the so-called “link variance” (NAKAGAWA & SCHIELZETH, 2010). The term  $t$  is the probability density of a standard normal distribution evaluated at the  $p$ th quantile. Hence (standardising the latent distribution) :

$$t = f_{\mathcal{N}}(\Phi^{-1}(p)) = f_{\mathcal{N}}\left(\frac{\mu}{\sqrt{V_P + 1}}\right) \tag{S6.7}$$

On the other hand, using our framework, one would compute the observed heritability by computing  $\Psi$  as in EQ. A6.16, and combine it with EQ. A6.8 & A6.18 :

$$h_{obs}^2 = \frac{\Psi^2 V_A}{p(1-p)} \tag{S6.8}$$

By comparing EQ. S6.6 & S6.8, one can see the identity holds if, and only if :

$$\Psi = \frac{t}{\sqrt{V_P + 1}} = f_{\mathcal{N}}(0, \mu, V_P + 1) \tag{S6.9}$$

In order to prove this identity, let us compute the ratio between the two :

$$\begin{aligned}
 \Psi / f_{\mathcal{N}}(0, \mu, V_P + 1) &= \frac{\int f_{\mathcal{N}}(x) f_{\mathcal{N}}(x, \mu, V_P) dx}{f_{\mathcal{N}}(0, \mu, V_P + 1)} \\
 &= \sqrt{\frac{V_P + 1}{V_P}} \frac{1}{\sqrt{2\pi}} \int \exp\left[-\frac{1}{2} \left(x^2 + \frac{(x - \mu)^2}{V_P} - \frac{\mu^2}{V_P + 1}\right)\right] dx \\
 &= \sqrt{\frac{V_P + 1}{V_P}} \frac{1}{\sqrt{2\pi}} \int \exp\left[-\frac{1}{2} \left(\frac{(x + x V_P - \mu)^2}{V_P (V_P + 1)}\right)\right] dx \\
 &= \sqrt{\frac{V_P + 1}{V_P}} \frac{1}{\sqrt{2\pi}} \int \exp\left[-\frac{1}{2} \left(\frac{(x - (\mu/(V_P + 1)))^2}{V_P/(V_P + 1)}\right)\right] dx \\
 &= \int_1 f_{\mathcal{N}}\left(x, \frac{\mu}{V_P + 1}, \frac{V_P}{V_P + 1}\right) dx \\
 &= 1
 \end{aligned} \tag{S6.10}$$

Note that, if fixed effects are included in the model,  $\Psi$  can be computed by averaging over them, as in EQ. S6.4 :

$$\Psi = \frac{1}{N} \sum_i f_{\mathcal{N}}(0, \mu + (\mathbf{X}\hat{\mathbf{b}})_i, V_P + 1) \tag{S6.11}$$

## S6.2 Proof for analytical solutions for Poisson traits

**Observed phenotypic mean** The observed phenotypic mean, hereafter noted  $\lambda$ , is the quantity (in absence of fixed effects) :

$$\lambda = \int \exp(l) f_N(l, \mu, V_P) dl = \exp(\mu + V_P/2) \quad (S6.12)$$

after simplifications in Mathematica. Again, when fixed effects are included, it suffices to average over them :

$$\lambda = \frac{1}{N} \sum_i \exp(\mu + (\mathbf{X}\hat{\mathbf{b}})_i + V_P/2) = \frac{1}{N} \sum_i \lambda_i, \quad (S6.13)$$

where  $\lambda_i$  is the observed mean conditional to  $(\mathbf{X}\hat{\mathbf{b}})_i$ .

**Expected-scale variance** In absence of fixed effects, the expected-scale variance is defined as follow (using Keonig's formula) :

$$V_{P,\text{exp}} = \int \exp(l)^2 f_N(l, \mu, V_P) dl - \lambda^2, \quad (S6.14)$$

which, using Mathematica, simplifies into :

$$V_{P,\text{exp}} = \lambda^2 [\exp(V_P) - 1]. \quad (S6.15)$$

When fixed effects, are included in the model, the formula does not simplify as much, because of the averaging :

$$V_{P,\text{exp}} = \exp(V_P) \frac{1}{N} \sum_i \exp(2\mu + 2(\mathbf{X}\hat{\mathbf{b}})_i + V_P) - \lambda^2 \quad (S6.16)$$

Note that  $\lambda$  should be computed according to Eq. S6.15 in that case. Let us define the parameter  $\Lambda$  as

$$\Lambda = \frac{1}{N} \sum_i \exp(2\mu + 2(\mathbf{X}\hat{\mathbf{b}})_i + V_P) = \frac{1}{N} \sum_i \lambda_i^2, \quad (S6.17)$$

so that  $V_{P,\text{exp}} = \Lambda \exp(V_P) - \lambda^2$ . The parameter  $\Lambda$  is the average of the squared values of  $\lambda_i$ . Unfortunately,  $\Lambda \neq \lambda^2$ , hence there is no further simplification.

**Distribution variance** Because the variance of a Poisson distribution is equal to the mean, the distribution variance reduces to  $\lambda$  :

$$V_{\text{dist}} = \int \exp(l) f_N(l, \mu, V_P) dl = \lambda \quad (S6.18)$$

When fixed effects are included in the model,  $\lambda$  should be computed as in Eq. S6.15.

**Observed-scale additive variance** In order to compute the observed-scale additive variance, we need the parameter  $\Psi$  defined in Eq. A6.16 in the main text. Again, because the derivative of an exponential is an exponential, the calculation reduces to  $\lambda$  :

$$\Psi = \int \exp(l) f_N(l, \mu, V_P) dl = \lambda \quad (S6.19)$$

Once again, when fixed effects are included in the model,  $\lambda$  should be computed as in Eq. S6.15.

The observed-scale additive variance is then computed as  $\Psi^2 V_A = \lambda^2 V_A$ .



**Observed-scale heritability** In presence of fixed effects in the model, the heritability is consequently the “simple” following ratio :

$$h_{\text{obs}}^2 = \frac{\lambda^2 V_A}{\Lambda \exp(V_P) - \lambda^2 + \lambda} \quad (\text{S6.20})$$

**Negative-Binomial distribution** All the above results can be extended to the Negative-Binomial distribution, which, compared to the Poisson distribution, includes an overdispersion parameter (hereby noted  $\theta$ ). The new parameter  $\theta$  only affects the distribution variance such as :

$$h_{\text{obs}}^2 = \frac{\lambda^2 V_A}{\Lambda \exp(V_P) - \lambda^2 + \lambda + \exp(2(\mu + V_P))/\theta} \quad (\text{S6.21})$$

### S6.3 Calculation of the Price-Robertson identity

The Price-Robertson identity (ROBERTSON, 1966 ; PRICE, 1970) is based on the covariance between the (latent) additive genetic values and the expected fitness. This covariance can be computed using Koenig’s formula for covariance :

$$\Delta\mu = \text{cov}(a, W_{\text{exp}}) = E[a \times W_{\text{exp}}] - E[a] E[W_{\text{exp}}] = E[a \times W_{\text{exp}}], \quad (\text{S6.22})$$

since, by definition,  $E[a] = 0$ . Hence we obtain the following complex calculation :

$$\text{cov}(a, W_{\text{exp}}) = \iint a(W_{\text{exp}}(l))f(l|a)f(a)dlda = \iint aW_{\text{exp}}(l)f_{\mathcal{N}}(l, \mu + a, V_{\text{RE}} + V_{\text{O}})f_{\mathcal{N}}(a, 0, V_A)dlda. \quad (\text{S6.23})$$

Double-integration can be difficult to solve, but we can simplify further the computation by noticing that :

$$\begin{aligned} \text{cov}(a, W_{\text{exp}}) &= \iint aW_{\text{exp}}(l)f_{\mathcal{N}}(l, \mu + a, V_{\text{RE}} + V_{\text{O}})f_{\mathcal{N}}(a, 0, V_A)dlda, \\ &= \int af_{\mathcal{N}}(a, 0, V_A) \int W_{\text{exp}}(l)f_{\mathcal{N}}(l, \mu + a, V_{\text{RE}} + V_{\text{O}})dlda. \end{aligned} \quad (\text{S6.24})$$

Noting that

$$E[W_{\text{exp}}|a] = \int W_{\text{exp}}(l)f_{\mathcal{N}}(l, \mu + a, V_{\text{RE}} + V_{\text{O}})dl, \quad (\text{S6.25})$$

we can simplify EQ. S6.24 into

$$\Delta\mu = \text{cov}(a, W_{\text{exp}}) = \int aE[W_{\text{exp}}|a]f_{\mathcal{N}}(a, 0, V_A)da. \quad (\text{S6.26})$$

