THÈSE

PRÉSENTÉE À

L'UNIVERSITÉ BORDEAUX 1

ECOLE DOCTORALE DES SCIENCES ET ENVIRONNEMENT

Par Jean-Paul SOULARUE

POUR OBTENIR LE GRADE DE

DOCTEUR

SPÉCIALITÉ : ÉCOLOGIE ÉVOLUTIVE, FONCTIONNELLE ET DES COMMUNAUTÉS

Évolution de la phénologie des arbres à l'échelle d'un paysage forestier

Directeur de recherche : Antoine KREMER

Soutenue le : 13 Décembre 2012

Devant la commission d'examen formée de :

M.	**LEFEVRE, François**	DR	INRA Avignon	Rapporteur
Mme.	**RONCE, Ophélie**	CR	CNRS Montpellier	Rapporteur
Mme.	**LE CORRE, Valérie**	CR	INRA Dijon	Examinateur
M.	**RODOLPHE, François**	DR	INRA Jouy en Josas	Examinateur
M.	**SHERMAN, David**	DR	INRIA Bordeaux	Président
M	**KREMER, Antoine**	DR	INRA Bordeaux	Directeur de thèse

Université Bordeaux 1
Les Sciences et les Techniques au service de l'Homme et de l'environnement

Remerciements

Je remercie chaleureusement mon directeur de thèse Antoine Kremer pour sa confiance, sa disponibilité, son enthousiasme, ses conseils et toutes les discussions que nous avons eues.

Je remercie grandement Sally Aitken et toute son équipe de m'avoir si bien accueilli pendant ces deux mois passés à Vancouver. Ce fut une expérience enrichissante dont je garde un excellent souvenir.

Je remercie toutes les personnes avec qui j'ai pu interagir durant ces trois années et qui ont contribué à la réalisation de ce travail par leur réponses à mes questions, leur aide et leurs conseils : Sally Aitken, David Sherman, Sophie Gerber, Alain Franc, Marie-Laure Desprez-Loustau, Valérie Le Corre, Frédéric Raspail, Arndt Hampe, Jean-Marc Frigério, Pauline Garnier-Géré, Stéphanie Mariette, Rémy Petit, Christophe Plomion, Virgil Fievet, Sylvain Delzon, Annabel Porté, Philippe Chaumeil, Jean-Marc Gion, Jean-Baptiste Lamy, Loïc Kerdraon, Thierry Labbé, Laetitia Pacaly, Susannah Tysor, Sam Yeaman, Véronique Lozano, Florence Le Pierres, Chantal Bouquet. En espérant n'oublier personne...

Venir travailler à l'UMR Biogeco durant ces trois années fut un plaisir. Je remercie mes collègues pour nos échanges et les moments partagés. Notamment Pierre, François, Eric, Sarah, Audrey A, Audrey JG, Emilie C, Emilie V, Christophe, Jean-Baptiste, Florian, Sophie, Stéphanie, Laure, Gilles et bien d'autres.

J'ai bien sûr une pensée spéciale pour mes proches.

Enfin, je remercie les membres du jury, notamment les deux rapporteurs, d'accepter d'examiner ce travail.

Résumé

La phénologie du débourrement est un caractère adaptatif majeur, sensible aux variations de température. Prédire l'évolution des forêts naturelles tempérées sous l'influence de changements environnementaux nécessite de pouvoir expliquer l'origine des patrons de différentiation clinaux observés pour ce caractère à l'échelle de paysages. Il a été démontré expérimentalement que le débourrement végétatif se produisait en même temps que la floraison. Cela suggère que les croisements se font préférentiellement entre arbres présentant des phénologies du débourrement similaires ; c'est ce que l'on appelle l'homogamie. Alors que la plupart des interprétations de clines de différenciation génétique soulignent l'influence de la sélection divergente, les spécificités de la phénologie du débourrement et ses conséquences sur le système de reproduction sont rarement considérées.

A travers une approche par modélisation nous montrons ici dans un premier temps que la seule interaction entre homogamie et flux de pollen peut générer une différentiation génétique clinale à l'échelle d'un paysage sans aucune pression de sélection. Dans un tel contexte théorique, le filtrage des flux de pollen réalisé par l'homogamie en présence d'un gradient environnemental différencie progressivement et durablement les populations.

Dans un second temps, nous montrons que l'homogamie amplifie la réponse adaptative des populations à la sélection *co-gradient* alors qu'il la contraint dans le cas de sélection *contre-gradient*. Nous montrons par ailleurs que l'homogamie peut induire une différentiation clinale en cas de sélection uniforme.

Mots-clés : Phénologie, flux de gènes, homogamie, métapopulation, differenciation, adaptation, modélisation.

Abstract

Timing of bud burst (TBB) is a key adaptive trait affected by temperature variations. Predicting the evolution of natural forests undergoing environmental variations requires to understand the evolutionary dynamics that have resulted in the strong patterns of differentiation characterized for this trait. It has been shown experimentally that the TBB was strongly correlated with the timing of flowering. This suggests that trees having similar TBB tend to mate preferentially, making assortative mating at TBB the default reproduction regime within tree species. Clinal patterns of genetic differentiation have been mostly interpreted as resulting from divergent selection, however, few studies have considered the peculiar features of timing of bud burst.

Through a modelling approach based on quantitative genetics models, we first demonstrate here that the sole interaction between assortative mating at TBB and pollen flow can induce a clinal differentiation among populations without any selection pressure. In a such theoretical context, assortative mating filters pollen flow in presence of environmental gradients and progressively shifts the genetic values of populations. Then, we demonstrate that assortative mating amplifies the adaptive response of populations to *co-gradient* selection, and constrains it in the case of *counter-gradient* selection. Finally, we show that assortative mating differentiates populations even in the case of uniform selection.

Keywords : Phenology, gene flow, assortative mating, metapopulation, differentiation, adaptation, modelling.

Content

Introduction

The importance of forests

Forests ecosystems cover more than a third of the earth's terrestrial surface *[FAO]*. The numerous tree species they include host abundant biodiversity and provide keystone ecological services through tight interactions with other biosphere's ecosystems; contributing to climate regulation via carbon sequestration, protecting soils from erosion, regulating water cycle, enriching soils by supplying organic matter. Forests occupy an important place in human societies, mainly because of the renewable raw material they provide, as wood, fiber and biomass. Forests represent also three quarters of the earth's terrestrial biomass *[Aitken et al., 2008]*, their advised exploitation could be an opportunity to reduce the use of fossil energies in the coming years. As the rapid change in climatic conditions occurring at a global scale clearly threatens forests lands *[Keenan et al., 2010]*, there is an urgent need for governments and practitioners to know accurately how forests can respond to environmental changes in order to set up relevant management programs.

Predicting the fate of forests in the context of climate change: ecological modelling

Forests and their environments compose complex sets made of numerous interacting entities that can be grasped from multiple perspectives. Understanding such sets and predicting their evolution requires to integrate huge quantities of data available at multiple levels. Nowadays, ecological modelling is the most commonly used approach to predict the fate of tree species at large scales *[Cheaib et al., 2012]*. Based on sensitivity of species to existing environmental conditions, it aims at predicting the evolution of species distributions by focusing on the change in several environmental factors. Over the last years, this kind of approach has mainly been implemented through two categories of models: niche models and process-based models *[Aitken et al., 2008] [Morin and Thuiller, 2009][Cheaib et al., 2012]*. Niche models produce bioclimatic envelops that define the abiotic environmental conditions suitable for the survival of a species of interest. These envelops are computed from multivariate statistics models accounting for factors characterizing environmental conditions -temperature, precipitations...- and the species considered. The future distribution of the species is then directly established from the transposition of predicted changes in environmental variables for the period targeted. Process-based models focus rather on the influence of environment on physiological processes thought of primary importance for the survival of a species. The process targeted can be related to a phenological response (*Phenophit*) *[Chuine and Beaubien, 2001]* or to any other kind of biological process like photsynthesis, C02 allocation (*Castanea*) *[Le Dantec et al., 2000]*. Reproducing success probabilities (*Phenophit*) or growth rates (*Castanea*) are then used for the prediction of future distributions. Interestingly, niche and process-based models can be fed with various sources that can directly originate either from static datasets (*e.g.* produced by meteorological stations) or from other models such as greenhouse gaz emission models. Recent articles devoted to the analysis of the frameworks designed for the prediction of the evolution of the distributions of tree species *[Cheaib et al, 2012]*, or more generalist review *[Bellard et al., 2012]*, yield more exhaustive and accurate descriptions of the main tools available in this field.

Overall, the benefits of ecological modelling are considerable. First of all, it allows to make predictions over large scales encompassing areas where populations might have to face important environmental changes. This approach has also an interesting integrative potential as some models can be coupled with other models providing predictions of changes in environmental variables. For instance, Thuiller *et al.* (2003) combined successfully their niche model *Biomod* with different greenhouse gaz emission and climatic models in order to predict the evolution of *Quercus petraea* up to the end of the century at the paneuropean scale *[Thuiller et al., 2003]*.

Nonetheless, the recent review of Bellard *et al.* (2012) and a comparative analysis realized by Cheaib *et al.* (2012) set the limits of this approach. The ecological models available are indeed often strongly specialized and show limitations *[Bellard et al., 2012]*. The analysis of Cheaib et *al. (2012)* showed also that the simulations of similar scenarios from available models can lead to very contrasted predictions depending on the species and the scenarios considered. A telling example in that study, is the impossibility for all of the tested frameworks to reconstruct the current distribution of Scots Pine in France *[Cheiab et al., 2012]*.
A striking characteristic of these tools is that they do not, or very indirectly, account for the evolutionary potential of populations *[Soberon and Nakmura, 2009][Cheaib et al., 2012][Hoffman and Sgrö, 2012]* despite the fact that previous works stressed the importance of adaptation under radical environmental changes *[Davis and Shaw, 2001][Chevin et al., 2012]*. Likewise, Chevin *et al.* (2010) stated more radically that niche- and process-based models cannot predict efficiently the evolution of species distributions owing to the too many bias they include *[Chevin et al., 2010]*. Of course, correlative approaches implicitly account for past evolutionary processes but, as reminded by Elith *et al* (2010), the relationship between environmental conditions and presence of species is not stable *[Elith et al. 2010]*. As a result, the elusive and contrasted predictions available cannot be of great help for practitioners *[Bellard et al., 2012][Cheaib et al., 2012]*. These criticisms and observations underline the necessity of improving the quality of predictions and indicate clearly the direction of future improvements to be done: the theoretical frameworks used should be more integrative and consider explicitly the evolutionary potential of natural forests.

Evolutionary potential of tree populations

Rapid and sustainable environmental changes are likely to induce new selection pressures over long periods on tree populations *[Lynch and Lande, 1993][Hoffman and Sgro, 2012]* that can respond in three different ways *[Aitken et al., 2008]*. First, they can adapt and persist locally, either by adaptation involving genetic changes or by phenotypic plasticity that can also be heritable *[Chevin and Lande, 2009]*. The resulting phenotypic changes can be *temporal* when the adaptive response concerns a phenological trait or truly *local* when a physiological adaptive character is concerned *[Bellard et al., 2012]*. Second, populations can migrate to more suitable habitats via seed dispersal. Third, populations can become extinct. Overall, the persistence of tree populations in the context of environmental change is likely to result from a combination of adaptive and migration mechanisms *[Davis and Shaw, 2001]*.
Assessing the evolutionary potential of tree populations requires to identify the relevant adaptive characters sensitive to the change in environmental conditions predicted. In the context of current global warming, phenological traits are receiving considerable interest *[Donnely et al., 2012] [Rutishauer and Stockli, 2012]*, and changes have already been noticed for multiple species *[Parmesan et al. 2006]*. Regarding temperate tree species, the timing of bud burst (TBB), has received much attention over the last years *[Vitasse et al., 2011][Alberto et al., 2011]*. TBB is indeed a highly heritable trait (*e.g.* $h^2 > 0.8$ for oaks *[Alberto et al., 2011]*), sensitive to temperature variations, and it strongly influences fitness values. Indeed, TBB is a major component of growth that determines the length of the growing season of trees *[Alberto et al., 2011]*. Moreover, TBB is also tightly correlated with the timing of flowering *[Franjic et. al, 2011]*. This synchronization was observed experimentally by Franjic *et al.* (2011), who monitored the periods of shedding of pollen, receptivity of female flowers and leave unfolding within a population of 100 clonal oaks.
In situ monitoring of TBB across the distribution of various tree species has revealed that current global warming trigged earlier leaf flushing *[Menzel et al., 2006][Bertin et al, 2008][Nordli et al., 2008]*. These observations were confirmed by short-term warming experiments that aimed at assessing the temperature sensitivity of phenology of hundreds of plant species *[Morin et al., 2010]*. Even though it is difficult to relate short-term responses simulated in warming experiments and *in situ* observations, the results obtained converged. Interestingly, warming experiments yield

supplemental assessments of the average shift in phenology induced per degree. Globally, change in phenology due to temperature was estimated to be included between 1.9 and 3.3 days per degree in warming experiments for the tested plants. In comparison, it was monitored to range from 2.5 to 5 days per degree for plants *in situ [Wolkovitch et al., 2012]*. In the context of climate change, the magnitude of these estimations kindles worries about the capacity of tree populations to cope with the temperature increases predicted since the migration speed of tree species clearly not allow them to track in space the shifts of bioclimatic envelopes predicted *[Aitken et al.,2008]*. Furthermore, even if populations can adapt locally, the important shift in TBB induced by changes in environments are susceptible to expose them to drastic phenomena such as late frosts or ravaging insects, that could severely affect populations *[Alberto et al., 2011]*.

TBB shows high level of genetic variability

Numerous experiments based on common gardens have highlighted high levels of variability at the QTLs associated with TBB for multiple tree species. High heritabilities exceeding 0.8 have indeed been reported in oaks *[Alberto et al., 2011][Baliuckas and Pliura, 2003]*. Interestingly, high levels of differentiation have been characterized among populations *[Savolainen et al., 2007]*: Qst values computed over wide areas including numerous populations considerably exceed the level of differentiation usually observed for neutral markers (0.024) (table 1).

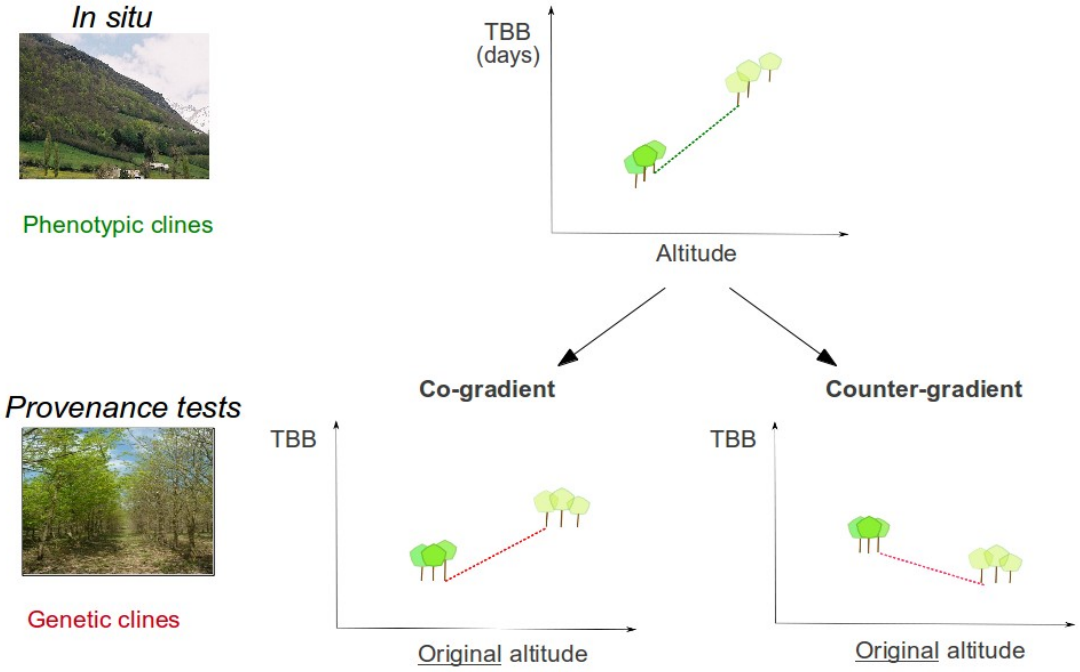
Species	Range	Qst	Populations	Reference
Quercus Petraea	Transect	0.20	10	*[Alberto et al., 2011]*
Quercus Robur	Wide	0.27	23	*[Jensen and Hansen, 2008]*
Quercus Petraea	Wide	0.55	115	*[Ducousso et al., 2005]*

Table 1 *Some examples of differentiation levels for TBB.*

Furthermore, the differentiation characterized often follows strong geographic clinal patterns varying with temperature along landscapes *[Savolainen et al., 2007][Aitken et al., 2008]*. Namely, the clines observed exhibit latitudinal, altitudinal, longitudinal directions or combinations of these directions. In addition, two kinds of clines have been characterized: phenotypic clines and genetic clines. Phenotypic differentiation is easily assessed trough *in situ* observations across the distribution of species of interest, and owing to the temperature effect on TBB *[Chuine and Cour, 1999]*, the slope of the phenotypic clines is of the same trend than the temperature clines associated to the variation of geographic variables. Thus, on average, natural populations occupying cold habitats, flush later than those occupying warm habitats. For instance, across altitudinal gradients, populations from low altitude flush earlier than populations from high altitude. On the other hand, genetic clines can be characterized within provenance tests where the TBB of distinct natural populations can be monitored under homogeneous environmental conditions. This approach has highlighted a linear relationship among the genetic values of populations and the temperature regimes reported in the habitats from which populations originate. Interestingly, two kinds of genetic clines have been characterized, depending on the species considered. In the first case genetic clines covary positively with phenotypic clines; namely, within provenance tests, the provenances originating from the warmer environments still flush earlier than the provenances originating from the colder environments. A simple illustration of this case can be obtained from two populations occupying different altitude levels: *in situ*, the population from the lowest altitude level flush earlier than the population from the highest altitude level and in provenance tests, the population from the lowest altitude level still flush earlier than the population from the highest altitude level (figure 1). Cases where the genetic and phenotypic clines exhibit similar trends correspond to co-gradient variations *[Conover, 1995][Conover et al., 2009],* which have mainly

been observed for oak species. However, this case is exceptional. Mostly, genetic clines are opposed to phenotypic clines: in provenance tests, the provenances coming from warm environments flush later than the provenances from cold environments, which corresponds to counter-gradient variation (figure 1) *[Conover, 1995]*.

Figure 1. Illustration of the co- and counter-gradient variations along an altitudinal gradient.



Adaptive potential of tree populations

Because the colonization ability of trees, depending on seed dispersal, is predicted to be lower than the shift of habitat *[Aitken et al., 2008]*, the persistence of tree populations is expected to be mainly ensured through phenotypic changes of adaptive traits supported both by phenotypic plasticity and evolutionary adaptation. In what follows, we will mainly focus on the adaptation ability of populations for TBB.

The high degree of polymorphism characterized within natural populations at the loci involved in the variations at TBB yields a general insight about their adaptive potential. Though in some cases, such high amounts of variability can be considered as a genetic load decreasing the mean fitness of populations owing to the existence of phenotypes diverging from local optimal values *[Bridle et al., 2009]*, the existence of genetic variants is a necessary condition for genetic adaptation. When the occurrence of environmental variations are accounted for, this variability is a crucial resource, allowing the evolutionary adaptation of populations undergoing new selection pressures induced *[Lynch and Lande, 1993]*. Since ongoing climate changes are expected to decrease the genetic variance within populations *[Botkin et al., 2007]*, this characteristic at phenological traits might be a key ensuring the persistence of temperate tree species.

In order to predict accurately the adaptive responses of tree populations to environmental changes,

it is necessary to understand how the clinal patterns of genetic variations observed have been shaped by the evolutionary processes. Of course, the wealth of evolutionary theory and the many existing analytical models give qualitative insights into the dynamics of populations undergoing environmental changes. As examples, well-known models formalizing the limits of evolutionary changes that can be stood by small populations *[Lynch and Lande 1993],* focusing on the effects of gene flow on adaptation *[Kirkpatrick and Barton, 1997]*, on the relation between demography and adaptation *[Gomulkievitz et Houle, 2009]*, or investigating the role of heritable phenotypic plasticity *[Chevin et al., 2010],* have clearly improved our understanding of the evolutionary dynamics of populations.
However, this kind of pure analytical approach does not allow easily for detailed predictions of population persistence at large scale and over long periods *[Chevin et al., 2010][Hoffman et Sgrö, 2012]*. Above all, these models are often highly specialized, which constrains the analysis and the inferences they allow for to the close neighborhoods of the hypothesis tests they were originally created for. Enhancing these models in order to make them more integrative is sometimes possible, for instance Chevin *et al.* (2010) extended the original model of Lynch and Lande (1993) by including phenotypic plasticity and finer environmental variations *[Chevin et al., 2010]*. However, mostly, describing mathematically the influence of multiple evolutionary processes without too stringent assumptions is challenging, and the general inferences produced are of low resolution. More importantly, existing models rarely account for essential peculiar features of tree species, particularly those associated with TBB. Therefore, the first step required before designing high integrative frameworks allowing for accurate predictions at the species scale, is to improve our understanding of the evolutionary mechanisms occurring within land forests by considering precisely the evolutionary consequences of the specific features of TBB.

Assortative mating at TBB

In trees and more generally in plants, mating individuals must have overlapping flowering periods; the emission of pollen by male flowers have to occur at the same time with the receptivity of female flowers. The resulting synchronization constraints induces a prezygotic reproducing isolation that makes assortative mating at flowering time the default reproduction regime within natural plants and trees species *[Kirkpatrick, 2000][Fox, 2003]*. Franjic et al. (2011) have observed that the flowering and the leafing periods are strongly correlated and that the difference in TBB within a single population may extend to several weeks *[Franjic et al., 2011]*. Interestingly, the receptivity period of female flowers extend to the best to few days *[Franjic et al., 2011]*, which suggests that crosses involve preferentially trees having similar TBB. Thereby, assortative mating occurring according to TBB is the default reproduction regime within natural forests, in what follows, *assortative mating* will systematically refer to assortative mating at TBB.
Two kinds of assortative mating can occur: positive and negative assortative mating. Positive assortative mating occurs within the limits of a single population, when early flushing individuals mate preferentially with other early individuals. Conversely, negative assortative mating occurs between individuals from distinct populations when late individuals from a population mate preferentially with early individuals of other populations, owing to the *in situ* phenotypic populations differences of TBB. The evolutionary consequences of assortative mating at quantitative traits under polygenic inheritance have largely been explored since the first works of Wright and Fisher on inbreeding *[Fisher, 1918][Wright, 1921][Fox, 2003][Devaux and Lande, 2008]*, however, mostly, the previous theory was established from investigations realized within single populations. Generally speaking, assortative mating generates linkage disequilibrium within populations at the loci involved in the trait variations. The increase in linkage disequilibrium generates an increase in genetic variance that amplifies the response to selection pressures within single populations *[Gianola, 1982][Jorjani, 1997]*. The evolutionary consequences of this mating system might therefore be crucial for tree populations and it might be interesting to further consider

them in realistic contexts relevant with tree species.

Gene flow

Gene flow through seed and pollen strongly impacts the adaptation and persistence of populations *[Kirkpatrick and Barton, 1997]*. It is thus an evolutionary process of key importance, however often roughly modeled *[Bellard et al., 2012],* partly because of uncertainties *[Kremer et al., 2012]*. Seed dispersal allows for the colonizations of new areas presenting suitable conditions for the survival of founding populations. On the other hand, pollen can disperse over very long distances that can exceed hundreds of kilometers *[Nathan et al. 2008][Bohrerova et al., 2009]* while remaining viable *[Schueler et al, 2005]*. Interestingly, the evolutionary consequences of gene flow are antagonistic *[Kremer et al., 2012]*. Overall, the introduction of non locally adapted genes within populations can induce a migration load diminishing the mean fitness values which may challenge recipient populations facing extensive selection pressures. Such constraining effects on adaptation also called swamping effect have been demonstrated theoretically several times *[García-Ramos and Kirkpatrick1997][Kirkpatrick and Barton, 1997]*, and suspected in empirical works *[Holliday et al., 2012]*. On the other hand, extensive gene flow does maintain high levels of variability within populations, which improve their adaptive potential *[Bridle et al., 2010]*.

Objectives

This thesis aims at yielding new elements improving our understanding of the interactions among the evolutionary processes that have shaped the spatial clinal patterns observed for TBB within tree species. Through a modelling approach, we focus on the interplay among assortative mating, gene flow and natural selection in realistic contexts regarding tree species. The relevance of the context relies on two essential points; first, our theoretical investigations involve multiple populations, a point often neglected within the previous theoretical works. Indeed, distributions of tree species range over very large areas and are often fragmented. Second, variations in position along large areas are associated to variations in temperature regimes, which probably result in contrasting selection pressures along environmental gradients. Indeed, the most common interpretations of the clinal differentiation observed among populations for TBB usually suspect divergent selection induced by biotic or abiotic selection pressures *[Alberto et al. 2011]*. However little is known about the selection pressures occurring *in natura* (which intensity ? Which optima ? Which model ?); we consider here clinal patterns of divergent selection.
Here, we first review the simulation packages allowing to simulate the evolution of quantitative characters in realistic ecological contexts, namely, allowing to consider evolution of multiple populations undergoing divergent selection, and connected by gene flow. We then explore deeply the evolutionary consequences of phenological assortative mating in a theoretical context excluding natural selection in order to know whether the sole interaction between assortative mating and gene flow could generate the clinal patterns of differentiation observed for TBB. Then, we consider more realistic settings where divergent selection, gene flow and assortative mating do interact. In this last study we aim to know whether assortative mating can mediate the adaptive response of tree populations by investigating the interplay among assortative mating, gene flow and divergent selection, with a peculiar focus on the prevalent directions of these three drivers.

References

Aitken SN, Yeaman S, Holliday JA, Wang T, Curtis-McLane S: Adaptation, migration or extirpation: **Climate change outcomes for tree populations**. *Evolutionary Applications* 2008, 1:95–111

Alberto F, Bouffier L, Louvet J-M, Lamy J.-B, Delzon S., Kremer A: **Adaptive responses for seed and leaf phenology in natural populations of sessile oak along an altitudinal gradient**. *J Evol Biol* 2011, 24:1442-1454

Baliuckas V, Pliura A: **Genetic variation and phenotypic plasticity of Quercus robur populations and open-pollinated families in Lithuania**. *Scand. J. Forest Res.* 2003, 18:305–319

Bellard C, Bertelsmeier C, Leadley P, Thuiller W, Courchamp F: **Impacts of climate change on the future of biodiversity**. *Ecology Letters, Blackwell Publishing Ltd* **2012***, 15:* 365-377

Bertin R I:**Plant phenology and distribution in relation to recent climate change**. *J. Torrey Bot. Soc.* 2008*, 135:*126-146

Bohrerova Z, Bohrer G, Cho KD, Bolch MA, Linden KG:**Determining the viability response of Pinus taeda pollen under abiotic stresses typical to atmospheric conditions during long distance dispersal**. *Ecol. Appl.* 2009, 19:656–667

Botkin DB, Saxe H, Araujo MB, Betts R, Bradshaw RH W, Cedhagen T et al*:* **Forecasting the effects of global warming on biodiversity**. *Bioscience* 2007*, 57:227–236*

Bridle JR, Polechová J, Kawata M and Butlin RK: **Why is adaptation prevented at ecological margins? New insights from individual-based simulations**. *Ecology Letters* 2010, 13:485–494, doi: 10.1111/j.1461-0248.2010.01442.x

Bridle JR, Polechova J, Vines TH: **Limits to adaptation and patterns of diversity**. In: Speciation and Patterns of Diversity (eds Butlin RK, Bridle JR, Schluter D). *Cambridge University Press* 2009 77–102

Cheaib A, Badeau V, Boe J, Chuine I, Delire C, Dufrêne E, François C, Gritti ES, Legay M, Pagé C, Thuiller W, Viovy N, Leadley P*:* **Climate change impacts on tree ranges: model intercomparison facilitates understanding and quantification of uncertainty**. *Ecology Letters 2012, 15:*533-544

Chevin L-M, Collins S and Lefèvre F: **Phenotypic plasticity and evolutionary demographic responses to climate change: taking theory out to the field**. *Functional Ecology* 2012*, doi: 10.1111/j.1365-2435.2012.02043.x*

Chevin L-M, Lande R, Mace GM: **Adaptation, Plasticity, and Extinction in a Changing Environment: Towards a Predictive Theory**. *PLoS Biol* 2010, *8(4): e1000357, doi:10.1371/journal.pbio.1000357*

Chevin LM and Lande R. **When do adaptive plasticity and genetic evolution prevent extinction of a density-regulated population?**. *Evolution,* 2009, 64:1143-1150

Chuine I and Beaubien E: **Phenology is a major determinant of tree species range**. *Ecology Letters* 2001, 4:500—510

Chuine I and Cour P: **Climatic determinants of budburst seasonality in four temperate-zone tree species**. *New Phytol.* 1999*,* 143*:*339-349

Conover DO and Schultz ET: **Phenotypic similarity and the evolutionary significance of counter-gradient variation**. *Tree* 1995*,* 10*:*248-252

Conover DO Duffy TA, Hice LA:**The covariance between genetic and environmental influences across ecological gradients: reassessing the evolutionary significance of counter-gradient and co-gradient variation**. *Ann N Y Acad Sci* 2009*,* 1168*:*100-129

Davis MB and Shaw RG*:* **Range shifts and adaptive response to quaternary climate change**. *Science* 2001*, 292:673-679*

Derory J: **Génomique et diversité du débourrement chez les chênes**. *PhD Thesis Université de Bordeaux* 2005

Devaux C, Lande R: **Incipient allochronic speciation due to non-selective assortative mating by owering time, mutation and genetic drift**. *Proc Biol Sci* 2008, 275(1652):2723

Donnelly A, Caffarra A, Kelleher CT, O'Neill BF and others: **Surviving in a warmer world: environmental and genetic responses**. *Clim Res* 2012, 53:245-262

Ducousso A, Louvet JM, Jarret P and Kremer A: **Geographic variations of sessile oaks in French provenance tests**. In Proceedings of the Joint Meeting of IUFRO working groups Genetic of oaks and Improvement and Silviculture of oaks, ed. R. Rogers, A. Ducousso and A. Kanazashi, FFPRI (Forestry and Forest Products Research Institute) *Scientific Meeting Report 3,* Tsukuba, Japan*,* 2005*, pp 128-138*

Elith J, Kearney M, Phillips S: **The art of modelling range-shifting species**. *Methods in Ecology and Evolution* 2010, 1: 330–342

Fisher RA: **The correlation between relatives on the supposition of Mendelian inheritance**. *Trans. Roz. Soc. Edinb.* 1918, 52:339-433

(FAO) Food and Agriculture Organization of the United Nations: Forest cover, **http://www.fao.org/forestry/28808/en**/, 2012/10/23

Fox GA: **Assortative mating and plant phenology: evolutionnary and pratical consequences**. *Evolutionnary Ecology Research 2003, 5:*1-18

Franjic J, Sever K, Bogdan S, Skvorc Z, Krstonosic D, Aleskovic I: **Phenological asynchronization as a restrictive factor of efficient pollination in clonal seed orchards of Pedunculate Oak (Quercus robur L.)**. *Croat. j. for. eng.* ***2011****, 31:*141-156

García-Ramos G and Kirkpatrick M: **Genetic models of adaptation and gene flow in peripheral populations**. *Evolution* 1997, 51:21-28

Gianola D: **Assortative mating and the genetic correlation**.*Theor Appl Genet* 1982*, 62:* 225-231

Gomulkiewicz R, Houle D: **Demographic and genetic constraints on evolution**. *Am. Nat*. 174:E218–E229

Hoffmann AA, Sgrö C M: **Climate change and evolutionary adaptation**. *Nature* 2012, 470:479-485

Holliday JA, Suktan H, Aitken SN: **Divergent selection and heterogeneous migration rates across the range of Sitka spruce (Picea sitchensis)**. *Proceedings of the Royal Society Series B* 2012, 279: 1675-1683.

Jensen JS, Hansen JK: **Geographical variation in phenology of *Quercus petraea* (matt) and *Quercus robur L*. Oak grown in a greenhouse**. *Scandinavian Journal of Forest Research* 2008, 23: 179-188

Jorjani H, Engström G, Strandberg E, Liljedahl LE: **Genetic studies of assortative mating - a simulation study. III. Assortative mating in selected populations**. *Acta Agric. Scand.* 1997*, 47:* 129-137

Keenan T., Maria Serra J., Lloret F., Ninyerola M., Sabate S: **Predicting the future of forests in the Mediterranean under climate change, with niche- and process-based models: CO2 matters!**. *Global Change Biology* 2010, 17:565-579

Kirkpatrick M, Barton NH: **Evolution of a species range**. *Am. Nat. 1997.* 150:1–23.

Kirkpatrick M: **Reinforcement and divergence under assortative mating**. *Proc Biol Sci, Section of Integrative Biology* 2000*, 267:*1649-1655

Kremer A, Ronce O, Robledo-Arnuncio JJ, Guillaume F, Bohrer G, Nathan R, Bridle JR, Gomulkiewicz R, Klein EK, Ritland K, Kuparinen A, Gerber S, Schueler S: **Long-distance gene flow and adaptation of forest trees to rapid climate change**. *Ecology Letters* 2012, 15: 378–392, doi: 10.1111/j.1461-0248.2012.01746.x

Le Dantec V, Dufrêne E, B Saugier: **Interannual and spatial variations of maximum leaf area index in temperate deciduous stands**. *For Ecol. Manage.* 2000, 134 : 71-81

Loustau D, Pluviaud F, Bosc A, Porté A, Berbigier P, Déqué M, Pérarnaud V: **Sub regional climate changes impacts on the water balance, carbon balance and primary productivity of maritime pine in south-west France**. In : Carnus, J.M., Dewar, R., Loustau, D., Tomé, M. and Orazio, C. (Eds.). Models for sustainable management of temperate plantation forests. *EFI proceedings*, 2001, 41: 45-57.

Lynch M, Lande R: **Evolution and extinction in response to environmental change**. In P. Kareiva, J. Kingsolver, and R. Huey (eds.) *Biotic Interactions and Global Change. Sinauer Assocs., Inc. Sunderland, MA*, 1993, pp. 234-250.