# RÉFÉRENCES BIBLIOGRAPHIQUES

---

- ÅGREN, J. & D. W. SCHEMSKE (2012). Reciprocal transplants demonstrate strong adaptive differentiation of the model organism *Arabidopsis thaliana* in its native range. *New Phytologist* 194, p. 1112–1122. (cf. p. 74).
- AITKEN, S. N., S. YEAMAN, J. A. HOLLIDAY, T. WANG & S. CURTIS-MCLANE (2008). Adaptation, migration or extirpation : climate change outcomes for tree populations. *Evolutionary Applications* 1, p. 95–111. (cf. p. 12, 111).
- AKEY, J. M., G. ZHANG, K. ZHANG, L. JIN & M. D. SHRIVER (2002). Interrogating a high-density SNP map for signatures of natural selection. *Genome Research* 12, p. 1805–1814. (cf. p. 53).
- ALBANI, M. C., L. CASTAINGS, S. WÖTZEL, J. L. MATEOS, J. WUNDER, R. WANG, M. REYMOND & G. COUPLAND (2012). PEP1 of *Arabidopsis thaliana* is encoded by two overlapping genes that contribute to natural genetic variation in perennial flowering. *PLoS Genet* 8, e1003130. (cf. p. 22).
- ALBERTO, F., L. BOUFFIER, J.-M. LOUVET, J.-B. LAMY, S. DELZON & A. KREMER (2011). Adaptive responses for seed and leaf phenology in natural populations of sessile oak along an altitudinal gradient. *Journal of Evolutionary Biology* 24, p. 1442–1454. (cf. p. 73, 112, 128).
- ALBERTO, F. et al. (2013). Potential for evolutionary responses to climate change – evidence from tree populations. *Global Change Biology* 19, p. 1645–1661. (cf. p. 111, 112).
- ALTSCHUL, S. F., T. L. MADDEN, A. A. SCHÄFFER, J. ZHANG, Z. ZHANG, W. MILLER & D. J. LIPMAN (1997). Gapped BLAST and PSI-BLAST : a new generation of protein database search programs. *Nucleic Acids Research* 25, p. 3389–3402. (cf. p. 117).
- ANSELL, S. W., M. GRUNDMANN, S. J. RUSSELL, H. SCHNEIDER & J. C. VOGEL (2008). Genetic discontinuity, breeding-system change and population history of *Arabidopsis thaliana* in the Italian peninsula and adjacent Alps. *Molecular Ecology* 17, p. 2245–2257. (cf. p. 22, 90, 113).
- ANSELL, S. W., H. K. STENØIEN, M. GRUNDMANN, S. J. RUSSELL, M. A. KOCH, H. SCHNEIDER & J. C. VOGEL (2011). The importance of Anatolian mountains as the cradle of global diversity in *Arabidopsis thaliana*, a key arctic–alpine species. *Annals of Botany* 108, p. 241–252. (cf. p. 90).
- ASSEFA, A., D. EHRICH, P. TABERLET, S. NEMOMISSA & C. BROCHMANN (2007). Pleistocene colonization of afro-alpine ‘sky islands’ by the arctic-alpine *Arabidopsis thaliana*. *Heredity* 99, p. 133–142. (cf. p. 90).
- AUSTIN, M. P. & K. P. VAN NIEL (2011). Improving species distribution models for climate change studies : variable selection and scale. *Journal of Biogeography* 38, p. 1–8 (cf. p. 102).
- AYERS, D. R., R. J. PEREIRA, A. A. BOLIGON, F. F. SILVA, F. S. SCHENKEL, V. M. ROSO & L. G. ALBUQUERQUE (2013). Linear and Poisson models for genetic evaluation of tick resistance in cross-bred Hereford x Nelor cattle. *Journal of Animal Breeding and Genetics* 130, p. 417–424 (cf. p. 166).
- BAILEY, C. D., M. A. KOCH, M. MAYER, K. MUMMENHOFF, S. L. O’KANE, S. I. WARWICK, M. D. WINDHAM & I. A. AL-SHEHBAB (2006). Toward a global phylogeny of the Brassicaceae. *Molecular Biology and Evolution* 23, p. 2142–2160. (cf. p. 22).
- BAIRD, N. A., P. D. ETTER, T. S. ATWOOD, M. C. CURREY, A. L. SHIVER, Z. A. LEWIS, E. U. SELKER, W. A. CRESKO & E. A. JOHNSON (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* 3, e3376. (cf. p. 18, 76).
- BAKA, Z. (2008). Occurrence and ultrastructure of *Albugo candida* on a new host, *Arabidopsis thaliana* in Saudi Arabia. *Micron* 39, p. 1138–1144. (cf. p. 115).
- BALDING, D. J. & R. A. NICHOLS (1995). A method for quantifying differentiation between populations at multi-allelic loci and its implications for investigating identity and paternity. *Genetica* 96, p. 3–12. (cf. p. 26, 50).
- BARNARD, J. & D. B. RUBIN (1999). Small-sample degrees of freedom with multiple imputation. *Biometrika* 86, p. 948–955. (cf. p. 95).
- BASSAR, R. D. et al. (2010). Local adaptation in Trinidadian guppies alters ecosystem processes. *Proceedings of the National Academy of Sciences* 107, p. 3616–3621. (cf. p. 73).
- BATES, D., M. MÄCHLER, B. BOLKER & S. WALKER (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67, p. 48 (cf. p. 95).
- BEAUMONT, M. A. & D. J. BALDING (2004). Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology* 13, p. 969–980. (cf. p. 27, 32, 36, 51, 53–55, 78).
- BEAUMONT, M. A. & R. A. NICHOLS (1996). Evaluating loci for use in the genetic analysis of population structure. *Proceedings : Biological Sciences* 263, p. 1619–1626 (cf. p. 27, 32).
- BEAVIS, W. D. (1994). « The power and deceit of QTL experiments : lessons from comparative QTL studies ». *Proceedings of the forty-ninth annual corn and sorghum industry research conference*. Chicago, IL, p. 250–266 (cf. p. 142).
- BELLARD, C., C. BERTLESMEIER, P. LEADLEY, W. THUILLER & F. COURCHAMP (2012). Impacts of climate change on the future of biodiversity : Biodiversity and climate change. *Ecology Letters* 15, p. 365–377. (cf. p. 12).
- BELLO, F. de, S. LAVOREL, S. LAVERGNE, C. H. ALBERT, I. BOULANGEAT, F. MAZEL & W. THUILLER (2013). Hierarchical effects of environmental filters on the functional structure of plant communities : a case study in the French Alps. *Ecography* 36, p. 393–402. (cf. p. 85, 136).

- BENJAMINI, Y. & Y. HOCHBERG (1995). Controlling the false discovery rate : a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)* 57, p. 289–300 (cf. p. 33).
- BÉRÉDOS, C., P. A. ELLIS, J. G. PILKINGTON & J. M. PEMBERTON (2014). Estimating quantitative genetic parameters in wild populations : a comparison of pedigree and genomic approaches. *Molecular Ecology* 23, p. 3434–3451. (cf. p. 171).
- BERG, J. J. & G. COOP (2014). A Population Genetic Signal of Polygenic Adaptation. *PLoS Genet* 10, e1004412. (cf. p. 20).
- BERGELSON, J. & F. ROUX (2010). Towards identifying genes underlying ecologically relevant traits in *Arabidopsis thaliana*. *Nature Reviews Genetics* 11, p. 867–879. (cf. p. 91).
- BERGONZI, S., M. C. ALBANI, E. V. L. van THEMAAT, K. J. V. NORDSTROEM, R. WANG, K. SCHNEEBERGER, P. D. MOERLAND & G. COUPLAND (2013). Mechanisms of age-dependent response to winter temperature in perennial flowering of *Arabis alpina*. *Science* 340, p. 1094–1097. (cf. p. 90, 100, 101).
- BIGHAM, A. et al. (2010). Identifying signatures of natural selection in Tibetan and Andean populations using dense genome scan data. *PLoS Genet* 6, e1001116. (cf. p. 54).
- BLANQUART, F., O. KALTZ, S. L. NUISMER & S. GANDON (2013). A practical guide to measuring local adaptation. *Ecology Letters* 16, p. 1195–1205. (cf. p. 53).
- BONHOMME, M., C. CHEVALET, B. SERVIN, S. BOITARD, J. ABDALLAH, S. BLOTT & M. SANCRISTOBAL (2010). Detecting selection in population trees : the Lewontin and Krakauer test extended. *Genetics* 186, p. 241–262. (cf. p. 27, 53, 78).
- BOUCHER, F. C., W. THULLER, C. ROQUET, R. DOUZET, S. AUBERT, N. ALVAREZ & S. LAVERGNE (2012). Reconstructing the origins of high-alpine niches and cushion life form in the genus *Androsace* s.l. (Primulaceae). *Evolution* 66, p. 1255–1268. (cf. p. 113, 136).
- BOURRET, V., M. DIONNE, M. P. KENT, S. LIEN & L. BERNATCHEZ (2013). Landscape genomics in Atlantic salmon (*Salmo salar*) : searching for gene–environment interactions driving local adaptation. *Evolution* 67, p. 3469–3487. (cf. p. 75).
- BOZINOVIC, F., J. M. ROJAS, B. R. BROITMAN & R. A. VÁSQUEZ (2009). Basal metabolism is correlated with habitat productivity among populations of degus (*Octodon degus*). *Comparative Biochemistry and Physiology Part A : Molecular & Integrative Physiology* 152, p. 560–564. (cf. p. 73).
- BRACHI, B., R. VILLOUTREIX, N. FAURE, N. HAUTEKEETE, Y. PIQUOT, M. PAUWELS, D. ROBY, J. CUGUEN, J. BERGELSON & F. ROUX (2013). Investigation of the geographical scale of adaptive phenological variation and its underlying genetics in *Arabidopsis thaliana*. *Molecular Ecology* 22, p. 4222–4240. (cf. p. 11, 73, 78, 112, 139).
- BROOKS, S. P., P. GIUDICI & G. O. ROBERTS (2003). Efficient construction of reversible jump Markov chain Monte Carlo proposal distributions. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)* 65, p. 3–39. (cf. p. 157).
- BROOKS, S. (1998). Markov chain Monte Carlo method and its application. *Journal of the Royal Statistical Society : Series D (The Statistician)* 47, p. 69–100. (cf. p. 57, 157).
- BROWNE, W. J., S. V. SUBRAMANIAN, K. JONES & H. GOLDSTEIN (2005). Variance partitioning in multilevel logistic models that exhibit overdispersion. *Journal of the Royal Statistical Society* 168, p. 599–613 (cf. p. 178).
- BUEHLER, D., R. GRAF, R. HOLDEREGGER & F. GUGERLI (2012). Contemporary gene flow and mating system of *Arabis alpina* in a Central European alpine landscape. *Annals of Botany* 109, p. 1359–1367. (cf. p. 22, 86, 90, 106).
- BUEHLER, D., R. HOLDEREGGER, S. BRODBECK, E. SCHNYDER & F. GUGERLI (2014). Validation of outlier loci through replication in independent data sets : a test on *Arabis alpina*. *Ecology and Evolution* 4, p. 4296–4306. (cf. p. 75, 90, 126, 139).
- BUEHLER, D., B. N. PONCET, R. HOLDEREGGER, S. MANEL, P. TABERLET & F. GUGERLI (2013). An outlier locus relevant in habitat-mediated selection in an alpine plant across independent regional replicates. *Evolutionary Ecology* 27, p. 285–300. (cf. p. 22, 86, 87, 106).
- BULMER, M. G. (1971). The effect of selection on genetic variability. *The American Naturalist* 105, p. 201–211 (cf. p. 177).
- BUUREN, S. van & K. GROOTHUIS-OUUDSHOORN (2011). mice : multivariate imputation by chained equations in R. *Journal of Statistical Software* 45, (cf. p. 95).
- BYARS, S. G., W. PAPST & A. A. HOFFMANN (2007). Local adaptation and cogradient selection in the alpine plant, *Poa hiemata*, along a narrow altitudinal gradient. *Evolution* 61, p. 2925–2941. (cf. p. 128).
- CASWELL, H. (2001). *Matrix population models : construction, analysis, and interpretation*. Sunderland, Massachusetts (US) : Sinauer Associates. (cf. p. 101).
- CATCHEN, J. M., A. AMORES, P. HOHENLOHE, W. CRESKO & J. H. POSTLETHWAIT (2011). Stacks : Building and Genotyping Loci De Novo From Short-Read Sequences. *G3 : Genes, Genomes, Genetics* 1, p. 171–182. (cf. p. 19, 115).
- CATCHEN, J. M., P. A. HOHENLOHE, S. BASSHAM, A. AMORES & W. A. CRESKO (2013). Stacks : an analysis tool set for population genomics. *Molecular Ecology* 22, p. 3124–3140. (cf. p. 19, 115).
- CHARLESWORTH, B. (1998). Measures of divergence between populations and the effect of forces that reduce variability. *Molecular Biology and Evolution* 15, p. 538–543 (cf. p. 54).
- (2013). Background Selection 20 Years on : The Wilhelmine E. Key 2012 Invitational Lecture. *Journal of Heredity* 104, p. 161–171. (cf. p. 49, 50, 56, 63).
- CHOLER, P., R. MICHALET & R. M. CALLAWAY (2001). Facilitation and competition on gradients in alpine plant communities. *Ecology* 82, p. 3295–3308. (cf. p. 85, 113).
- CHOPRA, D., H. WOLFF, J. SPAN, S. SCHELLMANN, G. COUPLAND, M. C. ALBANI, A. SCHRADER & M. HÜLSKAMP (2014). Analysis of TTG1 function in *Arabis alpina*. *BMC Plant Biology* 14, p. 16. (cf. p. 90).
- COLAUTTI, R. I., C. G. ECKERT & S. C. H. BARRETT (2010). Evolutionary constraints on adaptive evolution during range expansion in an invasive plant. *Proceedings of the Royal Society B-Biological Sciences* 277, p. 1799–1806. (cf. p. 112).
- COOP, G., D. WITONSKY, A. DI RIENZO & J. K. PRITCHARD (2010). Using environmental correlations to identify loci underlying local adaptation. *Genetics* 185, p. 1411–1423. (cf. p. 27, 32, 33, 36, 37, 54, 65, 78, 79, 81, 160).
- CORNUET, J.-M., F. SANTOS, M. A. BEAUMONT, C. P. ROBERT, J.-M. MARIN, D. J. BALDING, T. GUILLEMAUD & A. ESTOUP (2008). Inferring population history with DIY ABC : a user-friendly approach to approximate Bayesian computation. *Bioinformatics* 24, p. 2713–2719. (cf. p. 19).
- CUSHMAN, S. A. (2014). Grand challenges in evolutionary and population genetics : the importance of integrating epigenetics, genomics, modeling, and experimentation. *Frontiers in Genetics* 5, (cf. p. 75).