Menzel A, Sparks TH, Estrella N, Koch E, Aasa A, Ahas R, Alm-Kubler K, Bissolli P: **European phenological response to climate change matches the warming pattern**. *Global Change Biology* 2006*,* 12:1969-1976

Morin X, Roy J, Sonié L, Chuine I: **Changes in leaf phenology of three European oak species in response to experimental climate change**. *New Phytologist* 2010*,* 186*:*900-910

Morin X, Thuiller W: **Comparing niche- and process-based models to reduce prediction uncertainty in species range shifts under climate change**. Ecology 2009, 90(5):1301:1313

Nathan R, Schurr FM, Spiegel O, Steinitz O, Trakhtenbrot A, Tsoar A: **Mechanisms of long-distance seed dispersal**. *Trends Ecol. Evol.* 2008, 23:638-647

Nordli O, Wielgolaski F, Bakken A, Hjeltnes S, Mage F, Sivle A, Skre O: **Regional trends for bud burst and flowering of woody plants in Norway as related to climate change**. *International Journal of Biometeorology* 2008*, 52:*625-639

Parmesan C: **Ecological and Evolutionary Responses to Recent Climate Change**. *Annual Review of Ecology, Evolution, and Systematics* 2006, 37: 637-669

Rutishauser T, Stöckli R, Harte J, Kueppers L: **Climate change: Flowering in the greenhouse.** *Nature* 2012, 485:448-449

Savolainen O, Pyhäjärvi T, Knürr T: **Gene flow and local adaptation in trees**. *Annu Rev Ecol Evol Syst* 2007, 38:595-619

Schueler S, Schlünzen K, Scholz F: **Viability and sensitivity of oak pollen and its implications for pollen-mediated gene flow.** *Trends Ecol. Evol.* 2005*,* 19*:*154-161

Soberón J and Nakamura M: **Niches and distributional areas: Concepts, methods and assumptions.** Proceedings of the National Academy of Sciences USA. 2009, 106(suppl. 2):19644-19650

Thuiller, W: **BIOMOD: Optimising predictions of species distributions and projecting potential future shifts under global change**. *Global Change Biology* 2003 9:1353-1362

Vitasse Y., François C, Delpierre N, Dufrêne E, Kremer A, Chuine I, Delzon S: **Assessing the effects of climate change on the phenology of European temperate trees**. *Agricultural and Forest Meteorology* **2011***, 151(7):*969-980

Wolkovich EM, Cook BI, Allen JM, Crimmins TM, Betancourt JL, Travers SE, Pau S, Regetz J, Davies TJ, Kraft NJB, Ault TR, Bolmgren K, Mazer SJ, McCabe GJ, McGill BJ, Parmesan C, Salamin N, Schwartz MD, Cleland EE: **Warming experiments underpredict plant phenological responses to climate change**. *Nature* 2012, 485:494-497, doi:10.1038/nature11014

Wright S: **Systems of mating. III. Assortative mating based on somatic resemblance**. *Genetics* 1921, 6:144-161

Chapter 1

Forward simulation software for the evolution of adaptive traits in multiple populations

Jean-Paul Soularue* and Antoine Kremer
INRA, UMR 1202 BIOGECO, Cestas F-33610, France

Univ. Bordeaux, BIOGECO, UMR 1202, Talence F-33400, France

***Corresponding author**: jp.soularue@gmail.com, phone: +33 5 40 00 36 13, fax: +33 5 40 00 36 57

Running title: Modelling in evolutionary genetics

Abstract

Understanding the effects of interactions among evolutionary processes is crucial when trying to predict accurately the possible evolutionary fates of species facing environmental changes. To this end, simulation software is increasingly used to address evolutionary questions too complex to be solved analytically. Though the range of simulation programs available is continuously growing up, few of them really allow simulation of the evolution of quantitative phenotypes facing selection pressures within a connected metapopulation. We propose in this article a technical review of 6 simulation programs specialized in this area. Starting from a classification of features usually necessary to evolutionary biologists and ecologists aiming at realizing simulation studies, we first examined here the models and the built-in features provided by each package. We then compared their behaviors and their efficiency by simulating a simple common evolutionary scenario with each package. Interestingly, the sets of evolutionary models proposed by the programs show few overlaps. We observed also strong differences in terms of sensitivity, efficiency and flexibility among the simulators. These characteristics have to be carefully considered, prior to plan any kind of simulation work, in order to choose the most relevant tool. Overall, this review might be a resource helping in identifying future developments to be done regarding simulation software in evolutionary biology.

Keywords: modelling, selection, adaptation, migration, quantitative genetics.

Introduction

Large structured populations disseminated across heterogeneous landscapes constitute complex systems often studied by evolutionary biologists through the quantitative genetics paradigm. Disentangling the interplay among the evolutionary processes that shape the complex patterns of genetic variability observed is a key goal when trying to understand the evolutionary dynamics. However, despite the strong mathematical foundations of evolutionary theory and the development of advanced statistical methods, this understanding is far from being achieved and formulating evolutionary predictions under realistic scenarios remains a challenge. Of course, numerous analytical models formalizing the evolution of quantitative phenotypes in response to selection

since the earlier works of Fisher, Wright and Haldane have addressed a range of key issues *[Rice, 2004],* but most of these models describe simplified systems: single or limited sub-population number, simple genetic architectures, simplified migration models, limited population sizes, stereotyped demographic processes, no influence of landscapes characteristics. In addition to the stringent and simplifying assumptions usually inherent to pure analytical models, theoretical inferences are often specific to the issues addressed and can hardly be directly extended to many other situations.

The rapid increase in the computing power during the last three decades was followed by the development of simulation programs that are taking an increasing importance within the biologists community. When they are developed upon desired specifications, simulation packages easily provide varying numerical abstractions of complex systems that exhibit two interesting properties. First of all, they are integrative: they allow to model and implement simultaneously several processes each one describing either entities or interactions among entities constituting the systems. Generally, they include stochasticity, which is inherent to the biological systems considered and increases the range of approaches. Second, they are highly *customizable* and allow to explore wide ranges of settings, which contrast strongly with the unavoidable restrictive assumptions of purely mathematical models. As a result, simulators greatly facilitate multiple hypothesis testing and might yield significant insights on distinct levels of the systems studied, either at a global scale or at the level of specific entities.

Nonetheless, even if programming languages help in describing systems with less assumptions or more intuitively than mathematics, thus allowing to explore situations too complex to be solved analytically, the design of software is time consuming and requires technical competence. There are also numerous cases in which mathematical models give faster and better results than simulation software, especially in the case of specific studies involving limited numbers of evolutionary processes or when general assumptions or representations are sufficient. In addition, though more powerful and rapid CPU, memory and hard-drives are now available at reduced cost, numerous simulations come against the limits of the computers. Noticeably, stochastic processes particularly useful in the simulations of phenomena resulting from complex relationships among numerous processes, require often extensive CPU-time. For these reasons, numerical models should be seen as tools complementing pure mathematical models. Thereby, an increasing number of theoretical studies combine now the two approaches *[Le Corre and Kremer, 2003][Austerlitz and Garnier-Géré, 2003][Yeaman and Whitlock, 2011]*, simulations are tailored to mathematical models in order to validate them or draw wider and more accurate conclusions. Numerical models can as well greatly help in the analysis and the exploitation of the huge amount of genetic raw data now accumulating at whole population scale *[Epperson et al., 2010]*, see *[Excoffier and Heckel, 2006]* for a review.

In evolutionary biology, two main categories of simulation programs have emerged: backward simulators and forward simulators *[Carvajal-Rodriguez, 2010][Hoban et al., 2012]*. Backward simulators consider whole lineages and adopt the coalescent approach: starting from each gene, they search for the most recent common ancestor and reconstruct gene trees. Models of mutation, selection and recombination can then be applied. Forward software focus rather on individuals and performs predictive simulations by assembling evolutionary mechanisms and starting from complex initial configurations. In theory, this kind of approach allows to stick to realistic configurations since all kinds of evolutionary processes can be included and simulated. In that sense, forward simulators are more powerful and flexible than backward simulators under complex ecological scenarios. Conversely, because backward simulators are less CPU-time and memory demanding, they might be an interesting choice for the simulation of less realistic scenarios involving less entities.

Several recent review papers have already been written, either helping the user in choosing the most suitable tool *[Hoban et al., 2012]*, detailing computer engineering frameworks *[Carjaval-Rodriguez, 2010],* or discussing the utility of simulations in landscape genetics *[Epperson et al., 2010]*. As these reviews are rather general and consider all kinds of simulators, we are primarily

interested here in computer programs that meet expectations of evolutionary biologists and practitioners in charge of conservation or management of natural populations. We therefore focused on individual-based and forward-time software that simulates the evolution of quantitative phenotypes within multiple populations exchanging migrants across patchy heterogeneous landscapes. We propose in this paper a thorough technical and comparative analysis of the packages available: general presentation and technical requirements, accurate description of the main features and output proposed. We further evaluate their tractability and their efficiency within a given evolutionary scenario that was implemented with each simulator. Beside proposing precise insights on the features of the simulators to potential users, this work might be a resource for evolutionary biologists who intend to develop new simulating software or upgrade existing packages. We further hope that the comparative analysis would help practitioners interested by predictions for specific case studies.

Packages overview

Packages introduction

We found six simulators fitting our requirements. (1) *Forsim [Lambert et al., 2008]* was designed by Brian Lambert within the department of anthropology of Penn State University, it aims to simulate the evolution of complex traits and their correlations to some diseases. It allows to easily set up complex genetic architectures of multiple interacting traits under selection within several populations. The simulations always start from a single founder population whose migrants progressively establish new populations during the simulation process. (2) *Metapop* was originally created by Nathalie Machon and Didier Baradat at the *University of Orsay* to monitor the evolution of neutral nuclear and organelle markers. The package was extended to quantitative traits by Le Corre and Kremer (2003) *[Le Corre and Kremer, 2003]*. Though it has been used for various purposes *[Le Corre et al., 1997][Austerlitz and Garnier-Géré, 2003][Le Corre and Kremer, 2003] [Machon et al., 2003][Soularue and Kremer, 2012]*, no publication focused on this tool has been yet produced. In what follows, we will only mention the features introduced in the original version of the program but it should be noted that it has been greatly expanded by different users. A new official version strongly upgraded is planned to be released within the next months. (3) *Nemo [Guillaume and Rougemont, 2006]* simulates life-history and phenotypic traits within structured populations. It provides flexible life cycle events and population genetic models, and allows to consider multiple correlated quantitative traits. (4) *Quantinemo [Neuenschwander et al., 2008]* simulates the effects of evolutionary processes on complex architectures of quantitative traits involving epistasis. Those simulations include several populations exchanging migrants. (5) *Simupop [Peng and Kimmel, 2005]* is a general purpose simulation environment composed of numerous libraries that can be exploited by users via python scripts to simulate any kind of evolutionary scenario. Though its main focus is on population genetics, it is highly flexible and grants numerous built-in libraries and operators allowing for the simulation of simple quantitative architectures within contexts involving multiple populations. At last (6) *Aladyn [Schiffers et al., 2012]* is a program developed by Katja Schiffers and Justin Travis. It simulates the evolution of quantitative characters affected simultaneously by climatic and other environmental conditions varying continuously over the landscape. Of course, other existing packages can simulate the response of quantitative traits under selection, however they are often limited to single isolated populations (Capsis, ForwSim). On the other hand, many simulators accommodating multiple populations focus on population genetics of genetic markers and do not handle genetic architectures of quantitative traits (KernelPop, Cdpops, Vortex), see *[Hoban et al., 2012]* and Bo Peng's genetic simulation resources (http://popmodels.cancercontrol.cancer.gov/gsr/) for reviews of the features proposed by the best-known simulation programs.

Overall technical aspects
The 6 packages we selected were written in fast low-level languages; C++ was used for all of them, except Metapop coded in C. Executables of these programs as well as their source codes are available under the GNU General Public License via dedicated websites or upon requests (table 1). Depending on the targeted operating system, the installation process may require the compilation of the source files via the *gcc* or *g++* compilers and makefiles provided, or can be directly performed from an easy built-in installer. The *Simupop* platform requires as well the Python scripting language to be installed on the targeted system. Because they are constantly upgraded by communities of users, the versions of the programs used within published scientific works might differ strongly and include numerous features not available within the original versions available for download. The study we propose here is based on the versions *officially* released.
All of the six programs work from the command line. Starting configurations as well as simulation options can be initialized from structured input files, or in the case of *Simupop*, from python scripts. The structure and the size of the input files vary among the packages proposed. Mostly, the programs parse the input files that should fit a pre-established syntax and structure. Once the parsing of the configuration file has been done, the modules activated simulate the scenario defined over a given number of generations and build output files that consist either in formatted numerical values or graphical outputs. In the case of *Simupop*, users have to write full python scripts describing the configurations to simulate and computing the needed output.
Peer-reviewed publications as well as dedicated websites, user-guides, tutorial and sometimes forums are of great help during the phase of familiarization with these tools. *Metapop*, available on request, is still not exhaustively documented. Some information is available in studies exploiting this platform *[Austerlitz and Garnier-Géré, 2003][Le Corre and Kremer, 2003][Machon et al., 2003]* and on the Evoltree website (*http://www.evoltree.eu/index.php/modelling-platform/models/5272?task_view*). Supplementary Table 1 gives general information about the programs.

Phenotypes

Beyond the technical OS compatibilities, output produced and other general characteristics of the frameworks, the main criteria that should drive the choice of a simulation package are both the nature of the built-in models provided and the flexibility offered. In this part, we examine and compare in detail the features and models proposed by each simulator useful when modelling and simulating the evolution of quantitative characters within subdivided populations across patchy landscapes. Within the kind of simulators we are interested in, individuals are basically assimilated to their phenotypes and the major starting point of the simulations is the relationship between the phenotypic values of the individuals, the underlying genetic architecture and the environmental influence. Since *Simupop* is rather a set of Python libraries, we will only present here the functionalities and the models that can be directly implemented from those libraries with relatively few development efforts. However, theoretically this framework allows to describe any kind of situation.

Genetic architecture
The simplest quantitative model defines the genetic value *G* of an individual simply as the sum of the additive effects of the alleles presents at the *n* loci contributing to the variation of the character(s) studied. In the diploid case we have :

$$G = \sum_{l=0}^{n} (\alpha_1 + \alpha_2)_l \qquad (1)$$

where α_1 and α_2 are the allelic effects of the two alleles present at each locus *l*. In this model, the number of α values per locus is related with the degree of ploidy, thus, in the case of haploid organism it becomes:

$$G=\sum_{l=0}^{n}(\alpha)_l$$

In what follows, we will mainly illustrate the diploid case, however haploid (*Metapop*, *Simupop*) and polyploid populations (*Simupop*) can sometimes also be simulated (see table 1).

Table 1. Genetic architecture.

	Trait nb	Pleiotropy	Epistasis	Dominance	Genetic map	Allele definition	Ploidy	Heritability
Aladyn	- 2	- Indirectly from full linkage	- No	- No	- No	- No but initial phenotypic variance can be set	- Diploid	- Fixed to 1
Forsim	- n	- User defined GxG interactions	- User-defined GxG interactions	- User-defined GxG interactions	- Yes* in bp	- Allele number defined indirectly from the number of SNP per site (2 or 4) and the length of genes - No allelic effect can be assigned explicitly	- Diploid	- Broad sense - Narrow sense
Metapop	- 1	- No	$E_p=\sum_{l=0}^{n}\sum_{k=0,k\neq l}^{n} ep_{lk}$ $G=\sum_{l=0}^{n}(\alpha_1+\alpha_2)_l+E_p$	- *d* sampled from Normal law at each locus $G=\sum_{l=0}^{n}(\alpha_1+\alpha_2^{'}+d)_l$	- No	- Number for each locus - Additive effects following Normal distribution	- Nuclear loci: diploid - Cytoplasmic loci: haploid	- Broad sense - Narrow sense
Nemo	- 2	- Indirectly from full linkage	- No	- No	- Yes in cM	- No	- Diploid	- No direct option
Quantinemo	- n	- Full: when simulated, all loci contribute to the variations of all traits	E_p predefined or drawn randomly for each genotype *[11', 22', .., ii']*: $G=\sum_{l=0}^{n}(\alpha_1+\alpha_2)_l+E_p$	At each locus *l*: $d=k\lvert\alpha_1-\alpha_{2l}\rvert$ where k is user-defined $G=\sum_{l=0}^{n}(\alpha_1+\alpha_2^{'}+k\lvert\alpha_1-\alpha_2\rvert)_l$	- Yes in cM	- Number for each locus - Additive effects following Normal distribution or set explicitly	- Diploid	- Narrow sense - Broad sense - Constant
Simupop	- n	- Yes, with additional developments	- Yes, with additional developments	- Yes, with additional developments	- Yes in bp or cM	- Number for each locus - Allelic effects can be assigned easily with additional developments	- Haploid, - Diploid, - Haplo -diploid - Triploid - Tetraploid	- No direct option

Symbols and abbreviations. [11', 22', .., ii']: full genotype where aa' are the two alleles at each locus. α_l : additive effect of the allele i at a given locus. G: genetic value. E_p : overall epistatic effect. ep_{lk} : effect of epistatic interaction between the loci l and k. d : dominance effect at a given locus.*: mandatory.

Basically, the model (1) can be implemented from all the simulators examined here. Along with this model, indicating a number of QTLs is a prerequisite handled by all simulators while the definition of initial and maximal numbers of alleles at each QTL is only proposed by *Metapop, Quantinemo and Simupop.* When this feature is available, the initial additive contribution of the alleles may be set explicitly (*Quantinemo*, *Simupop*) or may follow statistic distributions (*Metapop*, *QuantiNemo*, *Simupop)* (table 1). When this feature is not available, the genetic variability among individuals

relies upon mutational effects also drawn randomly according to the mutational model chosen (*Aladyn*, *Forsim*, *Nemo*). Though the framework *Simupop* does not originally allow to directly specify the additive effects of alleles, it grants operators making the programming of initializing functions pretty fast and easy.
The initial proportion of the phenotypic variation that owes to genetic factors can be user-defined by providing narrow-sense or broad sense heritability values (*Forsim, Metapop* and *Quantinemo*). The heritability of the trait evolves then along the simulation process, owing to the change in genetic variance that is likely to occur in the scenarios simulated. A constant heritability combined with a variable environmental variance can alternatively be simulated in *Quantinemo*. Surprisingly, *Simupop* and *Nemo* do not provide any option to explicitly set the heritability of characters. In these cases, the initial heritability is implicitly inferred from the additive variance of the mutational effects and the environmental variance indicated (*Nemo, Simupop*) or from the initial additive variance of the allelic effects and the variance of the environmental influence (*Simupop*). In *Aladyn*, environmental effects on phenotypic values are neglected and heritability is fixed to 1.
The simple genetic additive model (1) can be greatly expanded from the built-in features provided by each simulator. First of all, interactions effects between alleles at contributing loci can be simulated in several ways. Looking at epistasis, users can specify within *Forsim* algebraic functions using standard mathematical functions (addition, division, multiplication, subtraction, absolute value, exponential) in order to describe *GxG* interactions with an interesting degree of flexibility. For example, *Forsim* allows to easily define the following relationship among contributing loci:

$$G = 2G_1^2 + \frac{G_2}{G_3 + G_4}$$

where G_i represents the additive effect at the locus *i*. On the other hand, *Metapop* and *Quantinemo* simulate epistatic effects through an additional additive epistatic component E_p, resulting in

$$G = \sum_{l=0}^{n} (\alpha_1 + \alpha_2)_l + E_p$$

Metapop simulates epistatic interactions of order 2 from random epistatic effects assigned to each pairwise of loci. In that case we have

$$E_p = \sum_{l=0}^{n} \sum_{k=0, k \neq l}^{n} ep_{lk}$$

where ep_{lk} sums the epistatic effects of all possible allelic interactions between the loci *l* and *k,* and where each epistatic effect is drawn at random from a Normal distribution scaled by the number of locus and an initial variance. Interestingly, *Quantinemo* simulates epistatic interactions of order *n* from global epistatic effects for each multilocus genotype *[11', 22', ..., ii']* assigned explicitly or drawn randomly from a Normal distribution $N(0, \sigma^2)$ where σ^2 is initially set by the user.
Dominance effects at each locus are only explicitly simulated by *Metapop* and *Quantinemo*. Both add a dominance effect to the global additive value computed. *Metapop* assigns a random dominance effect for each combination of the two alleles at each given locus, implementing

$$G = \sum_{l=0}^{n} (\alpha_1 + \alpha_2' + d)_l$$

Where d follows a Normal law scaled both by the variance of dominance effects indicated and the number of loci. The dominance component computed by *Quantinemo* depends on a coefficient *k*

$$G=\sum_{l=0}^{n}(\alpha_1+\alpha_2^{'}+k|\alpha_1-\alpha_2|)_l$$

where *k=-1* designates *allele 1* (of locus *l*) as dominant, k=1 indicates that *allele 2* (of locus *l*) is dominant, *k=0* cancels any dominance relationship between *allele 1* and *allele 2* and $|k| > 1$ stands for under-dominance or over-dominance depending on the sign of *k*. Users have here also the choice either to assign explicitly a dominance effect for each allele pair within a separate file, or to define a Normal distribution used to draw randomly the dominance effects.
The general quantitative model of genetic effects can be extended to the case of phenotypes composed of multiple traits (all simulators excluding *Metapop*). In this case, it might be useful to describe and monitor the effects of pleiotropic interactions among loci. *Forsim* accommodates pleiotropic interactions among loci based on simple algebraic definitions similar to the *GxG* interactions presented above. The simple example given in *[Lambert et al., 2008]* illustrates the kind of pleiotropic interactions that can be defined. Interestingly, when a single trait is explicitly set up, *Forsim* provides the alternative to simulate the global influence of the rest of the genome with a random polygenic background effect affecting the expression of genotypes. *Aladyn* and *Quantinemo* do not explicitly simulate pleiotropic interactions among the loci involved in the variations of the characters simulated, however full linkage among loci may induce a similar effect. Finally, full pleiotropy is considered in *Nemo:* all the loci contribute equally to the variation of the two traits simulated (table 1).
Beyond the general quantitative model of genetic effects, some studies might require the specification of underlying physical genetic architectures. Physical position of loci can be indicated to account for recombination events and linkage relationships in different ways. Such map description may range from a simple indication of a chromosome number associated with global recombination rates among loci (all simulators excluding *Forsim*) to the full description of genetic or physical maps explicitly positioning each locus along chromosomes either in basepair (*Forsim, Simupop*) or in centiMorgan (*Nemo*, *Quantinemo*, *Simupop*). A detailed genetic map description is mandatory within *Forsim* prior to launch any simulation, while it is optional within *Nemo*, *Quantinemo* and *Simupop*, and not handled by *Metapop*. Output allowing to monitor the evolution of the physical architecture is sometimes granted (see table 1). Interestingly, the genetic architectures can be enriched with neutral markers in addition to QTLs (*Nemo, Quantinemo, Simupop, Metapop*) and chromosomes may exhibit distinct inheritance patterns in addition to the standard nuclear autosomal type: cytoplasmic (*Metapop*), mitochondrial, sexual or user-defined types (*Simupop*).

Environmental effects
The overall phenotypic value of a given individual results from the joint contribution of genetic and environmental effects. Environmental effects are also understood as the source of phenotypic plasticity in the broad ecological sense (table 2). Apart from *Aladyn* that ignores environmental effects on phenotypes, all of the simulators basically implement the well-known and widespread model that expresses a micro-environmental influence ϵ drawn independently for each individual from a Gaussian distribution, and added to their genotypic value:

$$Z_{ij}=G_{ij}+\epsilon_{ij}$$

This expression of phenotypic values fits particularly with situations in which there is no need to explicitly associate environmental variations with variations in space. In *Metapop* and *Forsim* the variance of the distribution is set to 1, while other packages allow users to set that variance either by specifying directly a value (*Nemo, Quantinemo*) or from the initial heritability (*Quantinemo*). Though no built-in options is proposed within *Simupop* regarding this feature, it is very easy to include any kind of random effect from the quantitative operators available or from the Python libraries.

Table 2. Environmental contribution.

	Genotypic plasticity $Z_{ij}=G_{ij}+G'_{ij}\epsilon_j$	Macro-environmental effects $Z_{ij}=G_{ij}+\epsilon_j$	Micro-environmental effects $Z_{ij}=G_{ij}+\epsilon_{ij}$
Aladyn	- No	- No	- No
Forsim	- Only at family scale, using algebraic functions	- Only at family scale. A random variate is drawn randomly and assigned to each individual of a sibship	- Yes, with ϵ following $N(0,1)$
Metapop	- No	- No	- Yes, with ϵ following $N(0,1)$
Nemo	- No	- No	- Yes, with ϵ following $N(0,\sigma_\epsilon^2)$
Quantinemo	- No	- Almost: σ_ε^2 assigned to each patch (see third column)	- Yes, with ϵ following $N(0,\sigma_\epsilon^2)$ - σ_ε^2 variable, calculated each generation from h² or H² constants - σ_ε^2 constant - Individuals affected by their natal or current environment
Simupop	- Yes, with additional developments	- Yes, with additional developments	- Yes with ϵ following any kind of distribution

Symbols and abbreviations. Z_{ij} : phenotypic value of individual i from population j. G_{ij} : genetic value at the non-plastic loci of individual i from population j. G'_{ij} : genetic value at the plastic loci of individual i from population j. ϵ_j : macro-environmental effect in population j. ϵ_{ij} : micro-environmental effect for individual i in population j. N: normal distribution.separately for each patch.

Two of the simulators (*Quantinemo* and *Simupop*) grant the possibility to assign environmental effects to populations or patches rather than individuals, according to the variations of environmental conditions occurring across the landscape simulated, implementing the following model:

$$Z_{ij}=G_{ij}+\epsilon_j$$

where Z_{ij} stands for the phenotypic values of individual *i* from population *j*, G_{ij} for its genetic value, ϵ_j for the environmental conditions assigned to the population *j*. Hence, ϵ_j stands for the macro-environmental influence that affects uniformly all individuals of a given population. This second class of model is close to the previous one, however it defines the variate epsilon at the population level, which is likely to be useful when the landscapes simulated exhibit significant environmental variation among patches or niches. This latter feature is crucial within the kind of simulations we are interested in, as spatial patterns of environmental variations are omnipresent *in natura [Gienapp et al, 2008][Aitken et al, 2008]*. However, among the simulators considered, the explicit assignment of macro-environmental effects to patches is only supported by *Simupop* that provides the necessary libraries for explicit or random initializations. *Quantinemo* only enables the user to set the variance of the distribution in which environmental values are drawn at random. *Forsim* gets close to this second model by making possible to account for a family environmental effect that can be added to the individual random deviation drawn for each sibship. However, family structures differ from the population structure defined by the user and it is impossible to define any relation among these family effects and the populations.
The last feature related to environmental variations is genotypic plasticity. A simple model inspired from *[Lande, 2009]* could illustrate that feature:

$$Z_{ij} = G_{ij} + G'_{ij}\epsilon_j$$

In this latter equation, the interaction among the plastic loci G'_{ij} and the macro-environmental conditions undergone by population *j* relies on the second additive component $G'_{ij}\epsilon_j$. It can be simulated by *Forsim* from the algebraic formalism provided that enables to combine in a simple way environmental and genetic effects. However, as mentioned earlier, this tool does not allow to account for any structured environmental variation among patches. Users can also define *GxE* interactions within *Simupop*, with environmental effects assigned to individuals or populations. Since no dedicated library is provided, it might require some time to be implemented.
Finally, the example of models including genetic and environmental effects we used as illustrations in this part are excessively simple and generic. When they are implementable within the simulators, they can be merged with others, resulting in more complex models fitting the configurations to describe.

Evolutionary processes

Once the phenotypes of the individuals constituting populations have been defined, the second step consists in setting up relevant evolutionary processes regarding the scenarios to simulate. We examine in this part the genetic and demographic processes that directly alter the genetic values of individuals and the populations' composition along simulations.

Natural selection
Natural selection may be defined from built-in predefined models (*Aladyn*, *Forsim, Metapop, Nemo, Quantinemo*) or from user-defined functions (*Forsim, Simupop*). Both approaches result in fitness values affecting the survival of individuals (*Aladyn*, *Nemo, Simupop*) or their ability to reproduce (*Aladyn*, *Nemo, Metapop, Quantinemo, Forsim, Simupop*). In addition, *Nemo* and *Quantinemo* offer to compute fitness values according to 3 strategies involving distinct scales: (1) each patch can be considered separately, (2) fitness values can be computed at the meta-population level, or (3) fitness values can be assigned to individuals relatively to the mean fitness of the meta-population.
The widely used Gaussian stabilizing selection model can be easily implemented in the 6 simulators from equivalents of Turelli's model *[Turelli, 1984]*:

$$W = e\left(\frac{-(Z - Z_{opt})^2}{2\omega^2}\right)$$

where W_t is the fitness value assigned at a given generation, ω^2 is the intensity of selection, *Z* the trait's value and Z_{opt}, the optimal trait value. When this model is not available, stabilizing selection can be specified via a mean and a standard deviation of a fitness function that can be gaussian (*Forsim, Simupop*). Some simulators offer to describe more complex selection schemes; thereby, *Nemo* and *Quantinemo* provide stabilizing selection models applicable on multiple traits. In the case of *Nemo*, the fitness values are then computed according to

$$W(z) = \exp\left[-\frac{1}{2}(Z - V_{Z_{opt}})^T \omega^{-1}(Z - V_{Z_{opt}})\right]$$

Where $V_{Z_{opt}}$ and ω^{-1} constitute a vector of optima assigned to each trait *Z* and an associated matrix of selection intensities, respectively. On the other hand, within *Quantinemo* the fitness of individuals is the product of the fitness of each trait composing the phenotype. In the case of a phenotype being composed of *n* traits:

$$W = \prod_i^n W_i$$

Similar models yielding fitness values for multivariate phenotypes can be simulated by *Simupop*, however the programming of Python functions slightly more evolved than those mentioned up to now is required. At last, *Aladyn*, offers to mimic the simultaneous influence of two environmental factors associated to the two adaptive characters simulated, respectively. The resulting model implemented is close to Turelli's model:

$$W = e^{\left(\frac{-(Z_{C,E} - Z_{opt_{C,E}})^2}{2\omega_{C,E}^2}\right)}$$

In this last model, *C* refers to the climatic phenotype of the individuals simulated while *E* refers to their environmental phenotype. All of the software packages simulating discrete landscapes (all excluding *Aladyn*) enable to assign phenotypic optima and selection intensities separately to each patch defined. This yields the possibility to simulate divergent selection among populations according to varying spatial patterns, which is of primary importance within the kind of simulation we are interested in here. Such possibility is also possible in *Aladyn* which mimics continuous landscapes. In this simulator, users are asked to define a number and a size of patches dividing latitudinally and longitudinally the entire landscape as well as patterns of variations in the phenotypic optima to be simulated over the landscape. The resulting variations simulated over the landscape can be clinal, fractal or null. Of course, owing to the continuity of the landscapes in *Aladyn*, the values of the phenotypic optima will vary more or less within each patch, depending on the patch size defined.
Besides stabilizing and divergent selection, several other selection models are available and can be set independently for each population when it is relevant (table 3). Directional selection is handled, either from asymmetric cut-offs specified on user-defined fitness functions (*Forsim*) or from a generalized logistic curve *[Richards, 1959]* (*Quantinemo*):

$$W = min + \frac{max - min}{\left(1 + s * e^{r(P_{r_{max}} - P)}\right)^{1/s}}$$

where *min* and *max* are the lower and upper asymptotes, *r* stands for the growth rate, $P_{r_{max}}$ for the phenotypic value with the highest slope and *s* for the degree of symmetry of the curve. *Nemo* offers to assign fitness equaling the phenotypic value of the adaptive trait, selection on deleterious mutations traits, or the "Fix" model that assigns fitness values on the basis of the pedigree of individuals:

$$W_F = W_0 * e^{-F\lambda}$$

In this latter model, W_F is the fitness of individuals with the inbreeding coefficient *F*, W_0 is the base fitness of the population and λ is the number of lethal equivalents in the population. Finally, *Forsim* provides the necessary means to easily implement the truncation selection model and *Simupop* yields operators making the development of fitness functions easy.

Mutation
Three main mutations models are simulated by the simulation programs considered here: the Random Mutation model -RMM-, the K-allele mutation model -KAM- and the Increment mutation model -IMM-. Other models like the deleterious mutation model (*Nemo*), the Diallelic Mutations Model (DAM) (*Nemo*, *Simupop*), or model allowing for mutations resulting in the modification of the locus number (*Simupop*) are also available (table 3). On top of the mutation models that can be implemented, mutation rates can be assigned either globally (all of the packages), independently to

each locus (*Quantinemo, Forsim, Simupop*), to each allele (*Simupop*) or according to the sex of individuals (*Forsim*). Interestingly, mutations rates can be sometimes changed easily during the simulation process (*Quantinemo, Forsim, Simupop*). Depending on the simulators and the models of mutation simulated, the mutational effects either follow statistical distributions (all packages), or they are specified by users (*Metapop*, *Nemo*, *Quantinemo*). *Simupop* clearly provides the largest range of mutation models with an impressive number of options and allow to combine in different ways the numerous features available, however, the additive mutational effects have to be manually managed through the quantitative operators provided. Table 3 summarizes all of the built-in features regarding mutations.

Recombination

The probability of crossing-over segregating adjacent loci across the chromosomes can be established according to three ways. Either it is deduced from a genetic map expressing the genetic distances between loci within each chromosomes (*Forsim, Nemo, Quantinemo, Simupop*), either it is explicitly set (*Aladyn*, *Metapop*, *Nemo*, *Simupop*), or it can follow Poisson distributions with a mean corresponding to the length of each chromosome (*Nemo*). Recombination rates may be assigned globally (*Aladyn*, *Metapop, Nemo, Simupop*), specifically to each pairs of loci or chromosomes (*Nemo, Simupop*) or assigned separately to subpopulations (*Simupop*). Additional possibilities are also provided by *Simupop*, such as the simulation of double recombination events leading to gene conversion or the definition of several types of chromosome (autosomal, sexual, mitochondrial, customized) associated with distinct patterns of inheritance.

Demography

A starting number of populations can be defined from all of the programs excluding *Forsim* that systematically starts simulations from a single founder population. Proportion of female and male individuals can be specified (all packages excluding *Aladyn* and *Metapop*) and sometimes controlled over the landscapes and successive generations (*Simupop*). In addition, hermaphroditism is nearly systematically handled (all simulators excluding *Forsim*). Though growth and extinction of populations are naturally related to selection pressures, migration and overall demographic conditions, users may need the population growth to follow specific models or trends. When explicit growth models are provided, size of populations evolves each generation from three main variables -current size, growth rate, carrying capacity- according to deterministic (*Forsim*, *Quantinemo*) or stochastic (*Metapop*, *Quantinemo*) strategies. Alternatively, fecundity rates or overall offspring population size can be specified (*Aladyn*, *Forsim*, *Nemo*, *Quantinemo, Simupop*). They can be fix or they can follow user-specified distributions (see table 4). Since stochasticity is liable to make the number of individuals exceeding the carrying capacity of patches and saturate populations, the number of individuals may be post-regulated (*Aladyn*, *Nemo*, *Quantinemo*).

Some studies might require to explicitly position populations across the landscape. This feature is proposed by *Forsim* and *Simupop,* only at the population level. Alternatively, when this feature is absent, the environmental influence, the selection patterns and the migration rates are settings that implicitly position populations across the landscapes simulated.

Colonization events are supported by all programs allowing for the creation of empty patches filled by migrants from existing populations. Extinctions related to harvesting or other critical events are simulated mostly through patch-dependent extinction rates (*Metapop, Nemo, Quantinemo*) or by extinction events of populations specified at given time-points (*Forsim, Simupop*). *Nemo* yields the most interesting possibilities: the user can indeed set extinction sizes and densities, either via fixed sizes or via sizes following distributions as well as a minimal density below which too small populations wipe out. Of course, populations may wipe out when they face too stringent selection or demographic conditions.

Several mating systems can be simulated, ranging from the standard random mating to complex heterogeneous mating schemes. The versions of the packages we considered here only simulate discrete non-overlapping generations apart from *Simupop* that additionally propose aged-structured

Table 3. Evolutionary processes acting on the phenotypes.

	Selection	Mutation	Recombination
Aladyn	- Stabilizing - Divergent	- RMM	- Overall rate
Forsim	- Stabilizing - Divergent - Uniform - Directional - Neutral - Truncation - User-defined	- RMM - IMM - LM - Mutation rates can be gene or sex-specific - Effects drawn from a Gamma distribution	- Full genetic map (mandatory) + recombination rates expressed via nb basepair / cM - Can be sex-specific
Metapop	- Stabilizing - Divergent - Uniform - Neutral	- KAM - Only global mutation rates - Effects drawn from a Normal distribution	- Overall rate + number of loci
Nemo	- Stabilizing - Stabilizing multivariate - Divergent - Uniform - Neutral - Fix model (for deleterious) - Fitness relative-local, absolute, or relative-global	*Neutral markers:* - SSM - KAM *Quantitative traits:* - RMM (Normal distribution) - DAM (user-defined value) - Deleterious mutation (Constant, Exponential, Gamma or Log-normal distribution) - Pleiotropic mutation (Bivariate, Normal distribution) - Global mutation rates	- Overall rate - Rate for each chromosome - Recombination map (map distance between each locus) - If not specified, the loci are assumed to be independent
Quantinemo	- Stabilizing - Divergent - Uniform - Directional - Neutral - Heterogeneous multivariate - Combinations of distinct models are possible - Fitness values interpreted relative-local to the patch, absolute, or relative-global	*Neutral markers:* - KAM - SSM *Quantitative traits:* - RMM - IMM - Probability to mutate to certain alleles may be set explicitly	- Genetic map. If not specified, the loci are assumed to be independent
Simupop	- Fitness values assigned according to genotypes - User defined selection function - Heterogeneous models	- KAM - DAM - IMM - SSM - Nucleotide mutation model - Mixed mutation model - Mutations resulting in new loci... (non exhaustive list) - Rates can be assigned at locus, at alleles or globally - Distinct models can be set on distinct locus	- Overall rate: can be specific to pairs of loci or populations - Rates between adjacent loci - Gene conversion according to the Holliday model (double recombination event) - Specification of chromosome type: autosomal, mitochondrial, sexual and user-defined

Mutation models. RMM: Random Mutation Model, the mutational effect of the new allele is drawn randomly from a given distribution and is added to the existing allelic value. IMM: Increment Mutation Model, the mutational effect depends on the original state of the mutant allele. DAM: Di-Allelic Mutation Model, mutations effects can only increase or decrease a given value, resulting in symmetrical effects. KAM: K-Allele Mutation model assumes at most k alleles at a given locus. Each mutation replaces the existing allele by another allele from a predefined set of alleles ranging from 0 to k. SSM: given a predefined index of alleles, Single Step Mutation model replaces a mutating allele k by one of the "adjacent" allele k+1 or k-1 randomly. LM: Lethal Mutations. SMM: Stepwise Mutator Model for microsatellite markers.

mating systems. Random mating schemes can be implemented within all of the simulators with

some variations depending on the program. Selfing is proposed by all the simulators excluding *Forsim* while multiple other possibilities are provided by *Nemo*, *Quantinemo* and *Simupop* such as cloning, polygamy, monogamy that can be combined within heterogeneous mating schemes (*Simupop*) (table 4). Finally, the succession of life cycle events undergone by the populations within each generation is fixed within *Aladyn*, *Forsim* and *Metapop*, globally implementing similar basic cycles: (1) migration, (2) selection of parents, (3) matings, (4) occurrence of mutations and (5) recombinations within the offspring genotypes. Interestingly, *Nemo, Quantinemo* and *Simupop* propose some additional events and allow to modify the sequence of life cycle events simulated within each generation. As a result, they provide more flexibility, the only limit being the relevance of the life cycles defined.