- DAUB, J. T., T. HOFER, E. CUTIVET, I. DUPANLOUP, L. QUINTANA-MURCI, M. ROBINSON-RECHAVI & L. EXCOFFIER (2013). Evidence for polygenic adaptation to pathogens in the Human genome. *Molecular Biology and Evolution* **30**, p. 1544–1558. (cf. p. 54).
- DAVIS, J., J. SPAETH & C. HUSON (1961). A technique for analyzing the effects of group composition. *American Sociological Review* **26**, p. 215–225. (cf. p. 78).
- DAWKINS, R. (1982). *The extended phenotype : the long reach of the gene*. Revised. Oxford (UK) : Oxford University Press (cf. p. 8).
- DE KORT, H., K. VANDEPITTE, H. H. BRUUN, D. CLOSSET-KOPP, O. HONNAY & J. MERGEAY (2014). Landscape genomics and a common garden trial reveal adaptive differentiation to temperature across Europe in the tree species *Alnus glutinosa*. *Molecular Ecology* **23**, p. 4709–4721. (cf. p. 71, 76, 78, 79).
- DE MITA, S., A.-C. THUILLET, L. GAY, N. AHMADI, S. MANEL, J. RONFORT & Y. VIGOUROUX (2013). Detecting selection along environmental gradients : analysis of eight methods and their effectiveness for outbreeding and selfing populations. *Molecular Ecology* **22**, p. 1383–1399. (cf. p. 30, 32, 33, 44, 54).
- DEFAVERI, J. & J. MERILÄ (2014). Local adaptation to salinity in the three-spined stickleback? *Journal of Evolutionary Biology* **27**, p. 290–302. (cf. p. 73).
- DEMPSTER, E. R. & I. M. LERNER (1950). Heritability of Threshold Characters. *Genetics* **35**, p. 212–236 (cf. p. 166, 170, 187).
- DEVLIN, B. & K. ROEDER (1999). Genomic control for association studies. *Biometrics* **55**, p. 997–1004. (cf. p. 117).
- DEWITT, T. J., A. SIH & D. S. WILSON (1998). Costs and limits of phenotypic plasticity. *Trends in Ecology & Evolution* **13**, p. 77–81. (cf. p. 12, 85, 142).
- DOAK, D. F. & W. F. MORRIS (2010). Demographic compensation and tipping points in climate-induced range shifts. *Nature* **467**, p. 959–962. (cf. p. 102).
- DRAY, S. & A.-B. DUFOUR (2007). The ade4 package : implementing the duality diagram for ecologists. *Journal of statistical software* **22**, p. 1–20 (cf. p. 95, 114).
- DUFORET-FREBOURG, N., É. BAZIN & M. G. B. BLUM (2014). Genome scans for detecting footprints of local adaptation using a Bayesian factor model. *Molecular Biology and Evolution* **31**, p. 2483–2495. (cf. p. 27, 53, 69, 78, 79).
- ECKERT, A. J., A. D. BOWER, S. C. GONZÁLEZ-MARTÍNEZ, J. L. WĘGRZYN, G. COOP & D. B. NEALE (2010). Back to nature : ecological genomics of loblolly pine (*Pinus taeda*, Pinaceae). *Molecular Ecology* **19**, p. 3789–3805. (cf. p. 75).
- EDELAAR, P. & M. BJÖRKLUND (2011). If FST does not measure neutral genetic differentiation, then comparing it with QST is misleading. Or is it? *Molecular Ecology* (cf. p. 71, 76, 112).
- EDELAAR, P., P. BURRACO & I. GOMEZ-MESTRE (2011). Comparisons between QST and FST—how wrong have we been? *Molecular Ecology* **20**, p. 4830–4839. (cf. p. 49, 54, 71, 74, 112, 116).
- EDMONDS, C. A., A. S. LILLIE & L. L. CAVALLI-SFORZA (2004). Mutations arising in the wave front of an expanding population. *Proceedings of the National Academy of Sciences of the United States of America* **101**, p. 975–979. (cf. p. 33, 49, 54, 56, 63).
- EL RABEY, H., K. F. SALEM & M. Z. MATTAR (2013). The genetic diversity and relatedness of rice (*Oryza sativa* L.) cultivars as revealed by AFLP and SSRs markers. *Life Science Journal* **10**, p. 1471–1479 (cf. p. 76).
- ELSHIRE, R. J., J. C. GLAUBITZ, Q. SUN, J. A. POLAND, K. KAWAMOTO, E. S. BUCKLER & S. E. MITCHELL (2011). A robust, simple Genotyping-By-Sequencing (GBS) approach for high diversity species. *PLoS ONE* **6**, e19379. (cf. p. 17, 76).
- EXCOFFIER, L., T. HOFER & M. FOLL (2009). Detecting loci under selection in a hierarchically structured population. *Heredity* **103**, p. 285–298. (cf. p. 33, 54, 63, 115).
- FALCONER, D. S. (1960). *Introduction to Quantitative Genetics*. Edinburgh (UK) : Oliver & Boyd (cf. p. 164).
- FALCONER, D. S. & T. F. MACKAY (1996). *Introduction to Quantitative Genetics*. 4<sup>e</sup> éd. Harlow, Essex (UK) : Benjamin Cummings (cf. p. 16).
- FARIA, R., S. RENAULT, J. GALINDO, C. PINHO, J. MELO-FERREIRA, M. MELO, F. JONES, W. SALZBURGER, D. SCHLUTER & R. BUTLIN (2014). Advances in ecological speciation : an integrative approach. *Molecular Ecology* **23**, p. 513–521. (cf. p. 53).
- FELSENSTEIN, J. (1985). Phylogenies and the comparative method. *The American Naturalist* **125**, p. 1–15 (cf. p. 127).
- FIÉVET, J. B., C. DILLMANN & D. d. VIENNE (2009). Systemic properties of metabolic networks lead to an epistasis-based model for heterosis. *Theoretical and Applied Genetics* **120**, p. 463–473. (cf. p. 141).
- FISCHER, M. C., C. RELLSTAB, A. TEDDER, S. ZOLLER, F. GUGERLI, K. K. SHIMIZU, R. HOLDEREGGER & A. WIDMER (2013). Population genomic footprints of selection and associations with climate in natural populations of *Arabidopsis halleri* from the Alps. *Molecular Ecology* **22**, p. 5594–5607. (cf. p. 75).
- FISHER, R. A. (1924). The biometrical study of heredity. *Eugenics Review* **16**, p. 189–210 (cf. p. 167).
- (1930). *The Genetical Theory of Natural Selection*. Oxford (UK) : Oxford University Press. (Cf. p. 164).
- FISHER, R. A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh* **52**, p. 399–433 (cf. p. 12, 82, 141, 142, 164, 169).
- FOLL, M. & O. E. GAGGIOTTI (2006). Identifying the environmental factors that determine the genetic structure of populations. *Genetics* **174**, p. 875–891. (cf. p. 51).
- (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers : a Bayesian perspective. *Genetics* **180**, p. 977–993. (cf. p. 27, 32, 33, 36, 44, 51, 53–55, 57, 58, 69, 78, 79, 160).
- FOLL, M., O. E. GAGGIOTTI, J. T. DAUB, A. VATSIOU & L. EXCOFFIER (2014). Widespread signals of convergent adaptation to high altitude in Asia and America. *The American Journal of Human Genetics* **95**, p. 394–407. (cf. p. 54, 65, 70).
- FOULLEY, J. L. & S. IM (1993). A marginal quasi-likelihood approach to the analysis of Poisson variables with generalized linear mixed models. *Genetics Selection Evolution* **25**, p. 101. (cf. p. 166, 171).
- FRICHOT, É., F. MATHIEU, T. TROUILLON, G. BOUCHARD & O. FRANÇOIS (2014). Fast and efficient estimation of individual ancestry coefficients. *Genetics* **196**, p. 973–983. (cf. p. 115, 118, 184).
- FRICHOT, É., S. D. SCHOVILLE, G. BOUCHARD & O. FRANÇOIS (2013). Testing for associations between loci and environmental gradients using latent factor mixed models. *Molecular Biology and Evolution* **30**, p. 1687–1699. (cf. p. 28, 32, 33, 36, 45, 54, 65, 78, 79, 117, 160).

- FRICHOT, É., S. D. SCHOVILLE, P. de VILLEMEREUIL, O. E. GAGGIOTTI & O. FRANÇOIS (2015). Detecting adaptive evolution based on association with ecological gradients : Orientation matters ! *Heredity* **115**, p. 22–28. (cf. p. 47, 66).
- FUMAGALLI, M., M. SIRONI, U. POZZOLI, A. FERRER-ADMETTLA, L. PATTINI & R. NIELSEN (2011). Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through Human evolution. *PLoS Genet* **7**, e1002355. (cf. p. 54).
- GAGGIOTTI, O. E. & M. FOLL (2010). Quantifying population structure using the F-model. *Molecular Ecology Resources* **10**, p. 821–830. (cf. p. 27, 50, 51, 55, 77).
- GARAMSZEGI, L. Z., éd. (2014). *Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology*. Berlin, Heidelberg (Germany) : Springer (cf. p. 81, 127, 132, 138).
- GAUTIER, M. (2015). Genome-wide scan for adaptive divergence and association with population-specific covariates. *bioRxiv*, p. 023721. (cf. p. 27, 70).
- GAUZERE, J., S. ODDOU-MURATORIO, C. PICHOT, F. LEFÈVRE & É. KLEIN (2013). Biases in quantitative genetic analyses using open-pollinated progeny tests from natural tree populations. *Acta Botanica Gallica* **160**, p. 227–238. (cf. p. 77).
- GELMAN, A., J. B. CARLIN, H. S. STERN & D. B. RUBIN (2004). *Bayesian data analysis*. Second. Text in Statistical Science. Boca Raton, Florida (US) : Chapman & Hall/CRC Press. (Cf. p. 157).
- GHALAMBOR, C. K., J. K. MCKAY, S. P. CARROLL & D. N. REZNICK (2007). Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. *Functional Ecology* **21**, p. 394–407. (cf. p. 86, 107).
- GIMENEZ-BENAVIDES, L., A. ESCUDERO & J. M. IRIONDO (2007). Local adaptation enhances seedling recruitment along an altitudinal gradient in a high mountain Mediterranean plant. *Annals of Botany* **99**, p. 723–734. (cf. p. 128).
- GONDA, A., G. HERCZEG & J. MERILÄ (2011). Population variation in brain size of nine-spined sticklebacks (*Pungitius pungitius*) - local adaptation or environmentally induced variation ? *BMC Evolutionary Biology* **11**, p. 75. (cf. p. 73).
- GONZALO-TURPIN, H. & L. HAZARD (2009). Local adaptation occurs along altitudinal gradient despite the existence of gene flow in the alpine plant species *Festuca eskia*. *Journal of Ecology* **97**, p. 742–751. (cf. p. 128, 136).
- GOULD, S. J. & R. C. LEWONTIN (1979). The spandrels of San Marco and the Panglossian paradigm : a critique of the adaptationist programme. *Proceedings of the Royal Society of London B : Biological Sciences* **205**, p. 581–598. (cf. p. 7, 8, 11).
- GREEN, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82**, p. 711–732. (cf. p. 57).
- GUILLOT, G., R. VITALIS, A. I. ROUZIC & M. GAUTIER (2014). Detecting correlation between allele frequencies and environmental variables as a signature of selection. A fast computational approach for genome-wide studies. *Spatial Statistics*. *Spatial Statistics* Miami **8**, p. 145–155. (cf. p. 28, 54, 65, 66, 78).
- GÜNTHER, T. & G. COOP (2013). Robust identification of local adaptation from allele frequencies. *Genetics* **195**, p. 205–220. (cf. p. 53).
- HADFIELD, J. D. (2010). MCMC methods for multi-response generalised linear mixed models : The MCMCglmm R package. *Journal of Statistical Software* **33**, p. 1–22 (cf. p. 116, 172).
- HANCOCK, A. M., D. B. WITONSKY, G. ALKORTA-ARANBURU, C. M. BEALL, A. GEBREMEDHIN, R. SUKERNIK, G. UTERMANN, J. K. PRITCHARD, G. COOP & A. DI RIENZO (2011). Adaptations to climate-mediated selective pressures in Humans. *PLoS Genet* **7**, e1001375. (cf. p. 91).
- HAUTIER, Y., C. F. RANDIN, J. STOECKLIN & A. GUISAN (2009). Changes in reproductive investment with altitude in an alpine plant. *Journal of Plant Ecology* **2**, p. 125–134. (cf. p. 128, 136).
- HEIDELBERGER, P. & P. D. WELCH (1981). A spectral method for confidence interval generation and run length control in simulations. *Commun. ACM* **24**, p. 233–245. (cf. p. 116).
- HENDERSON, C. R. (1950). Estimation of genetic parameters. *Annals of Mathematical Statistics* **21**, p. 309–310 (cf. p. 164).
- HENDERSON, C. R. (1973). « Proceedings of the animal breeding and genetics symposium in honour of Dr. Jay L. Lush ». Published by the American Society of Animal Science & the American Dairy Science Association (cf. p. 164).
- HERMISSON, J. (2009). Who believes in whole-genome scans for selection ? *Heredity* **103**, p. 283–284. (cf. p. 33).
- HERNÁNDEZ-SERRANO, A., M. VERDÚ, L. SANTOS-DEL-BLANCO, J. CLIMENT, S. C. GONZÁLEZ-MARTÍNEZ & J. G. PAUSAS (2014). Heritability and quantitative genetic divergence of serotiny, a fire-persistence plant trait. *Annals of Botany* **114**, p. 571–577. (cf. p. 77).
- HIJMANS, R. J., S. E. CAMERON, J. L. PARRA, P. G. JONES & A. JARVIS (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25**, p. 1965–1978. (cf. p. 80).
- HILL, W. G. & M. KIRKPATRICK (2010). What animal breeding has taught us about evolution. *Annual Review of Ecology, Evolution, and Systematics* **41**, p. 1–19. (cf. p. 13).
- HODGE, J. & G. RADICK, éd. (2009). *The Cambridge companion to Darwin*. 2nd edition. Cambridge (UK) : Cambridge University Press (cf. p. 7).
- HOLDEREGGER, R. et al. (2008). Land ahead : using genome scans to identify molecular markers of adaptive relevance. *Plant Ecology & Diversity* **1**, p. 273–283. (cf. p. 75, 78).
- HONG, Z. et al. (2003). A unified nomenclature for Arabidopsis dynamin-related large GTPases based on homology and possible functions. *Plant Molecular Biology* **53**, p. 261–265. (cf. p. 126).
- HORVITZ, C. C. & D. W. SCHEMSKE (2002). Effects of plant size, leaf herbivory, local competition and fruit production on survival, growth and future reproduction of a neotropical herb. *Journal of Ecology* **90**, p. 279–290. (cf. p. 101).
- HOULE, D. (1992). Comparing evolvability and variability of quantitative traits. *Genetics* **130**, p. 195–204 (cf. p. 170).
- HSIEH, C.-C. S. (2009). « Characterization of FRIABLE1-like homologues in Arabidopsis using bioinformatics and reverse genetics ». Thesis (cf. p. 126, 137).
- HUXLEY, A. (1986). *Mountain flowers of Europe*. 2nd Edition edition. London (UK) : Cassell Illustrated. (Cf. p. 90).
- JONES, A. G., C. M. SMALL, K. A. PACZOLT & N. L. RATTERMAN (2010). A practical guide to methods of parentage analysis. *Molecular Ecology Resources* **10**, p. 6–30. (cf. p. 77).