Migration

Migration of individuals and gametes among populations constitutes an other crucial point of simulations. Basically, migration can be set through matrix specifying rates or proportions of migrants (zygotes) or gametes among populations. This possibility, yielded by all the simulators excluding *Aladyn*, allows to mimic any kind of dispersal model. On the other hand, *Nemo, Quantinemo* and *Simupop* provide built-in functions that automatically initialize the most well-known migration models. *Quantinemo* provides here again the most important number of models and settings, allowing for example to define how migration occurs at the border of the landscapes. Interestingly, *Nemo* enables the user to simulate the evolution of a quantitative heritable trait controlling dispersal in addition to the trait simulated. This built-in trait is coded by a single diploid locus that determine the dispersal rates of male and female individuals. *Aladyn* simulates continuous landscapes, it provides a unique dispersal model based upon log-normal and isotropic kernels. Migration rates of individuals follow then log-normal distributions whose mean distance is set by the user *[Schiffers et al. 2012]*. Table 4 gives an overall insight of the features related to migration provided by each package.

Table 4. Demography and migration.

	Population growth	Mating system	Position of populations	Migration	Life cycle events	Colonization	Extinction
Aladyn	- Fixed offspring number - Patch density regulation	- Random mating	- No, but continuous landscape and size of patches	- Log-normal isotropic dispersal kernel - BI (absorbing or torus)	- Fixed order	- Yes	- No
Forsim	- Fecundity follows Poisson distribution with user-specified mean - Maximal offspring number - Population size follows Verhulst model: $N_{t+1}=\frac{K*N_t*e^r}{K+(N_t*(e^r-1))}$ or can be constant - Simulations start from a single population	- Random mating with or without replacement - Assortative mating	- Yes - Location of offspring determined stochastically	- Migration matrix	- Fixed order	- Yes *	- Explicit extinction at a given point
Metapop	- Population size follows: $N_{t+1}=N_t+r*N_t((1-N_t)/K)+U(0,1)$	- Random mating without replacement - Selfing	- No, indirectly from the migration matrix	- Migration matrix	- Fixed order	- Yes	- Extinction rate
Nemo	- Fecundity follows Normal or Poisson distribution - No growth rate - Constant size - Fixed or random sex-ratio - Population resizing - Fusion/fission of patches - Patch density regulation	- Random mating with or without replacement - Polygamy - Monogamy - Selfing - Cloning - Heterogeneous mating schemes - Breed with backward migration - Mating males	- No, indirectly from the migration matrix	- Migration matrix - IMM - SST - DR, BI - Quantitative trait for dispersal	- Customizable	- Yes	- Extinction rate following Uniform, Poisson, Normal, Exponential or Log-Normal distribution - Minimal density below which populations wipe out
Quantinemo	- Population size computed either deterministically from: $N_{t+1}=\frac{N_tK(1+r)}{N_t(1+r)-N_t+K}$ or stochastically from: $N_{t+1}=P(\frac{N_tK(1+r)}{N_t(1+r)-N_t+K})$ from fecundity rate (fixed, P) or constant - Fixed or random sex-ratio - Population resizing - Patch density regulation	- Random mating with or without replacement - Selfing - Cloning - Polygamy - Monogamy - Heterogeneous mating scheme - Mating males	- No, indirectly from the migration matrix	- Migration matrix - IMM - SST - DR, BI	- Fixed order, but individual cycles can be left out	- Yes	- Extinction rate
Simupop	- Offspring number fixed or following Poisson, Binomial, Uniform, Geometric distribution - Fixed or random sex-ratio - Populations resizing - Fusion/fission of patches	- Random mating with or without replacement - Assortative mating - Selfing - Polygamy - Monogamy - Age-structured population - Heterogeneous mating scheme	- Yes	- Migration matrix - IMM: standard, hierarchical - SST - Global or patch-specific assignation of model	- Highly customizable	- Yes	- Explicit extinction at a given point -time

Symbols and acronyms used. N_t : size of a population at generation t. K: carrying capacity of a population. r: growth rate. P: Poisson distribution. U: Uniform distribution. IMM: Island migration model. SST: Stepping-stone model. DR: dispersal range. BI: Influence of landscapes' boundaries on migration: absorbing, reflective. *: mandatory.

Running simulations

Output

Overall, the most complete and detailed output dealing with population and quantitative genetics are provided by *Metapop*, *Nemo* and *Quantinemo*. On the other hand, *Forsim* is the only simulator offering to monitor accurately the evolution of genetic markers at the physical level by providing accurate descriptions of the history each SNP, haplotype and genes: localization, quantitative contribution, date of apparition, statistics. To this end it produces quantity of textual files and graphs facilitating analysis of the data generated at the end of each run. At last, the generic platform *Simupop* proposes interesting operators to generate additional information, but no built-in operator for the monitoring of quantitative genetic data. Dumping the genotypes in order to further restore interrupted simulations or to simulate multiple evolutionary scenarios from a single starting population is possible excepted in *Aladyn* and *Metapop*. Finally, *Nemo* and *Quantinemo* require the user to specify accurately the kinds of output to compute before launching any run. As this is done very quickly within the input files from few keywords, it allows to save memory space and CPU time and ensures a fast understanding and exploitation of the output provided. Table 5 lists output automatically computed by each simulator. However, post-processing the files of results produced by each simulator with external scripts (*e.g.* individuals genotypes in *Nemo* and *Quantinemo*) may allow to considerably extend the possibilities offered.

Generating initial state

Depending on the hypothesis to test or the scenario to simulate, the starting meta-populations simulated should often exhibit precise characteristics. The definition of the overall structure of the meta-population is the first crucial point: before choosing any simulator, it is essential to check whether the models and the features offered allow for a relevant description of the genetic, demographic and environmental components regarding the scenarios to simulate. But the initial state regarding the environmental conditions faced by the population requires particular attention since it might strongly affect the evolutionary dynamics in a given scenario. Overall, the main indicator conveying the state of a population is the distribution of the alleles carried by its individuals. Hence, as example, some studies may require to start from a population at mutation-migration-drift equilibrium, others from a population composed of homozygous individuals or from a population exhibiting allelic frequencies corresponding to real observations. Three of the simulators considered here allow to specify initial allelic frequencies: *Metapop*, *Quantinemo* and *Simupop*. Initial allelic settings described within dedicated additional files in *Metapop* and *Quantinemo*. However the files required can be large and, often, they have to be generated automatically from external scripts. Interestingly, *Simupop* provides operators for generating initial genotypes and allelic frequencies at the population level.

Otherwise, some simulators construct directly the initial allelic frequencies from general characteristics indicated by the users. *Aladyn* also proposes to start simulations from populations either adapted or not to the selective pressures faced, but only for one of the two characters simulated. *Nemo and Quantinemo* allow to assign an initial genetic value to each patch or to the whole meta-population and to start from an initial meta-population either monomorphic or polymorphic for the specified trait value. However, none of the simulators examined here provides any built-in feature generating initial meta-populations at mutation-migration-drift equilibrium despite the theoretical interest of such initial state. All packages proceed empirically to generate initial settings corresponding to equilibrium status: they start from initial entirely homozygous population and go through thousands of generations letting mutation, migration and drift shape the distributions of alleles until their frequencies reach asymptotic values, which then corresponds to the initial settings for the evolutionary scenario to be tested.

Table 5. Output.

	Genetic marker monitoring	Quantitative genetics	Population genetics	Demography	Formats	Save and restore populations	Other
Aladyn	- No	- Components of phenotypic values	- No	- Population size	- .csv	- No	- None
Forsim	- SNP tracking - Haplogenes tracking - Recombination tracking	- Components of phenotypic values - Components of phenotypic variance - Heritability - Selection monitoring	- Nucleotidic diversity - Diversity statistics - Linkage disequilibrium	- Population size - Migration monitoring	- .txt - .svg - .xml	- Yes	- Elapsed time per generations - Output of summaries - Multiple graphs
Metapop	- No	- Components of phenotypic values - Components of phenotypic variance - Heritability, QST	- F-statistics - Genetic diversity - Allelic frequencies	- Population size	- .txt	- No	- None
Nemo	- No	- Components of phenotypic values - Components of phenotypic variance - Heritability, QST - Genetic correlations - Selection monitoring	- F-statistics - Genetic diversity - Allelic frequencies	- Population size - Migration monitoring - Fecundity monitoring - Matings - Population monitoring	- .txt - FSTAT	- Yes, binary files that can be compressed - Populations can be seeded from text files as well (e.g., FSTAT files)	*Statistics on:* - Dispersal - Wolbachia infection - Deleterious mutation
Quantinemo	- No	- Components of phenotypic values - Components of phenotypic variance - Heritability, QST - Genetic correlations - Coanscestry	- F-statistics - Genetic diversity - Allelic frequencies	- Population size - Migration monitoring - Fecundity monitoring - Matings	- .txt - FSTAT	- Yes: complete, genotypes, genotypic or phenotypic values uses FSTAT and FSTAT extended formats	- Allelic values - Dominance values - Epistatic values
Simupop	- Recombination tracking	- No	- F-statistics - Genetic diversity - Allelic frequencies - Nucleotidic diversity - Linkage disequilibrium	- Population size	- .txt	-Yes from multiple file formats (FSTAT, csv, Phylip, ped, MAP, Genepop, STRUCTURE)	- Hardy-Weinberg test

Definitions. Genetic correlations: genetic correlation between the genetic values of the traits when multiple traits are simulated. Matings: information about pedigree, kinship and co-ancestry. Nucleotidic diversity: count of snps, haplotypes and genes. Selection monitoring: mean fitness values. Population monitoring: patch age since last extinction, extinction rate...

Management of simulations

Nemo, *Quantinemo* and *Simupop* enable the user to save the meta-population data on the hard drive at any moment and to easily import them back. This procedure allows to stop the construction the initial meta-population, to interrupt and resume simulations, or to perform any kind of additional process. *Nemo* and *Quantinemo* propose different kinds of dumps (genotypes or populations), saved mostly within *Fstats* files. *Simupop* provides the opportunity to perform complete dumps saved within special binary files. Interestingly, apart from *Metapop*, all of the simulators offer to change the values of some settings at any moment of the simulation process, yielding the interesting

possibility to create temporal patterns of variation. Changes may concern the intensity of selection and phenotypic optima (all excluding *Metapop*), the migration settings (all excluding *Aladyn* and *Metapop*), or the structure of the whole meta-population (all excluding *Aladyn* and *Metapop*): extinction, fusion of patches... At last, efficient management of simulation campaigns are facilitated when exhaustive log files are produced, when replicates can be launched in batch mode, when the results folders are automatically created and when post-script execution can be indicated prior to launch simulations. In this field, *Forsim*, *Nemo* and *Quantinemo* provide interesting possibilities hence limiting confusion and possible critical errors during the exploitation of the many output produced.

Comparative test

To conclude, we conducted a comparative analysis by simulating a standard evolutionary scenario involving multiple populations with each of the 6 simulators. Assuming that the tools will produce similar results, the main objectives were (1) to assess the time we needed to set up the starting configuration and (2) to evaluate the efficiency of the computing means usage. Of course, these assessments are strongly related to the features of the scenario simulated.

Scenario content
The landscape comprised 11 populations of 1000 individuals. Each phenotype consisted in a single quantitative trait controlled by 10 QTLs. A local random environmental deviation drawn at random from the Normal distribution $N(0,1)$ was added to the phenotypic value of the trait. The QTLs were located on separate chromosomes and segregated independently, genes mutated according to the K-Allele model at a rate of $\mu = 10^{-5}$ at each QTL per generation. The populations underwent natural stabilizing selection towards distinct phenotypic optima that varied clinally across the landscape, hence mimicking divergent selection among populations. The strength of the stabilizing selection simulated was mostly moderate ($\omega^2 = 50$) while the variance of the phenotypic optima ($\sigma^2_{Z_{opt}}$) equalled 6. Migration occurred among populations according to the Wright Island migration model with *Nm*=12.7. Finally, individuals mated randomly.
Each simulation was started from monomorphic populations homozygous at each locus. Before applying natural selection on the populations, we systematically ran a pre-period free of selection over 10 000 generations in order to generate consistent variability, correspodning to mutation-migration-drift equilibrium. Once the populations were at equilibrium, selection was introduced during 1000 generations. Our objective was to monitor the evolution of the mean genetic values of the populations and to assess the relative efficiency of the simulators by monitoring the time needed to achieve the simulations. We were mostly interested by the building up of a genetic cline in response to divergent selection, and aimed as comparing the slope of the cline among the different simulators. The simulations were realized on a Linux Ubuntu system exploiting a 2.4 GHz Intel Core Duo P9400 processor coupled with 4 Gb of RAM.

Setting up the scenario
The time we needed for the setting up of relevant input files was clearly related both with (1) the degree of relevance of the simulation programs with the scenario simulated and (2) their overall degree of flexibility. Roughly speaking, when the scenario was clearly included in the original scope of the models tested, the time of configuration was short. Conversely, models proposing other kinds of features, not tested here, required longer time of configuration. *Quantinemo* is highly specialized in the simulation of quantitative traits and provides numerous options quickly accessible. As a result, the input file to specify is concise and can rapidly be set. We particularly liked the very easy way proposed for the selection of the outputs needed. Initialization was also rapid in *Nemo,* which is close to *Quantinemo*. *Aladyn* enabled also for an easy and fast initialization, owing to the specialization of this tool and its relevance with the scenario we intended

to simulate. *Metapop*, also providing many features for the simulation of quantitative traits, requires both to describe the genetic architecture of individuals and to indicate the initial allelic frequencies, which resulted in large input files. However, the structure of the input file is simple and *Metapop*'s users mostly generate initial allelic frequencies automatically from external scripts. Moreover, a graphical interface, that will be released soon in a next version of *Metapop*, will greatly shorten the initialization phase. Initializing the scenario in *Forsim* was more difficult; more time was necessary to set up our 11 populations undergoing divergent clinal selection, and this part of the input file was more complex than within *Metapop, Quantinemo* and *Nemo* for three main reasons. First, within *Forsim,* simulations always start from a single population, which requires to further create populations by allocating individuals from the main population. Second, the genetic structure to start from is much complex in *Forsim* than in the other packages; the user has indeed to specify clearly the length and the position of the QTL, as well as the length of the chromosomes. Third, the overall structure of *Forsim*'s input files and the formalism provided gives access to quantity of interesting features but they differ from the input files of the other simulators. In our case, the creation of the input files and familiarization with the simulator were very time consuming. Finally, *Simupop* was the tool that proposes the less built-in feature dedicated to the simulation of quantitative traits. As mentioned above, this framework requires the user to write full Python scripts exploiting the numerous libraries provided, offering an impressive collection of possibilities but at the same time, complicating steadily the initialization of simulations. Table 6 gives as an indication, the number of lines of the input files we set up for this test.

Table 6. Input file length, mean execution time and memory usage.

	Input file length (lines)	Execution time	Memory usage (Mb)
Aladyn	31	58'11	125,9
Forsim	334	56'07	38,2
Metapop	43 (+848)*	6'48	2,1
Nemo	52	12'47	39,3
Quantinemo	39	8'38	33,2
Simupop	157	7'22	4,3

** additional file specifying initial allelic frequencies.*

Efficiency

We were first interested in the duration necessary to achieve the simulations from each simulator. To this end, we monitored the duration of the process with the Linux command *time* during the simulation of the scenario involving the strongest selection pressure. *Metapop* was the fastest program, achieving the simulation in a mean time of 6'48 over the 10 repetitions simulated (table 6). It was closely followed by *Simupop;* on average our python script simulated the scenario in 7'22. Otherwise, *Quantinemo* simulated the 11 000 generations in 8'38, *Nemo* in 12'47, while *Forsim* and *Aladyn* were clearly slower: 56'07 and 58'11, respectively. Both C (*Metapop*) and C++ (others) are known to be fast languages, and though differences among the compilers might produce executables with contrasting efficiency, the great speed of *Metapop* might be rather explained by the architecture of the source code. First of all, *Metapop* is totally appropriate to the simulation of the scenario considered here, which allowed for an optimal use of the computing power available. Besides, in comparison with the two quite similar programs *Nemo* and *Quantinemo,* the range of features and output produced by the original version of *Metapop* is smaller. Interestingly, the exploitation of *Simupop*'s libraries by our python script led to short time of execution also. Learning

how to use the numerous possibilities offered by the *Simupop* platform took considerable time, however it then allowed to write scripts both simple and totally appropriate to simulate the scenarios we defined, with a very limited output set. In the case of *Aladyn*, the simulation of continuous landscapes and realistic dispersal kernels clearly increases the complexity of the computation and running time of the simulations. Similarly, the important flexibility offered by *Forsim* and the numerous output provided, both graphical and textual, resulted in more important length of simulation.