- JONES, O. R. & J. WANG (2010). COLONY : a program for parentage and sibship inference from multilocus genotype data. *Molecular Ecology Resources* **10**, p. 551–555. (cf. p. 77, 116).
- JOOST, S., A. BONIN, M. W. BRUFORD, L. DESPRÉS, C. CONORD, G. ERHARDT & P. TABERLET (2007). A spatial analysis method (SAM) to detect candidate loci for selection : towards a landscape genomics approach to adaptation. *Molecular Ecology* **16**, p. 3955–3969. (cf. p. 32).
- KÄLL, L., J. D. STOREY, M. J. MACCOSS & W. S. NOBLE (2008). Posterior error probabilities and false discovery rates : two sides of the same coin. *Journal of Proteome Research* **7**, p. 40–44. (cf. p. 57, 157).
- KANG, H. M., N. A. ZAITLEN, C. M. WADE, A. KIRBY, D. HECKERMAN, M. J. DALY & E. ESKIN (2008). Efficient control of population structure in model organism association mapping. *Genetics* **178**, p. 1709–1723. (cf. p. 117).
- KARHUNEN, M., J. MERILÄ, T. LEINONEN, J. M. CANO & O. OVASKAINEN (2013). DRIFTSEL : an R package for detecting signals of natural selection in quantitative traits. *Molecular Ecology Resources* **13**, p. 746–754. (cf. p. 77, 112, 132).
- KARHUNEN, M., O. OVASKAINEN, G. HERCZEG & J. MERILÄ (2014). Bringing habitat information into statistical tests of local adaptation in quantitative traits : a case study of nine-spined sticklebacks. *Evolution* **68**, p. 559–568. (cf. p. 77, 112, 132).
- KARHUNEN, M. & O. OVASKAINEN (2012). Estimating population-level coancestry coefficients by an admixture F model. *Genetics* **192**, p. 609–617. (cf. p. 77, 112, 115, 118).
- KARL, R., C. KIEFER, S. W. ANSELL & M. A. KOCH (2012). Systematics and evolution of arctic-alpine *Arabis alpina* (Brassicaceae) and its closest relatives in the eastern Mediterranean. *American Journal of Botany* **99**, p. 778–794. (cf. p. 90).
- KAWAKAMI, T., T. J. MORGAN, J. B. NIPPERT, T. W. OCHELTREE, R. KEITH, P. DHAKAL & M. C. UNGERER (2011). Natural selection drives clinal life history patterns in the perennial sunflower species, *Helianthus maximiliani*. *Molecular Ecology* **20**, p. 2318–2328. (cf. p. 73, 112).
- KAWECKI, T. J. & D. EBERT (2004). Conceptual issues in local adaptation. *Ecology Letters* **7**, p. 1225–1241. (cf. p. 10, 14, 15, 111, 124).
- KIMURA, M. (1985). *The neutral theory of molecular evolution*. Cambridge (UK) : Cambridge University Press. (cf. p. 7, 18).
- KOCH, M. A., C. KIEFER, D. EHRICH, J. VOGEL, C. BROCHMANN & K. MUMMENHOFF (2006). Three times out of Asia Minor : the phylogeography of *Arabis alpina* L. (Brassicaceae). *Molecular Ecology* **15**, p. 825–839. (cf. p. 22, 90, 91).
- KOFLER, R. & C. SCHLÖTTERER (2012). Gowinda : unbiased analysis of gene set enrichment for genome-wide association studies. *Bioinformatics* **28**, p. 2084–2085. (cf. p. 58).
- KOLAKSAZOV, M., F. LAPORTE, K. ANANIEVA, P. DOBREV, M. HERZOG & E. ANANIEV (2013). Effect of chilling and freezing stresses on jasmonate content in *Arabis alpina*. *Bulgarian Journal of Agricultural Science* **19**, p. 15–17 (cf. p. 23, 87).
- KOLAKSAZOV, M., F. LAPORTE, V. GOLTSEV, M. HERZOG & E. D. ANANIEV (2014). Effect of frost stress on chlorophyll a fluorescence and modulated 820 nm reflection in *Arabis alpina* population from Rila mountain. *Genetics and Plant Physiology*, p. 1–2 (cf. p. 23, 87).
- KOLLMANN, J. & M. J. BANUELOS (2004). Latitudinal trends in growth and phenology of the invasive alien plant *Impatiens glandulifera* (Balsaminaceae). *Diversity and Distributions* **10**, p. 377–385. (cf. p. 112).
- KÖRNER, C. (2003). *Alpine plant life : functional plant ecology of high mountain ecosystems*. Berlin, Heidelberg (Germany) : Springer (cf. p. 11, 21, 85, 101, 102, 105, 106, 112, 126, 127, 136).
- (2007). The use of ‘altitude’ in ecological research. *Trends in Ecology & Evolution* **22**, p. 569–574. (cf. p. 85, 101, 113, 124).
- KORTE, A. & A. FARLOW (2013). The advantages and limitations of trait analysis with GWAS : a review. *Plant methods* **9**, p. 29 (cf. p. 78).
- KRUUK, L. E. B., S. J. E. BAIRD, K. S. GALE & N. H. BARTON (1999). A comparison of multilocus clines maintained by environmental adaptation or by selection against hybrids. *Genetics* **153**, p. 1959–1971 (cf. p. 49, 54).
- KRUUK, L. E. B. (2004). Estimating genetic parameters in natural populations using the ‘animal model’. *Philosophical Transactions of the Royal Society of London. Series B : Biological Sciences* **359**, p. 873–890. (cf. p. 77, 164, 166).
- KULWAL, P., J. ROY, H. BALYAN & P. GUPTA (2003). QTL mapping for growth and leaf characters in bread wheat. *Plant Science* **164**, p. 267–277. (cf. p. 45).
- LAMESCH, P. et al. (2012). The Arabidopsis Information Resource (TAIR) : improved gene annotation and new tools. *Nucleic Acids Research* **40**, (D1), p. D1202–D1210. (cf. p. 117).
- LANDE, R. (1979). Quantitative genetic analysis of multivariate evolution, applied to brain :body size allometry. *Evolution* **33**, p. 402–416 (cf. p. 167, 174).
- LANDE, R. & S. J. ARNOLD (1983). The measurement of selection on correlated characters. *Evolution* **37**, p. 1210–1226 (cf. p. 167, 174).
- LARSSON, L. C., L. LAIKRE, S. PALM, C. ANDRÉ, G. R. CARVALHO & N. RYMAN (2007). Concordance of allozyme and microsatellite differentiation in a marine fish, but evidence of selection at a microsatellite locus. *Molecular Ecology* **16**, p. 1135–1147. (cf. p. 54).
- LATOUR, B. (1996). *Petites leçons de sociologie des sciences*. T. 114. Paris (France) : Seuil (cf. p. 139).
- LEGER, E. A. & K. J. RICE (2007). Assessing the speed and predictability of local adaptation in invasive California poppies (*Eschscholzia californica*). *Journal of Evolutionary Biology* **20**, p. 1090–1103. (cf. p. 112).
- LEINONEN, T., R. J. S. MCCAIRNS, R. B. O’HARA & J. MERILÄ (2013). QST–FST comparisons : evolutionary and ecological insights from genomic heterogeneity. *Nature Reviews Genetics* **14**, p. 179–190. (cf. p. 112).
- LEPAIS, O. & C. F. E. BACLES (2014). Two are better than one : combining landscape genomics and common gardens for detecting local adaptation in forest trees. *Molecular Ecology* **23**, p. 4671–4673. (cf. p. 78, 79).
- LEVIN, D. A. (1973). The role of trichomes in plant defense. *The Quarterly Review of Biology* **48**, p. 3–15 (cf. p. 137).
- LEWONTIN, R. C. & J. KRAKAUER (1973). Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics* **74**, p. 175–195 (cf. p. 26, 27, 32).
- LINHART, Y. B. & M. C. GRANT (1996). Evolutionary significance of local genetic differentiation in plants. *Annual Review of Ecology and Systematics* **27**, p. 237–277. (cf. p. 73).

- LOBRÉAUX, S., S. MANEL & C. MELODELIMA (2014). Development of an *Arabidopsis thaliana* genomic contig sequence data set and application to single nucleotide polymorphisms discovery. *Molecular Ecology Resources* **14**, p. 411–418. (cf. p. 22, 23, 87, 90, 129).
- LOTTERHOS, K. E. & M. C. WHITLOCK (2014). Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. *Molecular Ecology* **23**, p. 2178–2192. (cf. p. 78).
- (2015). The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Molecular Ecology* **24**, p. 1031–1046. (cf. p. 139).
- LUIKART, G., P. R. ENGLAND, D. TALLMON, S. JORDAN & P. TABERLET (2003). The power and promise of population genomics : from genotyping to genome typing. *Nature Reviews Genetics* **4**, p. 981–994. (cf. p. 26, 32, 53, 75, 76).
- LUO, Y., A. WIDMER & S. KARRENBERG (2015). The roles of genetic drift and natural selection in quantitative trait divergence along an altitudinal gradient in *Arabidopsis thaliana*. *Heredity* **114**, p. 220–228. (cf. p. 22).
- LUQUET, E., J.-P. LENA, C. MIAUD & S. PLENET (2015). Phenotypic divergence of the common toad (*Bufo bufo*) along an altitudinal gradient : evidence for local adaptation. *Heredity* **114**, p. 69–79. (cf. p. 128).
- LURIN, C. et al. (2004). Genome-wide analysis of *Arabidopsis* pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *The Plant Cell* **16**, p. 2089–2103. (cf. p. 123, 137).
- LUSH, J. L. (1937). *Animal breeding plans*. Ames, Iowa (US) : Iowa State College Press (cf. p. 167).
- LUTTIKHUIZEN, P. C., J. DRENT, W. VAN DELDEN & T. PIERSMA (2003). Spatially structured genetic variation in a broadcast spawning bivalve : quantitative vs. molecular traits. *Journal of Evolutionary Biology* **16**, p. 260–272. (cf. p. 73).
- LYNCH, M. & B. WALSH (1998). *Genetics and analysis of quantitative traits*. Sunderland, Massachusetts (US) : Sinauer Associates (cf. p. 16, 164).
- MACKAY, T. F. C. (2001). The genetic architecture of quantitative traits. *Annual Review of Genetics* **35**, p. 303–339. (cf. p. 142).
- MANEL, S., B. N. PONCET, P. LEGENDRE, F. GUGERLI & R. HOLDEREGGER (2010). Common factors drive adaptive genetic variation at different spatial scales in *Arabidopsis thaliana*. *Molecular Ecology* **19**, p. 3824–3835. (cf. p. 22, 23, 86, 87, 90).
- MANEL, S., F. GUGERLI, W. THUILLER, N. ALVAREZ, P. LEGENDRE, R. HOLDEREGGER, L. GIELLY, P. TABERLET & INTRABIO DIV CONSORTIUM (2012). Broad-scale adaptive genetic variation in alpine plants is driven by temperature and precipitation. *Molecular Ecology* **21**, p. 3729–3738. (cf. p. 90, 102).
- McKAY, J. K., J. G. BISHOP, J.-Z. LIN, J. H. RICHARDS, A. SALA & T. MITCHELL-OLDS (2001). Local adaptation across a climatic gradient despite small effective population size in the rare sapphire rockcress. *Proceedings of the Royal Society of London B : Biological Sciences* **268**, p. 1715–1721. (cf. p. 73).
- MELODELIMA, C. & S. LOBRÉAUX (2013). Complete *Arabidopsis thaliana* chloroplast genome sequence and insight into its polymorphism. *Meta Gene* **1**, p. 65–75. (cf. p. 90).
- MILLER, M. R., J. P. DUNHAM, A. AMORES, W. A. CRESKO & E. A. JOHNSON (2007). Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Research* **17**, p. 240–248. (cf. p. 18, 76).
- MILOT, É., F. M. MAYER, D. H. NUSSEY, M. BOISVERT, F. PELLETIER & D. REALE (2011). Evidence for evolution in response to natural selection in a contemporary human population. *Proceedings of the National Academy of Sciences* **108**, p. 17040–17045. (cf. p. 166).
- MORGENSTERN, M. (2011). *Geographic variation in forest trees : genetic basis and application of knowledge in silviculture*. UBC Press. (Cf. p. 112).
- MORRIS, M. R. J., R. RICHARD, E. H. LEDER, R. D. H. BARRETT, N. AUBIN-HORTH & S. M. ROGERS (2014). Gene expression plasticity evolves in response to colonization of freshwater lakes in threespine stickleback. *Molecular Ecology* **23**, p. 3226–3240. (cf. p. 69).
- MORRISSEY, M. B. (2014). In search of the best methods for multivariate selection analysis. *Methods in Ecology and Evolution* **5**, p. 1095–1109 (cf. p. 174).
- (2015). Evolutionary quantitative genetics of non-linear developmental systems. *Evolution in press*, (cf. p. 141, 164, 169, 179).
- MORRISSEY, M. B., C. A. WALLING, A. J. WILSON, J. M. PEMBERTON, T. H. CLUTTON-BROCK & L. E. B. KRUK (2012). Genetic Analysis of Life-History Constraint and Evolution in a Wild Ungulate Population. *The American Naturalist* **179**, E97–E114. (cf. p. 166).
- MOSER, G., S. H. LEE, B. J. HAYES, M. E. GODDARD, N. R. WRAY & P. M. VISSCHER (2015). Simultaneous Discovery, Estimation and Prediction Analysis of Complex Traits Using a Bayesian Mixture Model. *PLoS Genet* **11**, e1004969. (cf. p. 143).
- MOUSSEAU, T. A. & H. DINGLE (1991). Maternal effects in insect life histories. *Annual review of entomology* **36**, p. 511–534 (cf. p. 77).
- MULLER, P., G. PARMIGIANI & K. RICE (2006). FDR and Bayesian multiple comparisons rules. *Johns Hopkins University, Dept. of Biostatistics Working Papers* (cf. p. 149, 157).
- NAKAGAWA, S. & H. SCHIELZETH (2010). Repeatability for Gaussian and non Gaussian data : a practical guide for biologists. *Biological Reviews* **85**, p. 935–956. (cf. p. 166, 167, 170, 171, 187).
- NARUM, S. R. & J. E. HESS (2011). Comparison of FST outlier tests for SNP loci under selection. *Molecular Ecology Resources* **11**, p. 184–194. (cf. p. 33, 44, 45).
- NEI, M. & T. MARUYAMA (1975). Lewontin-Krakauer test for neutral genes. *Genetics* **80**, p. 395 (cf. p. 26).
- NICHOLSON, G., A. V. SMITH, F. JÓNSSON, O. GÚSTAFSSON, K. STEFÁNSSON & P. DONNELLY (2002). Assessing population differentiation and isolation from single-nucleotide polymorphism data. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **64**, p. 695–715 (cf. p. 27, 54, 78).
- OVASKAINEN, O., M. KARHUNEN, C. ZHENG, J. M. C. ARIAS & J. MERILÄ (2011). A new method to uncover signatures of divergent and stabilizing selection in quantitative traits. *Genetics* **189**, p. 621–632. (cf. p. 71, 77, 81, 82, 108, 112, 115, 116, 127, 132).
- PARDO-DIAZ, C., C. SALAZAR & C. D. JIGGINS (2015). Towards the identification of the loci of adaptive evolution. *Methods in Ecology and Evolution* **6**, p. 445–464. (cf. p. 75, 76, 78, 79, 112).