In addition to the CPU-time, the memory usage yields a wider insight of the exploitation of the resources by the programs. We hence monitored the quantity of memory used by each package several times during the simulation process. In this field, *Metapop* required the less amount of memory (2.1 Mb), twice less than *Simupop* (4.3 Mb) and much less than *Nemo, Forsim*, *Quantinemo and Aladyn* which respectively needed 33.2, 38.2, 39.3 and 125.9 Mb (table 6). Owing to the reasons presented above, the limited memory usage of *Simupop* was not surprising. However, the very few amount of memory usage shown by *Metapop* indicates that developers paid considerable attention to memory saving during its phase of conception. *Metapop* was indeed coded several years before the other programs, at a time where RAM was less available and more expensive. Overall, execution time and memory usage are important points for users who intend to simulate landscapes including large numbers of entities, such as populations, individuals, QTLs, generations…

Genetic differentiation

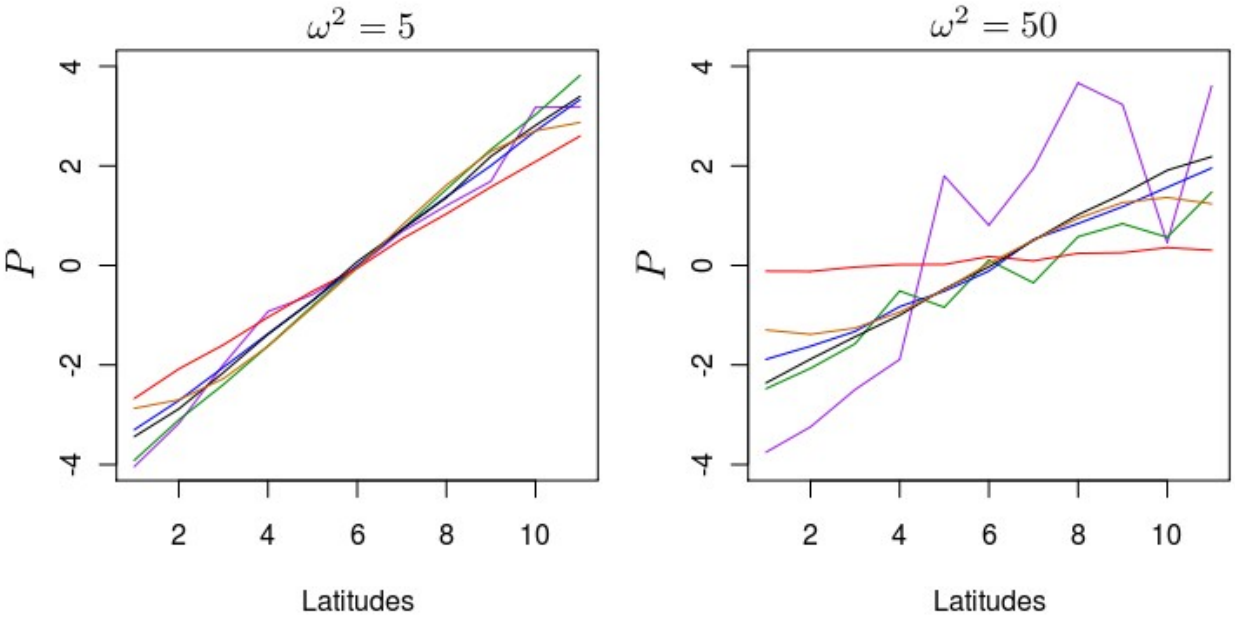
Though the evolutionary responses of populations showed similar trends among the simulators, we observed some important differences. First, the amount of genetic variability accumulated during the first 10 000 generations strongly varied. For example, with a mutation rate of 10-5, *Simupop* did not generate any significant variability while *Metapop* generated a mean phenotypic variability of 1.58 during the first phase of initialization. To ensure significant amounts of genetic variability to be generated in *Simupop*, we had to specify very high and unrealistic mutation rates in *Simupop* (10^{-3}). Moreover, because strong stabilizing selection ($\omega^2=5$) systematically led the populations to extinction in *Forsim* during the second part of the simulation, we simulated a less stringent selection intensity with $\omega^2=10$ in this scenario, only with *Forsim*. This two examples illustrate the difficulty to generate identical starting configurations and to simulate exactly equivalent evolutionary scenarios from several numerical models, owing to contrasts in effects of common settings and to the difference in the way the simulators can be parametrized.

The evolutionary responses of populations to intensive selection pressure involving a steep gradient of optimal values ($\sigma^2_{Zopt}=6$) and strong stabilizing selection ($\omega^2=5$) were pretty similar. But some differences in the extent of the differentiation were observed despite the strength of the selection pressures simulated. Hence, *Aladyn*, *Forsim*, *Metapop*, *Nemo* and *Simupop* generated the steeper phenotypic clines, close to the phenotypic optima defined, while *Quantinemo* led to slightly flatter phenotypic clines (figure 1). Surprisingly, the differentiation generated by *Aladyn* and *Forsim* were not exactly clinal. Moreover, the difference in the variance of the phenotypic values generated among repetitions was also very important in *Forsim*: the coefficient of variation of the mean phenotypic values of populations averaged -1.37 among the 10 repetitions simulated while it remained close to 0 in other simulators. Decreasing the strength of selection ($\omega^2=50$) amplified the differences in the patterns observed. In this latter case, the phenotypic differentiation generated by *Forsim* showed important fluctuating peaks at generation 11 000, while very limited differentiation was induced among the populations close to the edges of the landscape in *Aladyn* (figure 1). Similarly, the differentiation patterns generated by *Simupop* were less linear than under strong selection, but the overall differentiation produced remained roughly clinal. Overall, the phenotypic clines generated by *Aladyn*, *Metapop and Nemo* were here very similar, and here again, our populations responded less to selection in *Quantinemo* (figure 1).

It is interesting to see that *Aladyn*, *Metapop*, *Nemo* and *Simupop* produced phenotypic clines of very similar slopes. On the other hand, the populations differentiated less in *Quantinemo*, and more

differently in *Forsim*. Of course, our comparison is not exhaustive. A complete test would have required the consideration of much more scenarios considering variations in several other factors such as the number of QTLs, of alleles per QTL, of individuals, of populations, of generations

Figure 1. Variation in mean population phenotypic value across the landscape.



Phenotypic differentiation observed at the end of the simulation process, at generation 11 000. Moderate ($\omega^2=50$) and strong ($\omega^2=5$) stabilizing selection were associated with phenotypic optima (Z_{opt}) varying clinally over the landscape ($\sigma^2_{Z_{opt}}=6$). The green line refers to the differentiation generated by *Simupop*, black is for *Metapop*, blue for *Nemo*, red for *Quantinemo,* purple for *Forsim* and Orange for *Aladyn*. Each line is a mean of 10 repetitions.

Conclusion

All the simulation programs considered here enable to address numerous evolutionary questions regarding structured populations distributed in heterogeneous landscapes. They constitute valuable resources ideally complementing experimental or mathematical approaches. With the increasing availability of genetic data and consequent computing means, evolutionary simulation software will further greatly help us in understanding evolutionary dynamics of populations. Nonetheless, the features and models show few overlaps among the simulators, though the fields of application of the simulation packages considered in this study are close. Besides, our simple comparative test produced roughly similar evolutionary dynamics of the populations simulated in each tool, but it also revealed contrasted responses of populations to gradients of selection as well as differences in the amount of genetic variability that can be generated by mutations. These contrasts stem naturally from the differences in the original scope of each tool, but also, from the way in which the built-in models and features are implemented. Even when features are seemingly close, they are often handled singularly by each simulator. For example, epistatic interactions handled by 4 of the 6 simulators considered are systematically modeled according to distinct strategies (table 3). These intrinsic differences can add up and lead to important divergences at different points of a simulation process, but not necessarily. Hence, the evolutionary responses simulated by *Simupop* were here very close to those produced by *Metapop* and *Nemo,* despite the dissimilarity of this package

compared to others. Conversely, the gap between the differentiation simulated by the two seemingly closer tools *Quantinemo* and *Nemo* was slightly more important.
Obtaining similar evolutionary trajectories within a common scenario from all simulators is comforting. Nonetheless, using numerical models to address practical evolutionary issues and to help in determining relevant governance policies -noticeably regarding the management of ecosystems facing new threats related to climate change- might require very precise answers. The heterogeneity in results we observed here shows that the choice of the simulator clearly influence the possible outcomes. The required accuracy could be reached by exploring wider ranges of settings for a given evolutionary scenario, which is systematically done in works based on simulations *[Schiffers et al., 2012][Soularue & Kremer, 2012]*. Combining different models and approaches might be also an option *[Elith et al., 2010],* rarely considered with evolutionary numerical models. *In fine* these strategies might allow to formulate predictions about the extent of evolutionary change that can be expected from natural populations submitted to new selection pressures.
Finally, choosing the most relevant tool is a major prerequisite to any simulation study. Among the simulators tested here, *Aladyn*, *Metapop*, *Nemo* and *Quantinemo* offer the more relevant and complete features, including output, for straightforward modelling and monitoring quantitative phenotypes relying on complex genetic architecture. Choosing the most suitable among these 4 models will depend on the features required for a given study: number of traits to simulate, genetic models, dispersal models, overall complexity of the landscape, number of generations, output needed... Overall, once the hypothesis to test have been precisely outlined, the most appropriate tool is the one that offers the best trade-off between the length of the familiarization period and the range of the possibilities offered regarding the scenarios to simulate. This trade-off clearly depends on the degree of suitability between the nature of the built-in models provided by the simulator and the questions addressed. In the case of a very specific need, programming new modules in order to add required features to an existing familiar tool might be a suitable approach, even it is time consuming. In that latter case, the technical architecture of the program targeted is to consider, as well as the accessibility to code documentation. On the other hand, *Forsim* is ideal for users aiming at monitoring precisely the evolution of markers localization, it comes also with an interesting and flexible formalism enabling to model quantity of configurations involving complex genetic processes. Finally, there were great differences in terms of flexibility and evolvability among the packages. In these fields, *Simupop* and the numerous python libraries it provides have a very strong potential.

Acknowledgement

We sincerely thank Katja Schiffers, Bo Peng, Frédéric Guillaume, Samuel Neuenschwander for reviewing the lists of the features of their respective programs as well as the scripts we used in this work. We thank Brian Lambert for his help. Thanks to Jean-Baptiste Lamy for discussions and his comments about this manuscript. This study has been carried out with financial support from the commission of the European Communities (DG Research, Framework program 7) within the MOTIVE project (#226544).

References

Aitken SN, Yeaman S, Holliday JA, Wang T, Curtis-McLane S: **Adaptation, migration or extirpation: Climate change outcomes for tree populations.** *Evolutionary Applications* 2008, 1:95–111

Austerlitz A and Garnier-Géré: **Modelling the impact of colonisation on genetic diversity and differentiation of forest trees: interaction of life cycle, pollen flow and seed long-distance dispersal**. *Heredity* 2003, 90:282–290

Burger R, Lynch M: **Evolution and Extinction in a Changing Environment: A Quantitative-Genetic Analysis.** *Evolution* 1995**,** 49:151-163

Carvajal-Rodriguez A: **Simulation of genes and genomes forward in time.** *Current Genomics* 2010, 11:58-61

Epperson BK, Mcrae BH, Scribner K, Cushman SA, Rosenberg MS, Fortin MJ, James PM, Murphy M, Manel S, Legendre P, Dale MRT: **Utility of computer simulations in landscape genetics**. *Molecular Ecology* 2010, 19:3549–3564, doi: 10.1111/j.1365-294X.2010.04678.x

Elith J., Kearney M, Phillips S: **The art of modelling range-shifting species** *Methods in Ecology and Evolution* 2010, 1:330-342

Excoffier L and Heckel G: **Computer programs for population genetics data analysis: a survival guide. Nature Reviews Genetics**. 2006 7:745-758, doi:10.1038/nrg1904

Gienapp P, Teplitsky C, Alho JS, Mills JA, Merilä J: **Climate change and evolution: disentangling environmental and genetic responses**. *Molecular Ecology* 2008, 17:167–178, doi: 10.1111/j.1365-294X.2007.03413.x

Guillaume F, Rougemont J: **Nemo: an evolutionary and population genetics programming framework. *Bioinformatics* 2006 22(20): 2556-2557, doi: 10.1093/bioinformatics/btl415**

Hoban S, Bertorelle G, Gaggiotti OE: **Computer simulations: tools for population and evolutionary genetics**. *Nat Rev Genet.* 2012, 2012 13(2):110-22, doi: 10.1038/nrg3130

Lambert BW, Terwilliger JD, Weiss KM: **Forsim: a tool for exploring the genetic architecture of complex traits with controlled truth.** *Bioinformatics* 2008, 24 (16):1821-1822, doi: 10.1093/bioinformatics/btn317

Lande R: **Adaptation to an extraordinary environment by evolution of phenotypic plasticity and genetic assimilation**. *J. Evol. Biol.* 2009, 22:1435–1446

Le Corre V and Kremer A: Genetic Variability at Neutral Markers, Quantitative Trait Loci and Trait in a Subdivided Population Under Selection. Genetics 2003, 164(3):1205-1219

Le Corre V, Machon N, Petit RJ, Kremer A: **Colonization with long-distance**

seed dispersal and genetic structure of maternally inherited genes in forest trees: a simulation study. *Genet. Res.* 1997 69:117-125

Lynch M and Lande R: **Evolution and extinction in response to environmental change**. In P. Kareiva, J. Kingsolver, and R. Huey (eds.) Biotic Interactions and Global Change. Sinauer Assocs., Inc. Sunderland, MA, 1993, pp. 234-250.

Machon N, Bardin P, Mazer SJ, Moret J, Godelle B, Austerlitz F: **Relationship between genetic structure and seed and pollen dispersal in the endangered orchid Spiranthes spiralis**. *New Phytologist* 2003, 157:677-687

Neuenschwander S, Hospital F, Guillaume F, Goudet J: **Quantinemo: an individual-based program to simulate quantitative traits with explicit genetic architecture in a dynamic metapopulation. *Bioinformatics* 2008 24(13):1552-3**

Peng B, Kimmel M, Amos CI: **Forward-Time Population Genetics Simulations: Methods, Implementation, and Applications**. *Wiley-Blackwell* 2012, Oxford.

Peng B and Kimmel M: **Simupop: a forward-time population genetics simulation environment**. *Bioinformatics* 2005, 21(18):3686-3687, doi:10.1093/bioinformatics/bti584

***Rice SH: Evolutionary Theory: Mathematical and Conceptual Foundations*, *Sinauer* Associates Inc. Publishers, 2004, Sunderland.**

Richards FJ: **A Flexible Growth Function for Empirical Use**. *J. Exp. Bot* 1959 10(2):290-301, doi:10.1093/jxb/10.2.290

Schiffers K, Bourne EC, Lavergne S, Thuiller W and Travis J: **Limited evolutionary rescue of locally adapted populations faceing climate change**. Philosophical transactions of the royal society B, 2012.

Soularue JP and Kremer A: **Assortative mating and gene flow generate clinal phenological variation in trees**. *BMC Evolutionary Biology* 2012, 12:79

Turelli M: **Heritable genetic variation via mutation-selection balance: Lerch's zeta meets the abdominal bristle**. *Theor. Popul. Biol.* 1984, 25:138:193

Yeaman S. and Whitlock M.C: **The genetic architecture of adaptation under migration-selection balance**. *Evolution* 2011, 65:1897–1911, doi: 10.1111/j.1558-5646.2011.01269.x

Supplementary Table 1. Simulators accessibility: OS, requirement, formats and helping resources.

	Installation and requirements	Download	Input formats	Output formats	Helping resources
***Aladyn* v 1.0**	- Compilation needed - Requires *g*++ compiler	http://www.katja-schiffers.eu/docs/Allele_Model.zip	- Source code	- Text file	- Readme file
***Forsim* v. 2.0.0 beta**	- Compilation needed - Requires *g*++ compiler	On request : Brian Lambert bwl1@psu.edu	- Text file	- Text and graphic files	- Website: *http://www.anthro.psu.edu/weiss_lab/research.shtml* - (Lambert *et al.* Bioinformatics 2008) - User manual
***Metapop* v 1.0**	- Compilation needed. - Requires *gcc* compiler	On request : Antoine Kremer, *kremer@pierroton.inra.fr*; Frédéric Raspail *raspail@pierroton.inra.fr*	- Text file	- Text file	None
***Nemo* v2.2.0**	- Executable available - Compilation possible via *g*++ - Run only through Cygwin or MS-DOS console on Windows systems	*http://nemo2.sourceforge.net/download.html*	- Text file	- Text file	- Website: *http://nemo2.sourceforge.net/* (Guillaume and Rougemont 2006) - User manual
***Quantinemo* v.1.0.4**	- Executable available - Compilation via *g*++	*http://www2.unil.ch/popgen/softwares/Quantinemo/*	- Text file	- Text file	- Website: *http://www2.unil.ch/popgen/softwares/Quantinemo* (NeuenSchwander *et al,* 2008) - User manual
***Simupop* v1.0.7**	- Executable installer - Compilation recommended under Linux - Compilation via *g*++ - Requires Python	*http://Simupop.sourceforge.net/Main/Download*	- Python scripts	- Text file and graphics with *rpy*	- Website: *http://Simupop.sourceforge.net/Main/homepage* - (Peng and Kimmel 2005) - User-guide, exhaustive class documentation, source code documentation, tutorial, book, forum,

Chapter 2

Assortative mating and gene flow generate clinal phenological variation in trees

Jean-Paul Soularue and Antoine Kremer*

INRA, UMR 1202 BIOGECO, Cestas F-33610, France

Univ. Bordeaux, BIOGECO, UMR 1202, Talence F-33400, France

*Corresponding author: antoine.kremer@pierroton.inra.fr, phone: +33 5 57 12 28 32, fax: +33 5 57 12 28 81

Accepted in BMC Evolutionary Biology

Abstract

On-going climate change is shifting the timing of bud burst (TBB) of broad leaf and conifer trees in temperate areas, raising concerns about the abilities of natural populations to respond to these shifts. The level of expected evolutionary change depends on the level and distribution of genetic variation of TBB. While numerous experimental studies have highlighted the role of divergent selection in promoting clinal TBB differentiation, we explored whether the observed patterns of variation could be generated by the joint effects of assortative mating for TBB and gene flow among natural populations. We tested this hypothesis using an *in silico* approach based on quantitative genetic models. Our simulations showed that genetic clines can develop even without divergent selection. Assortative mating in association with environmental gradients substantially shifted the mean genetic values of populations. Owing to assortative mating, immigrant alleles were screened for proximal or distant populations depending on the strength of the environmental cline. Furthermore, we confirmed that assortative mating increases the additive genetic variance within populations. However, we observed also a rapid decline of the additive genetic variance caused by restricted gene flow between neighboring populations resulting from preferential matings between phenologically-matching phenotypes. We provided evidence that the patterns of genetic variation of phenological traits observed in forest trees can be generated solely by the effects of assortative mating and gene flow. We anticipate that predicted temperature increases due to climate change will further enhance genetic differentiation across the landscape. These trends are likely to be reinforced or counteracted by natural selection if phenological traits are correlated to fitness.