- PARMESAN, C. (2006). Ecological and Evolutionary Responses to Recent Climate Change. *Annual Review of Ecology, Evolution, and Systematics* 37, p. 637–669 (cf. p. 12).
- PENG, B. & M. KIMMEL (2005). simuPOP : a forward-time population genetics simulation environment. *Bioinformatics* 21, p. 3686–3687. (cf. p. 33, 58).
- PÉREZ-FIGUEROA, A., M. J. GARCÍA-PEREIRA, M. SAURA, E. ROLÁN-ALVAREZ & A. CABALLERO (2010). Comparing three different methods to detect selective loci using dominant markers. *Journal of Evolutionary Biology* 23, p. 2267–2276. (cf. p. 30, 33, 44).
- PETERSON, B. K., J. N. WEBER, E. H. KAY, H. S. FISHER & H. E. HOEKSTRA (2012). Double digest RADseq : An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE* 7, e37135. (cf. p. 108, 115).
- PIANKA, E. R. (1970). On r- and K-selection. *The American Naturalist* 104, p. 592–597 (cf. p. 104).
- PIGLIUCCI, M., C. J. MURREN & C. D. SCHLICHTING (2006). Phenotypic plasticity and evolution by genetic assimilation. *Journal of Experimental Biology* 209, p. 2362–2367. (cf. p. 127, 136).
- PIMM, S. L., G. J. RUSSELL, J. L. GITTLEMAN & T. M. BROOKS (1995). The Future of Biodiversity. *Science* 269, p. 347–350. (cf. p. 12).
- POL, M. van de & J. WRIGHT (2009). A simple method for distinguishing within- versus between-subject effects using mixed models. *Animal Behaviour* 77, p. 753–758. (cf. p. 78).
- PONCET, B. N., D. HERRMANN, F. GUGERLI, P. TABERLET, R. HOLDEREGGER, L. GIELLY, D. RIOUX, W. THUILLER, S. AUBERT & S. MANEL (2010). Tracking genes of ecological relevance using a genome scan in two independent regional population samples of *Arabis alpina*. *Molecular Ecology* 19, p. 2896–2907. (cf. p. 22, 23, 86, 90).
- PRICE, G. R. (1970). Selection and covariance. *Nature* 227, p. 520–521 (cf. p. 176, 189).
- PRITCHARD, J. K. & A. DI RIENZO (2010). Adaptation – not by sweeps alone. *Nature Reviews Genetics* 11, p. 665–667. (cf. p. 45).
- PRITCHARD, J. K., M. STEPHENS & P. DONNELLY (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, p. 945–959 (cf. p. 19).
- RAFTERY, A. E. (1995). Bayesian model selection in social research. *Sociological methodology* 25, p. 111–164 (cf. p. 45).
- RAJON, E. & J. B. PLOTKIN (2013). The evolution of genetic architectures underlying quantitative traits. *Proceedings of the Royal Society B : Biological Sciences* 280, p. 20131552. (cf. p. 143).
- RELLSTAB, C., F. GUGERLI, A. J. ECKERT, A. M. HANCOCK & R. HOLDEREGGER (2015). A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology* 24, p. 4348–4370. (cf. p. 69, 75, 78, 79, 81, 139).
- RIEBLER, A., L. HELD & W. STEPHAN (2008). Bayesian variable selection for detecting adaptive genomic differences among populations. *Genetics* 178, p. 1817–1829. (cf. p. 54).
- RITLAND, K. (2000). Marker-inferred relatedness as a tool for detecting heritability in nature. *Molecular Ecology* 9, p. 1195–1204 (cf. p. 76).
- ROACH, D. A. & R. D. WULFF (1987). Maternal effects in plants. *Annual Review of Ecology and Systematics* 18, p. 209–235 (cf. p. 77, 124).
- ROBERTSON, A. (1966). A mathematical model of the culling process in dairy cattle. *Animal Production* 8, p. 95–108 (cf. p. 176, 189).
- (1968). « Population biology and evolution ». New York (US) : Syracuse University Press, p. 5–16 (cf. p. 176).
- ROBERTSON, A. (1950). Heritability of threshold characters. *Genetics* 35, p. 212–236 (cf. p. 169, 170).
- ROCKMAN, M. V. (2012). The QTN program and the alleles that matter for evolution : all that’s gold does not glitter. *Evolution* 66, p. 1–17. (cf. p. 33, 45, 142).
- ROFF, D. A. (1997). *Evolutionary quantitative genetics*. New York (US) : Chapman & Hall. (Cf. p. 170).
- (2001). *Life history evolution*. Sunderland, MA (US) : Sinauer Associates, Inc. (Cf. p. 101).
- ROFF, D. A. (2007). A Centennial Celebration for Quantitative Genetics. *Evolution* 61, p. 1017–1032. (cf. p. 142).
- ROQUET, C., F. C. BOUCHER, W. THUILLER & S. LAVERGNE (2013). Replicated radiations of the alpine genus *Androsace* (Primulaceae) driven by range expansion and convergent key innovations. *Journal of Biogeography* 40, p. 1874–1886. (cf. p. 113, 136).
- SAVOLAINEN, O., M. LASCoux & J. MERILÄ (2013). Ecological genomics of local adaptation. *Nature Reviews Genetics* 14, p. 807–820. (cf. p. 71, 76, 111, 112).
- SCHMITZ-LINNEWEBER, C. & I. SMALL (2008). Pentatricopeptide repeat proteins : a socket set for organelle gene expression. *Trends in Plant Science* 13, p. 663–670. (cf. p. 123, 137).
- SHENDURE, J. & H. JI (2008). Next-generation DNA sequencing. *Nature Biotechnology* 26, p. 1135–1145. (cf. p. 32, 53).
- SLAVOV, G. T., R. NIPPER, P. ROBSON, K. FARRAR, G. G. ALLISON, M. BOSCH, J. C. CLIFTON-BROWN, I. S. DONNISON & E. JENSEN (2014). Genome-wide association studies and prediction of 17 traits related to phenology, biomass and cell wall composition in the energy grass *Miscanthus sinensis*. *New Phytologist* 201, p. 1227–1239. (cf. p. 78).
- SOBER, E. (2008). *Evidence and evolution : the logic behind the science*. Cambridge (UK) : Cambridge University Press (cf. p. 7, 138).
- SPLITZ, K. (1993). Population structure in *Daphnia obtusa* : quantitative genetic and allozymic variation. *Genetics* 135, p. 367–374 (cf. p. 16, 73, 74, 112, 116).
- STEIN, C. M. (1973). Estimation of the mean of a multivariate normal distribution. *Proceedings of the Prague Symposium on Asymptotic Statistics* 1, p. 345–381 (cf. p. 169, 176).
- STINCHCOMBE, J. R. (2014). « Cross-pollination of plants and animals : wild quantitative genetics and plant evolutionary genetics ». *Quantitative Genetics in the Wild*. Sous la dir. d’A. CHARMANTIER, D. GARANT & L. E. KRUK. Oxford (UK) : Oxford University Press, p. 128–146 (cf. p. 74).
- STINCHCOMBE, J. R., C. WEINIG, M. UNGERER, K. M. OLSEN, C. MAYS, S. S. HALLDORSOTTIR, M. D. PURUGGANAN & J. SCHMITT (2004). A latitudinal cline in flowering time in *Arabidopsis thaliana* modulated by the flowering time gene *FRIGIDA*. *Proceedings of the National Academy of Sciences of the United States of America* 101, p. 4712–4717. (cf. p. 112).



- STONE, G. N., S. NEE & J. FELSENSTEIN (2011). Controlling for non-independence in comparative analysis of patterns across populations within species. *Philosophical Transactions of the Royal Society B : Biological Sciences* **366**, p. 1410–1424. (cf. p. 127).
- STOREY, J. D. (2003). The positive false discovery rate : A Bayesian interpretation and the q-value. *Annals of Statistics*, p. 2013–2035 (cf. p. 30, 117, 157).
- STOREY, J. D. (2002). A direct approach to false discovery rates. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)* **64**, p. 479–498 (cf. p. 57, 157).
- STOREY, J. D., J. E. TAYLOR & D. SIEGMUND (2004). Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates : a unified approach. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)* **66**, p. 187–205. (cf. p. 33, 35–37, 46).
- STOREY, J. D. & R. TIBSHIRANI (2003). Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences* **100**, p. 9440–9445. (cf. p. 29, 33, 35–37, 46, 117, 149).
- STOUFFER, S. A., E. A. SUCHMAN, L. C. DEVINNEY, S. A. STAR & R. M. J. WILLIAMS (1949). *The American soldier, Vol.1 : adjustment during army life*. Princeton University Press. Princeton, New Jersey (US) (cf. p. 117).
- STRANGER, B. E., E. A. STAHL & T. RAJ (2011). Progress and promise of genome-wide association studies for Human complex trait. *Genetics* **187**, p. 367–383. (cf. p. 45).
- TEDDER, A., S. CARLEIAL, M. GOŁĘBIEWSKA, C. KAPPEL, K. K. SHIMIZU & M. STIFT (2015). Evolution of the selfing syndrome in *Arabis alpina* (Brassicaceae). *PLoS ONE* **10**, e0126618. (cf. p. 22, 90).
- TORÄNG, P., J. WUNDER, J. R. OBESO, M. HERZOG, G. COUPLAND & J. ÅGREN (2015). Large-scale adaptive differentiation in the alpine perennial herb *Arabis alpina*. *New Phytologist* **206**, p. 459–470. (cf. p. 22, 73, 90, 124).
- UPTMOOR, R., W. WENZEL, W. FRIEDT, G. DONALDSON, K. AYISI & F. ORDON (2003). Comparative analysis on the genetic relatedness of *Sorghum bicolor* accessions from Southern Africa by RAPDs, AFLPs and SSRs. *Theoretical and Applied Genetics* **106**, p. 1316–1325. (cf. p. 76, 77).
- « Use of CAD genes to increase nitrogen use efficiency and low nitrogen tolerance to a plant » (2015). US9018445 B2. B. J. VINOCUR, S. AYAL, A. DIBER, E. EMMANUEL, G. RONEN, M. GANG, D. DIMET, H. KARCHI, Y. HERSCHKOVITZ & E. LTD (cf. p. 126, 137).
- VATSIU, A. I., E. BAZIN & O. E. GAGGIOTTI (2015). Detection of selective sweeps in structured populations : a comparison of recent methods. *Molecular Ecology*. (cf. p. 70).
- VILAS, A., A. PÉREZ-FIGUEROA & A. CABALLERO (2012). A simulation study on the performance of differentiation-based methods to detect selected loci using linked neutral markers. *Journal of Evolutionary Biology* **25**, p. 1364–1376. (cf. p. 27, 33, 44, 45).
- VILLELLAS, J., D. F. DOAK, M. B. GARCÍA & W. F. MORRIS (2015). Demographic compensation among populations : what is it, how does it arise and what are its implications ? *Ecology Letters* **18**, p. 1139–1152. (cf. p. 102).
- VILLEMEREUIL, P. de, É. FRICHOT, É. BAZIN, O. FRANÇOIS & O. E. GAGGIOTTI (2014). Genome scan methods against more complex models : when and how much should we trust them ? *Molecular Ecology* **23**, p. 2006–2019. (cf. p. 54, 58, 65, 78, 158, 160–162).
- VILLEMEREUIL, P. de & O. E. GAGGIOTTI (2015). A new FST-based method to uncover local adaptation using environmental variables. *Methods in Ecology and Evolution*. (cf. p. 117, 122).
- VILLEMEREUIL, P. de, O. E. GAGGIOTTI, M. MOUTERDE & I. TILL-BOTTRAUD (2015). Common garden experiments in the genomic era : new perspectives and opportunities. *Heredity*. (cf. p. 112).
- VILLEMEREUIL, P. de, O. GIMENEZ & B. DOLIGEZ (2013). Comparing parent–offspring regression with frequentist and Bayesian animal models to estimate heritability in wild populations : a simulation study for Gaussian and binary traits. *Methods in Ecology and Evolution* **4**, p. 260–275. (cf. p. 131, 166, 170).
- VILLEMEREUIL, P. de, J. A. WELLS, R. D. EDWARDS & S. P. BLOMBERG (2012). Bayesian models for comparative analysis integrating phylogenetic uncertainty. *BMC Evolutionary Biology* **12**, p. 102. (cf. p. 132).
- VITALIS, R., K. DAWSON & P. BOURSOT (2001). Interpretation of variation across marker loci as evidence of selection. *Genetics* **158**, p. 1811–1823 (cf. p. 32).
- VITALIS, R., M. GAUTIER, K. J. DAWSON & M. A. BEAUMONT (2014). Detecting and measuring selection from gene frequency data. *Genetics* **196**, p. 799–817. (cf. p. 27).
- VOS, P. et al. (1995). AFLP : a new technique for DNA fingerprinting. *Nucleic Acids Research* **23**, p. 4407–4414. (cf. p. 17, 76).
- WADDINGTON, C. H. (1953). Genetic assimilation of an acquired character. *Evolution* **7**, p. 118–126. (cf. p. 127, 136).
- (1959). Canalization of development and genetic assimilation of acquired characters. *Nature* **183**, p. 1654–1655. (cf. p. 127, 136).
- WALSH, B. & M. LYNCH. « Evolution and selection of quantitative traits ». Sunderland, Massachusetts (US) : Sinauer Associates (cf. p. 164, 167, 174).
- WANG, J. (2004). Sibship reconstruction from genetic data with typing errors. *Genetics* **166**, p. 1963–1979. (cf. p. 77).
- (2012). Computationally efficient sibship and parentage assignment from multilocus marker data. *Genetics* **191**, p. 183–194. (cf. p. 77).
- (2014). Marker-based estimates of relatedness and inbreeding coefficients : an assessment of current methods. *Journal of Evolutionary Biology* **27**, p. 518–530. (cf. p. 74).
- WANG, R., M. C. ALBANI, C. VINCENT, S. BERGONZI, M. LUAN, Y. BAI, C. KIEFER, R. CASTILLO & G. COUPLAND (2011). Aa TFL1 confers an age-dependent response to vernalization in perennial *Arabis alpina*. *The Plant Cell* **23**, p. 1307–1321. (cf. p. 22).
- WANG, R., S. FARRONA, C. VINCENT, A. JOECKER, H. SCHOOF, F. TURCK, C. ALONSO-BLANCO, G. COUPLAND & M. C. ALBANI (2009). PEP1 regulates perennial flowering in *Arabis alpina*. *Nature* **459**, p. 423–427. (cf. p. 22, 90, 91, 100, 101).
- WEIR, B. S. & C. C. COCKERHAM (1984). Estimating F-statistics for the analysis of population structure. *Evolution* **38**, p. 1358–1370. (cf. p. 9, 19, 118).