Keywords: phenology, assortative mating, gene flow

Introduction

Apical bud phenology of temperate trees has been intensively studied in recent years owing to predicted shifts in the timing of bud development as a result of climate changes *[Bertin et al., 2008]*. Monitoring of leaf unfolding in various species across their distributions has shown that global warming will trigger earlier flushing *[Menzel et al., 2006][Nordli et al.2008][Vitasse et al., 2011]*. These observations have raised concerns about the capacity of tree populations to cope with changes in the timing of bud burst (TBB), which is related to the fitness of trees in two ways: *(i)* it establishes the length of the growing season and is a major determinant of growth *[Bennie et al., 2010]*, *(ii)* it determines the timing of flowering, so is related to fecundity *[Franjic et al., 2011]*. The adaptive response of TBB to global warming is dependent on the level and distribution of genetic variation within a species; the more variation, the larger the predicted genetic shift in TBB. Numerous investigations involving common garden experiments have demonstrated that TBB exhibits large intra- and inter-population differences, as shown by high population differentiation Q_{ST} associated with high heritability values *[Kremer et al. 2010]*. Additional genetic investigations indicated that juvenile-mature correlation in TBB is high and genotype-environment interactions are low *[Ekberg et al. 2010]*. Finally, genetic dissection by quantitative trait loci (QTLs) mapping has shown that many QTLs contribute to TBB, but these QTLs show stable expression over years and sites *[Derory et al. 2010]*.

Regardless of species, TBB follows strong geographic clinal patterns of variation, either altitudinal, latitudinal or longitudinal. Phenotypic clines revealed by *in situ* observations of TBB show congruent patterns across species: bud burst in southern latitudes or lower altitudes occurs earlier than in northern latitudes or higher altitudes *[Worall, 1983][Vitasse et al., 2009][Alberto et al., 2011]*, because TBB is triggered by heat sum *[Chuine and Cour, 1999]*. Genetic clines can be assessed in common garden experiments where TBB is observed under the same environmental conditions for all populations and are illustrated by the linear relationships between TBB of different populations and geographic variables. Interestingly, genetic clines vary across species and exhibit co-gradient variation or counter-gradient variation with geographic variables and associated phenotypic clines *[Conover, 1995][Conover et al., 2009]*. Co-gradient variation corresponds to clines of both phenotypic variation and genetic variation in a species that co-vary in the same way with the environmental gradient. Counter-gradient variation occurs when phenotypic and genetic clines vary in opposite directions. In the case of oak, genetic and phenotypic clines exhibit co-gradient variation; *e.g.* populations from southern latitudes flush earlier than populations from northern latitudes, when assessed under the same conditions in common gardens *[Jensen and Hansen, 2008][Vitasse, 2009 (CJFR)]*. In the case of beech, genetic clines are opposite to phenotypic clines and exhibit counter-gradient variation: provenances from northern latitudes flush earlier than populations from southern latitudes *[Wuelish et al., 1995][Chmura and Rozkowski, 2002]*.

Clinal variations, either co- or counter-gradient, have usually been interpreted as consequences of divergent selection among populations by either abiotic or biotic selection pressures. For example, late-flushing trees will not suffer the detrimental effects of late frosts *[Howe et al., 2003]* or may avoid damage by defoliating insects *[VanAsch et al., 2007][Ghelardini and Santini, 2009]*. However, few studies have considered the impacts of other evolutionary factors, such as gene flow in combination with the peculiar features of bud burst, in shaping the genetic variation of TBB. Indeed, because trees mate assortatively by flowering time *[Kirkpatrick, 2000][Fox, 2003]*, and because TBB is tightly linked to the timing of flowering, assortative mating is likely to shape the variation of TBB. Furthermore, under assortative mating, immigrant pollen will introduce genes likely to generate new allelic combinations for TBB, owing to the existence of environmental clines.

A number of theoretical studies have dissected the effects of assortative mating on the evolution of

quantitative traits under polygenic inheritance, beginning with the early investigations by Fisher (1918)*[Fisher, 1918]* and Wright (1921)*[Wright, 1921]*. All predicted that assortative mating will increase genetic variation as a result of the build up of genetic covariations among loci *[Fisher, 1918][Lande, 1977][Rosvall and Mullin, 2003][Weis and Kossler, 2004]*. Others demonstrated the amplifying role of assortative mating on natural selection *[Jorjani, 1997 (3)][Fox, 2003]*, as well as its contribution to allopatric speciation *[Caisse and Antonovics, 1978][Devaux and Lande, 2008]*. Finally, more recent studies aimed at predicting the effects of assortative mating on the genetic covariance of different trait*s [Gianola, 1982][Hayashi, 1998][Weis, 2005]*. No prior investigations, however, have considered the effects of assortative mating on a trait in multiple populations interconnected by extensive gene flow in the presence of environmental gradients. We tested whether interactions between gene flow and assortative mating under such circumstances could generate the distribution of genetic variation that is observed in common garden experiments, even in the absence of divergent selection. Our main hypothesis was that assortative mating, by filtering incoming alleles among interbreeding populations, will change the genetic composition and the genetic values of the phenological trait in recipient populations and hence generate population differentiation. We mainly focused on the maintenance of high within- and between-population genetic variation and on the build-up of genetic clines. There exists no available analytical theoretical prediction of genetic variation and differentiation taking into account assortative mating. We therefore used a simulation approach allowing us to monitor *in silico* the evolution of TBB under contrasting levels of assortative mating and environmental clines.

Methods

Components of population subdivision

Our main objective was to track components of genetic variation in phenology-related traits in a subdivided population that would mimic extant ecological settings. We were primarily interested in assessing the within- and between-population genetic variances (V_W and V_B) as well as the differentiation among populations as measured by Q_{ST}, which are standard genetic measurements used in quantitative genetics.

$$Q_{ST} = \frac{V_B}{V_B + 2V_W} \qquad (1)$$

where V_W is the within-population genetic variance, and V_B is the between-population genetic variance.

As suggested by recent QTL studies *[Derory et al., 2010][Saintagne et al., 2004]*, we assumed that phenological traits were controlled by multiple QTLs with only additive effects. Previous theoretical studies have also shown that the genetic variances V_B and V_W of multilocus traits can be substantially inflated by allelic covariations among loci *[Le Corre and Kremer, 2003]*.

$$V = \sum_i \sigma_i^2 + \sum_i \sum_{j \neq i} Cov_{ij} \qquad (2)$$

where σ_i^2 is the genic variance of locus *i* and Cov_{ij} is the covariance between allelic effects at locus *i* and *j*.

V stands for V_B or V_W with appropriate σ_i^2 and Cov_{ij} expressed either at within- or between-

population levels.

These covariations build up as a result of within- or between-gametic disequilibrium generated by different evolutionary forces and are scaled by the parameters θ_W and θ_B.

$$\theta = \frac{\sum_{i=1}^{n} \sum_{j \neq i}^{n} Cov_{ij}}{\sum_{i=1}^{n} \sigma_i^2} \qquad (3)$$

Le Corre and Kremer (2003) *[Le Corre and Kremer]* and Kremer and Le Corre (2011) *[Kremer and Le Corre]* showed how the θ values contributed to the final differentiation of the trait together with the genetic differentiation that also arises at the QTLs controlling the trait G_{ST_q}.

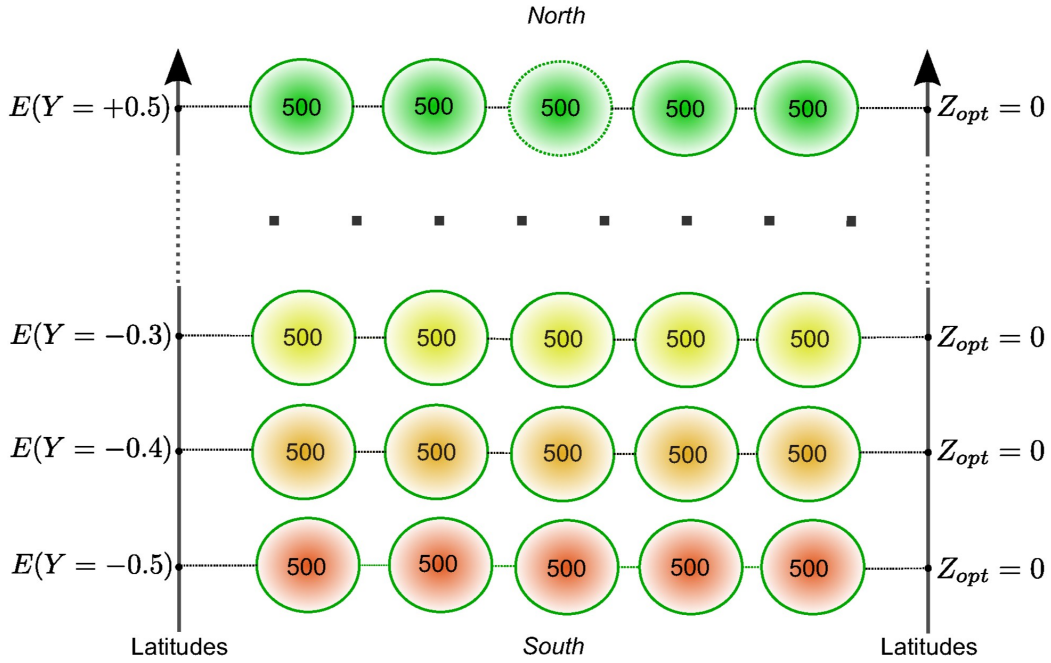
$$Q_{ST} = \frac{(1+\theta_B) G_{ST_q}}{(\theta_B - \theta_W) G_{ST_q} + 1 + \theta_W} \qquad (4)$$

A major finding of previous theoretical work was that divergent selection generates important between-population disequilibria that becomes a major driver of population differentiation (Q_{ST}) and has only a minor impact on differentiation at QTLs (G_{ST_q}). In the absence of selection and under random mating, θ_W and θ_B should be 0 and Q_{ST} equal to G_{ST_q}. We will explore in these simulations how assortative mating will shape the distribution of genetic variability by monitoring the different components of Q_{ST} (*e.g.* V_W, θ_W, θ_B, V_B, and G_{ST_q}) under different evolutionary scenarios.

Models and simulations

We used the Metapop simulation engine to assess evolutionary changes along successive generations in a subdivided population. Essential steps of the evolutionary processes included in the software - mutation, gene flow, selection, demographic growth - have been described in earlier papers *[Le Corre and Kremer, 2003][Le Corre et al., 1997][Le Corre and Kremer, 2011][Machon et al., 2003]*. We will only address here the changes introduced to account for assortative mating and phenotypic clines of phenological traits.

Figure 1. Spatial settings of populations and environmental effects.



Fifty-five populations of 500 individuals each were spread homogeneously on a 5 x 11 grid along 11 latitudinal positions. *E(Y)* represents the environmental effect at a given latitude *Y* and is scaled by k_E (see equation 8). No selection was introduced: stabilizing selection was canceled with $\omega^2 = 10^9$ and all populations shared a phenotypic optimum $Z_{opt} = 0$.

Phenotypic subdivision of phenological traits

Populations are positioned on a two-dimensional grid (figure 1) that mimics in a discrete way real situations showing continuous environmental variations. Each population is composed of N individuals. The overall phenotypic value Z'_{ij} of individual i from population j is composed of three components: the additive part G_{ij} of the genes contributing to the trait, the environmental component E_j and a random local environmental deviation ϵ_{ij}.

$$Z'_{ij} = G_{ij} + E_j + \epsilon_{ij} \quad (5)$$

And the within-population phenotypic value is

$$Z_{ij} = G_{ij} + \epsilon_{ij} \quad (6)$$

G_{ij} is the genetic value resulting from the sum of additive effects of alleles present at n QTLs controlling the trait.

$$G_{ij}=\sum_{l=1}^{n}(\alpha_1+\alpha_2)_l \qquad (7)$$

α values are drawn at loci from the distribution $N(0,\sqrt{W_l\times\sigma^2_{A_0}/2})$, where W_l is the level of contribution of the *l*th locus considered and $\sigma^2_{A_0}$ the initial variance of allelic effects based on estimated values of additive variance in experimental plantations. More details on the method are available in *[Kremer and Le Corre, 2011]*.

E_j represents the influence of environmental conditions at the location of population *j*. E_j is of the same magnitude for all individuals of population *j* located at latitude *Y*. In our study case, *E* accounts for the effect of temperature on TBB demonstrated in forest trees *[Vitasse et al., 2009]*; indeed, flushing dates of broadleaves and conifers are tightly dependent on the heat sum *[Chuine and Cour, 1999]* and exhibit continuous variation with latitude, resulting in environmental clines of *E* values. This is the rationale of assigning the same E_j value to all trees of population *j*. The linear variation of E_j along latitude, which corresponds to an environmental cline, results in the phenotypic cline as observed *in natura* (figure 2). The steepness of the environmental cline is scaled by k_E, a standardized measure of the between-environment variance relative to the within-population phenotypic variation. We considered different levels of steepness of the environmental cline by taking different values of k_E:

$$k_E=\frac{\sigma^2_E}{(\sigma^2_{G_0}+\sigma^2_\epsilon)} \qquad (8)$$

$\sigma^2_{G_0}$ being the total genetic variance observed within the initial population. Hence k_E is constant over the generations through the evolutionary process.

Given that E follows a linear relationship with latitude, we can assign environmental values E_j according to

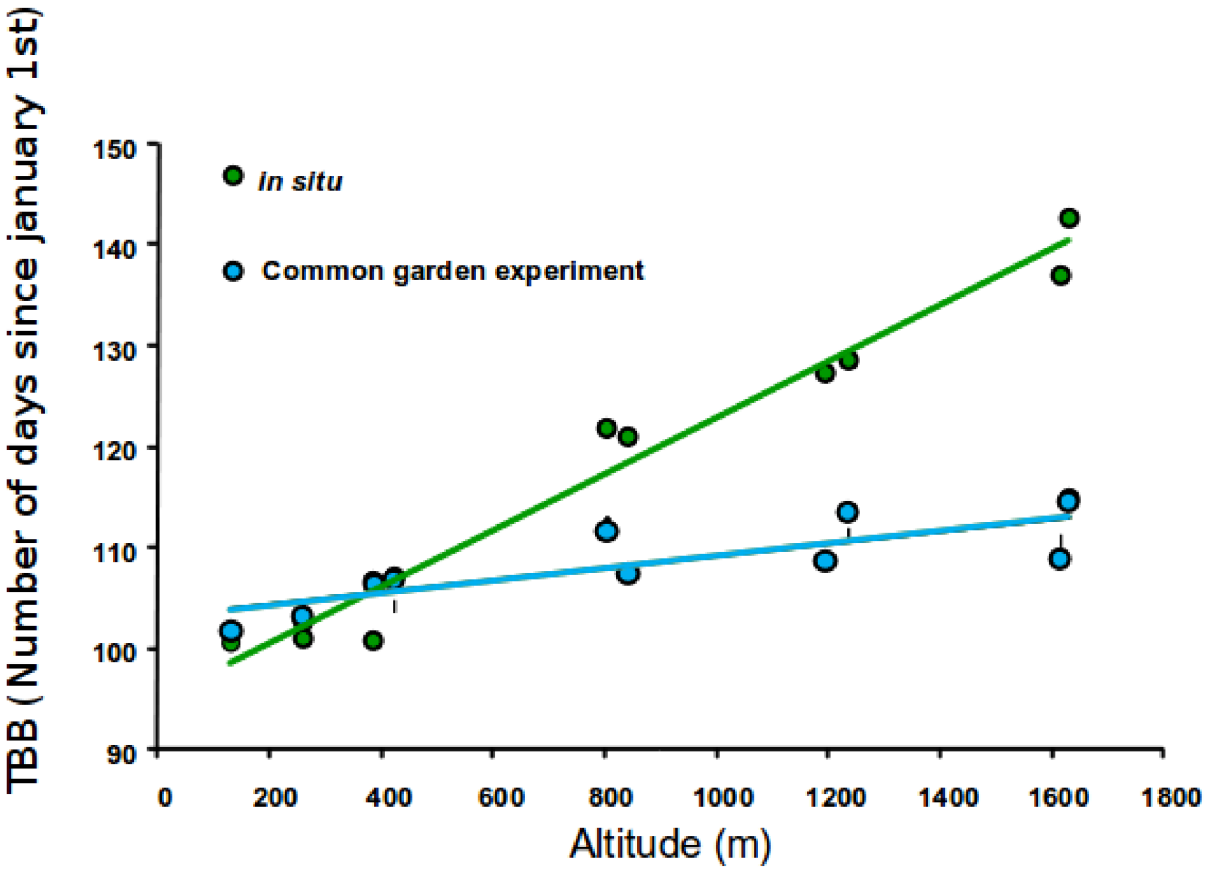
$$E_j=\sqrt{\frac{k_E\times(\sigma^2_{G_0}+\sigma^2_\epsilon)}{\sigma^2_Y}}\times Y_j \qquad (9)$$

Finally, ϵ_{ij} is a random local environmental deviation following the distribution $N(0,\sigma_\epsilon)$.

Sequence of evolutionary processes in Metapop

Metapop implements evolutionary processes over successive generations in a subdivided population. Within each generation, processes are simulated along four steps within a main loop, depicted in supplemental figure 1. First, fitness values of reproducing individuals are computed according to stabilizing and divergent selection. The level of stabilizing selection is scaled by the parameter ω^2 from Turelli's relation *[Turelli, 1984]* while the strength of divergent selection is scaled by $\sigma^2_{Z_{opt}}$, where Z_{opt} of a given population is the phenotypic value for which trees have the highest fitness in that population.

Figure 2. An example of environmental and genetic clines for time of bud burst in oaks (data of *[Alberto et al., 2011]*).