- WHITTAKER, R. J. (1993). Plant population patterns in a glacier foreland succession : pioneer herbs and later-colonizing shrubs. *Ecography* 16, p. 117–136. (cf. p. 22, 85).
- WIENKOOP, S., D. ZOELLER, B. EBERT, U. SIMON-ROSLIN, J. FISAHN, M. GLINSKI & W. WECKWERTH (2004). Cell-specific protein profiling in *Arabidopsis thaliana* trichomes : identification of trichome-located proteins involved in sulfur metabolism and detoxification. *Phytochemistry*. Proteomics 1 65, p. 1641–1649. (cf. p. 126, 137).
- WILLING, E.-M. et al. (2015). Genome expansion of *Arabidopsis thaliana* linked with retrotransposition and reduced symmetric DNA methylation. *Nature Plants* 1, p. 14023. (cf. p. 90, 108, 129).
- WILSON, A. J. (2008). Why  $h^2$  does not always equal VA/VP? *Journal of Evolutionary Biology* 21, p. 647–650. (cf. p. 167, 173, 178).
- WILSON, A. J., M. B. MORRISSEY, M. J. ADAMS, C. A. WALLING, F. E. GUINNESS, J. M. PEMBERTON, T. H. CLUTTON-BROCK & L. E. B. KRUK (2011). Indirect genetics effects and evolutionary constraint : an analysis of social dominance in red deer, *Cervus elaphus*. *Journal of Evolutionary Biology* 24, p. 772–783. (cf. p. 166).
- WILSON, A. J., D. RÉALE, M. N. CLEMENTS, M. M. MORRISSEY, E. POSTMA, C. A. WALLING, L. E. B. KRUK & D. H. NUSSEY (2010). An ecologist's guide to the animal model. *Journal of Animal Ecology* 79, p. 13–26. (cf. p. 77, 164, 166).
- WINGLER, A., M. JUVANY, C. CUTHBERT & S. MUNNÉ-BOSCH (2015). Adaptation to altitude affects the senescence response to chilling in the perennial plant *Arabidopsis thaliana*. *Journal of Experimental Botany* 66, p. 355–367. (cf. p. 22, 90, 100).
- WINGLER, A., E. J. STANGBERG, T. SAXENA & R. MISTRY (2012). Interactions between temperature and sugars in the regulation of leaf senescence in the perennial herb *Arabidopsis thaliana*. *Journal of Integrative Plant Biology* 54, p. 595–605. (cf. p. 90).
- WOLAK, M. E. & L. F. KELLER (2014). « Dominance genetic variance and inbreeding in natural populations ». *Quantitative Genetics in the Wild*. Sous la dir. d'A. CHARMANTIER, D. GARANT & L. E. KRUK. Oxford (UK) : Oxford University Press, p. 104–127 (cf. p. 74, 77).
- WRIGHT, S. (1934). An analysis of variability in number of digits in an inbred strain of Guinea pigs. *Genetics* 19, p. 506–536 (cf. p. 170).
- (1949). The genetical structure of populations. *Annals of eugenics* 15, p. 323–354 (cf. p. 19).
- YEAMAN, S. & M. C. WHITLOCK (2011). The genetic architecture of adaptation under migration–selection balance. *Evolution* 65, p. 1897–1911. (cf. p. 45, 143).
- ZULLIGER, D., E. SCHNYDER & F. GUGERLI (2013). Are adaptive loci transferable across genomes of related species? Outlier and environmental association analyses in Alpine Brassicaceae species. *Molecular Ecology* 22, p. 1626–1639. (cf. p. 90).

## Résumé de la thèse

L'adaptation locale est un phénomène micro-évolutif qui peut survenir lorsque des populations d'une même espèce sont exposées à des conditions environnementales différentes. Si cet environnement exerce une pression sous forme de sélection naturelle, qu'il existe un potentiel adaptatif au sein des populations et que le flux de gènes est suffisamment modéré, les populations vont alors tendre vers un optimum adaptatif local. Dans cette thèse, je m'intéresse aux moyens méthodologiques de l'étude de l'adaptation locale d'une part, et à l'étude de ce phénomène le long d'un gradient d'altitude chez la plante alpine *Arabis alpina* d'autre part.

Dans la première partie méthodologique, je montre que les méthodes de scan génomique pour détecter les marqueurs génétiques sous sélection peuvent souffrir de forts taux de faux positifs lorsqu'exposées à des jeux de données complexes, mais réalistes. Je présente ensuite une méthode statistique de détection de marqueurs génétiques sous sélection qui, contrairement aux méthodes existantes, utilisent à la fois la notion de différenciation génétique (ou  $F_{ST}$ ) et une information environnementale. Cette méthode a été développée de manière à limiter son taux de faux positifs de manière générale. J'offre enfin une perspective concernant les liens entre une expérience ancienne en biologie évolutive (l'expérience de jardin commun) et les nouveaux développements moléculaires et statistiques modernes.

Dans la seconde partie empirique, je présente une analyse de la démographie d'*A. alpina* dans six populations naturelles. Outre qu'elle révèle des caractéristiques biologiques intéressantes sur cette espèce (faible espérance de vie, reproduction et survie très différentielle...), cette analyse montre que la croissance diminue et la survie augmente chez cette espèce avec la baisse de la température moyenne (donc avec l'altitude). Puisque ces analyses ne permettent pas d'exclure des hypothèses de dérive et de plasticité phénotypique, je présente une étude en jardin commun sur *A. alpina* qui permet de lier les problèmes de plasticité phénotypique et qui, combinée à des analyses moléculaires, permettent d'exclure l'hypothèse de dérive. Les résultats montrent qu'il existe un syndrome phénotypique adaptatif lié à la température moyenne qui tend à des plantes plus petites, plus compactes, qui croissent et se reproduisent moins, dans les milieux froids. À l'aide des données moléculaires et de méthodes de scan génomique, je présente une liste de 40 locus qui peuvent être impliqués dans ce processus.

Pour finir, je discute l'ensemble de ces résultats empiriques dans un contexte plus général d'écologie alpine. Je résume ensuite les principaux obstacles méthodologiques à l'étude de l'adaptation locale et je fournis quelques perspectives méthodologiques.

## Abstract

Local adaptation is a micro-evolutionary phenomenon, which arises when populations of the same species are exposed to contrasted environmental conditions. If this environment exerts some natural selection pressure, if an adaptive potential exists among the populations and if the gene flow is sufficiently mild, populations are expected to tend toward a local adaptive optimum. In this thesis, I study the methodological means of the study of local adaptation on the one hand, and I investigate this phenomenon along an elevation gradient in the alpine plant *Arabis alpina* on the other hand.

In the first, methodological part, I show that the genome scan methods to detect selection using genetic markers might suffer strong false positive rates when confronted to complex but realistic datasets. I then introduce a statistical method to detect markers under selection, which, contrary to existing methods, make use of both the concept of genetic differentiation (or  $F_{ST}$ ) and environmental information. This method has been developed in order to reduce its global false positive rate. Finally, I present some perspectives regarding the relationships between the relatively old "common garden" experiment and the new developments in molecular biology and statistics.

In the second, empirical part, I introduce an analysis of the demographic characteristics of *A. alpina* in six natural populations. Besides providing interesting biological information on this species (low life expectancy, strongly contrasted reproduction and survival...), these analyses show that growth increase and survival decrease with the decrease of average temperature (hence with altitude). Since these analyses do not allow us to rule out hypotheses such as drift and phenotypic plasticity, I show the results of a common garden experiment which enable us to smooth phenotypic plasticity and, when combined with molecular data, enable us to rule out the hypothesis of drift. The results show the existence of an adaptive phenotypic syndrome, in which plants are smaller, are more compact, grow slower and reproduce less in cold temperature environments. Using the molecular data, I draw a list of 40 loci which might be involved in this adaptive process.

In the end, I discuss these empirical findings as a whole to place them in a more general context of alpine ecology. I sum up the main methodological challenges when studying local adaptation and offer some methodological perspectives.