The time of bud burst (TBB) was recorded in sessile oak stands located along two valleys on the northern side of the Pyrénées mountains. *In situ* observations (green dots on the graph) showed that trees located at higher elevations flushed much later then trees located at lower altitudes, as a result of strong correlations between TBB and heat sum *[Vitasse et al., 2011]*. This pattern of variation, the phenotypic cline, is clearly linear. Open-pollinated seeds were collected in each stand and were experimentally raised in a common garden at low altitude, and TBB was monitored (blue points). The TBB was plotted as a function of the altitudes where the seeds were collected. A linear pattern of variation corresponds to a genetic cline. This example illustrates a co-gradient pattern of variation, because the slopes of the phenotypic and genetic clines share the same sign. Counter-gradient variation corresponds to cases where the two clines vary in opposite directions.

Second, from the populations' growth settings and seed migration matrix, the number of individuals of each population contributing to the future generation is computed. Third, mates are chosen based on the constraints due to assortative mating scaled by the correlation between Z_i and Z_j, the overall phenotypic values of individuals *i* and *j*.

$$\rho = \frac{cov(Z_i', Z_j')}{\sigma_{Z_i'} \times \sigma_{Z_j'}} \qquad (10)$$

Following (10), the differences in phenotypic values of two mating parents are drawn from the distribution $N(0,\sigma_\delta)$ with

$$\sigma_\delta=\sqrt{\frac{\sigma^2_{Z_i'}}{\rho^2}-\sigma^2_{Z_i'}} \quad (11)$$

Fertilization occurs by drawing male and female gametes conditionally to ρ, fitness of the parents and seeds migration matrix. A proportion of male gametes, based on the pollen migration matrix, is drawn from other populations to account for pollen flow. Finally, mutation is also considered.

Monitoring of gene flow

We now consider how the interaction between gene flow and assortative mating may modify the genetic values in natural populations. Because assortative mating will filter immigrant alleles so that they can mate with trees of recipient populations, we compare the genetic values of immigrant alleles to local alleles to explore whether gene flow will modify the mean genetic value of populations.

In each generation, matings take place between trees of the same population, but a fraction m_p of matings involves pollen from other populations. We can subdivide the genetic value of the offspring into two components:

$$G_{t+1}=(1-m_p)(\frac{1}{2}G_t^{\varphi}+\frac{1}{2}G_t^{\sigma})+m_p(\frac{1}{2}G_t^{\varphi}+\frac{1}{2}G_t^{*}) \quad (12)$$

where G_t^{φ} and G_t^{σ} stand respectively for the mean genetic values of the female and male parents, and G_t^{*} stands for the mean genetic value of the male parents providing immigrant alleles at generation *t*. $(1-m_p)(\frac{1}{2}G_t^{\varphi}+\frac{1}{2}G_t^{\sigma})$ represents the component of the genetic value due to intra-population matings and $m_p(\frac{1}{2}G_t^{\varphi}+\frac{1}{2}G_t^{*})$ the component of the genetic value due to inter-population matings involving external incoming alleles. Each generation, $G_t^{\varphi}=G_t^{\sigma}=G_t$.

When assortative mating occurs within populations, mating parents share similar phenotypic values, and because they belong to the same population, they also share the same environmental values. In this case, the value of the first component of equation (12) will not change between successive generations. However, because male parents from the outside populations should share the same phenotypic value as the female parent, their genetic values are likely to be different from those of the female parents owing to the environmental gradient. Within a population, the mean phenotypic value of the male parents corresponding to the immigrant alleles is equal to

$$Z_t'^{*}=G_t^{*}+E^{*} \quad (13)$$

and the mean phenotypic value of the female parents is equal to

$$Z_t'=G_t+E \quad (14)$$

Because the phenotypic values of both parents should be similar owing to assortative mating, the mean genetic value of the male parents is

$$G_t^* \simeq G_t + E - E^* \qquad (15)$$

As a result, each generation the genetic value of the population is expected to shift by about $\Delta = G_{t+1} - G_t$, which can be expressed in

$$\Delta \simeq \frac{1}{2} m_p (E - E^*) \qquad (16)$$

More generally, matings that occur within populations can be subdivided in two different kinds: (1) matings between individuals sharing similar genetic values, which would correspond to positive assortative mating and (2) matings between individuals likely to have different genetic values resulting from gene flow. In the extreme case, these matings may result from negative assortative mating. The shift of the genetic value is therefore driven by the level of effective gene flow m_p and the difference in environmental values between the recipient and donors populations. Consequently, we monitored the effective pollen flow during the simulations by tracking its spatial origin.

Simulations settings

We simulated the evolution of 55 populations of 500 individuals each spread homogeneously on a 5x11 grid depicted in figure 1. We did not consider overlapping generations and the number of individuals per population was kept constant over successive generations. A fictive gradient of latitudes was set from latitude -0.5 to latitude $+0.5$ in steps of 0.1. Three levels of environmental clines were considered along the latitudinal gradient: $k_E = 1$, $k_E = 2$ and $k_E = 3$.

Recent observations in oak populations suggested that assortative mating for TBB is substantial *[Franjic et al. 2011]*. Indeed, the flowering time in oak may extend over several weeks within a population, but the receptive period of female flowers lasts only a few days at the individual level. We consequently investigated two strengths of assortative mating, encompassing the suspected range of variation, using $\rho = 0.3$ and $\rho = 0.8$ to model moderate and strong assortative mating, respectively. Random mating was considered as well with $\rho = 0$.We used Wright's island migration model to generate gene flow among populations located on the grid system, and considered two levels of gene flow: $N_m = 5.1$ and $N_m = 10.2$. These values fit the range of variation of F_{ST} values (2.4% to 4.7%) observed in natural oak populations [7]. Pollen and seed migration rates (m_p and m_s) were then inferred from N_m values and introduced in the simulations, assuming further that $m_p = 100 * m_s$ (table 1). In addition to the island model, we also designed gene flow via the stepping stone model using pollen and seed migration rates corresponding to $N_m = 5.1$.

Table 1. Initial simulation settings.

heritability h^2	0,83
selfing rate *s*	0,02
nb. of populations *d*	-55
nb. of ind. per pop. N_{ind}	500
pollen migration rates m_p	0.02, 0.04
seed migration rates m_s	0.0002, 0.0004
nb. of QTL *n*	10
mutation rate μ	10^{-5}
nb. of latitude levels *Y*	11
interval of latitudes *Y*	[-0.5, +0.5]
steepness of environmental cline *E* scaled by k_E	1, 2, 3
variance of Z_{opt} across latitudinal levels $\sigma^2_{Z_{opt}}$	0, 1
intensity of stabilizing selection ω^2	10^9, 5
assortative mating intensity ρ	0, 0.3, 0.8

Assuming that the starting populations were in mutation-migration-drift equilibrium, initial allelic frequencies in different populations were drawn from a Dirichlet distribution *[Kremer and Le Corre, 2011]*. We assumed that phenological traits were controlled by 10 QTLs. Additive values of alleles were chosen at random from a Gaussian distribution whose initial variance was adjusted to fit the heritability values observed in extant progeny plantations, 0.83 from *[Cornelius, 1994]*. Mutations at each QTL occurred across generations at a rate of $\mu = 10^{-5}$ per generation. The local environmental deviation was drawn at random from the distribution $N(0,1)$ (table 1).

Table 2. Evolutionary scenarios.

	$\rho=0$	$\rho=0.3$	$\rho=0.8$
$k_E=1$	X^*	X	X
$k_E=2$	X^*	X	X, X^s, X^m
$k_E=3$			X

** identical scenarios; because under random mating, phenotypic values of individuals have no impact on our simulation outcomes, variations in the environmental component have no influence when* $\rho=0$. X^s *and* X^m *stand respectively for scenarios simulated under the stepping-stone migration model and with a higher migration rate (* $N_m=10.2$ *) under the island migration model.*

We considered eight different evolutionary scenarios by combining unique slopes of environmental clines, levels of assortative mating, migration models, and levels of gene flow (table 2). Because our investigations were focused on the impact of gene flow and assortative mating on the evolution of TBB, we purposely excluded selection in the simulations. We consequently canceled stabilizing selection within all populations by setting all ω^2 values to 10^9, and we defined uniform selection with $\sigma^2_{Z_{opt}}$ set to 0.

However, as a control, we added one scenario including selection ($\omega^2=5$ and $\sigma^2_{Z_{opt}}=1$), corresponding to strong stabilizing selection and moderate divergent selection. This scenario did not consider assortative mating and was designed to compare the steepness of the genetic clines observed in the eight studied cases with a selective scenario. For each evolutionary scenario based on combinations of these settings (table 2), we performed 50 independent replicated simulations over 1000 generations.

Results

Within population genetic variance

Assortative mating substantially increased allelic covariances during the first generations (figure 3). After reaching maximum values, covariances decreased very rapidly and evolved to asymptotic levels. These patterns were more pronounced when assortative mating was strong and were only slightly modified by the magnitude of the environmental cline. Under strong assortative mating, covariances accounted for more than 1.5 of the genic variances relative to the total genetic variance, while under moderate assortative mating, the maximum value was only 0.28. Under steeper clines, the maximum values of θ_W were slightly higher, 1.5 *vs* 1.4, and its change over generations was slightly delayed. Overall θ_W values were always larger under assortative mating than under random mating.

The variations in θ_W had striking consequences on the genetic variances (equation (2)). Indeed, under assortative mating, genetic variances increased rapidly during the early generations, then they very rapidly dropped below even the level of genetic variance reached under random mating. As for covariances, there was a strong effect of the level of assortative mating and only a minor effect of the environmental cline. The decrease in genetic variance due to assortative mating could be dramatic after 400 generations. Furthermore, the final heritability for the trait was divided by a factor 2.5 at generation 500. As expected without selection in large populations, genetic variance was maintained under random mating and extensive gene flow.

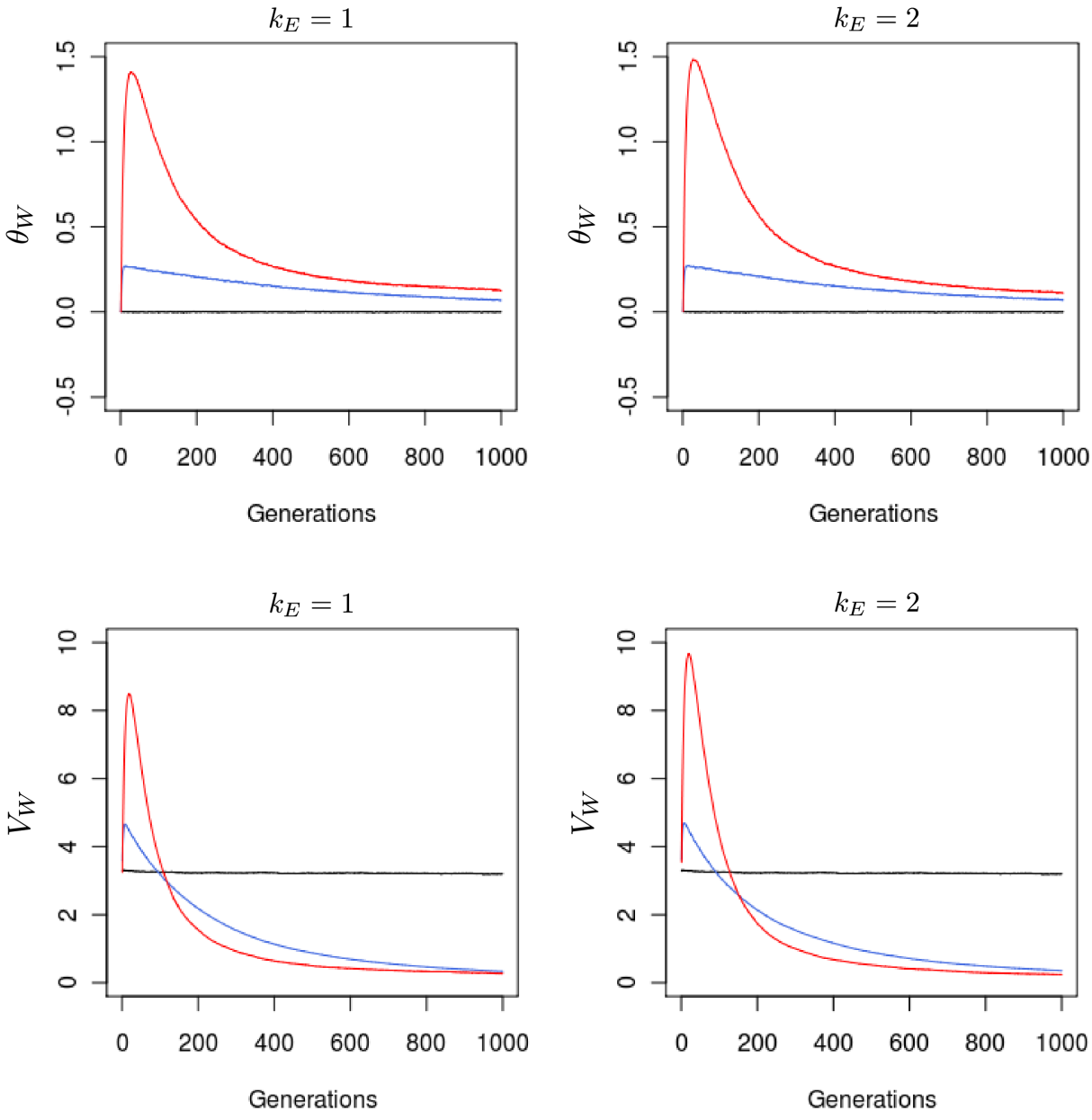
Between population genetic variance

Assortative mating had a strong effect on allelic covariances at the between-population level; θ_B increased during the early generations and was maintained at higher values through the 1000 generations, in contrast to θ_W values. There was a stronger impact when environmental clines were steeper. For example, under strong assortative mating, the maximum value of θ_B was 2.7 when $k_E=2$ *vs* 2.5 when $k_E=1$. The initial phase of increase lasted longer under moderate assortative mating than under strong assortative mating: 500 generations *vs* 230 generations when $k_E=1$ (figure 4).

Between-population variances of allelic frequencies at selected loci increased steadily over generations. They increased more rapidly under strong assortative mating, while no substantial differences were observed between random mating and moderate assortative mating. By generation 1000, differentiation at selected loci had reached 0.16, which could be compared with differentiation under random mating (0.03), which was very close to differentiation at neutral

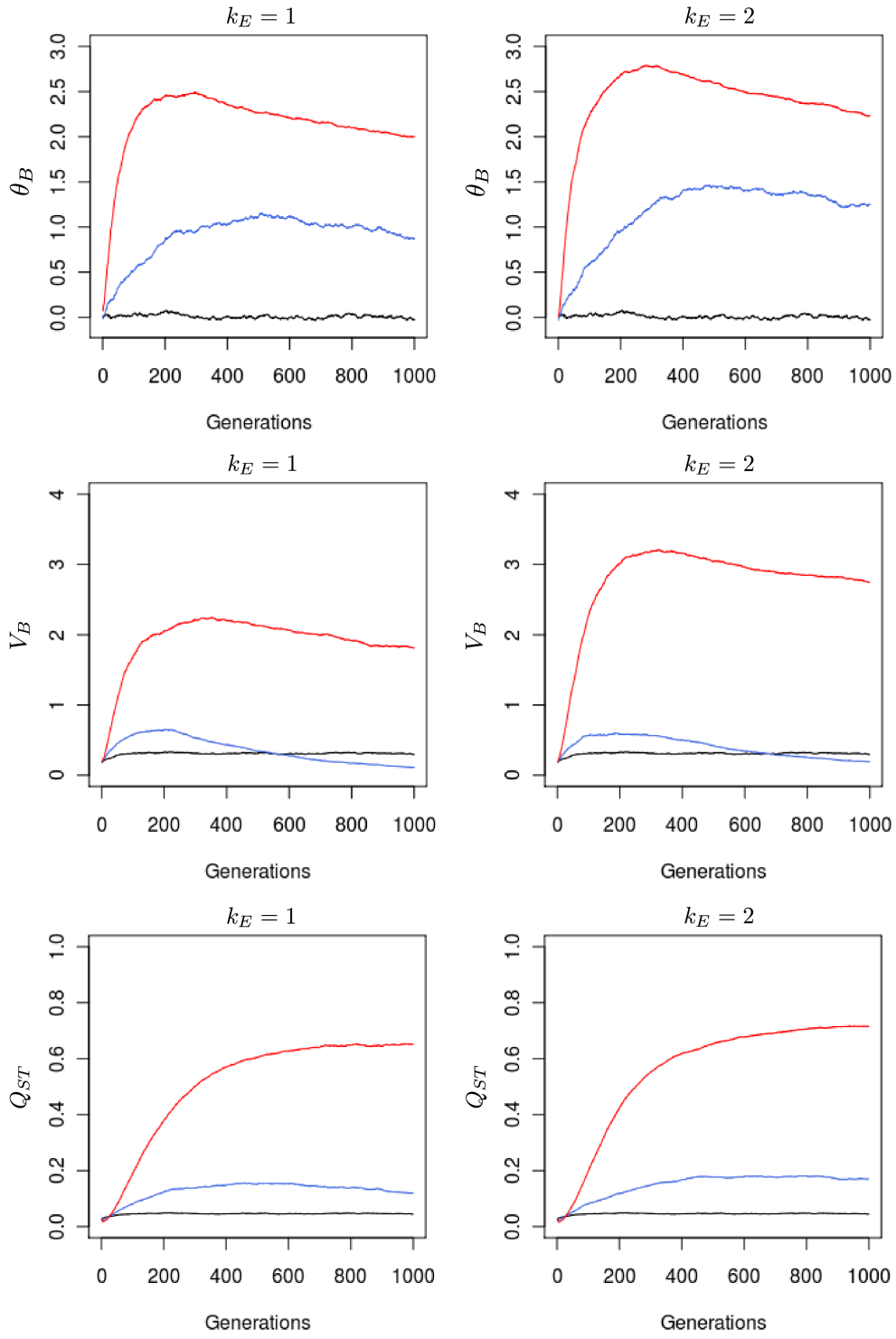
markers (0.024) (data not shown). Overall, between-population genetic variances exhibited strong differences between moderate and strong assortative mating and also between low and strong environmental clines (figure 4).

Figure 3. Variations in within-population allelic covariation (θ_w) and genetic variance (V_w) under different evolutionary scenarios.



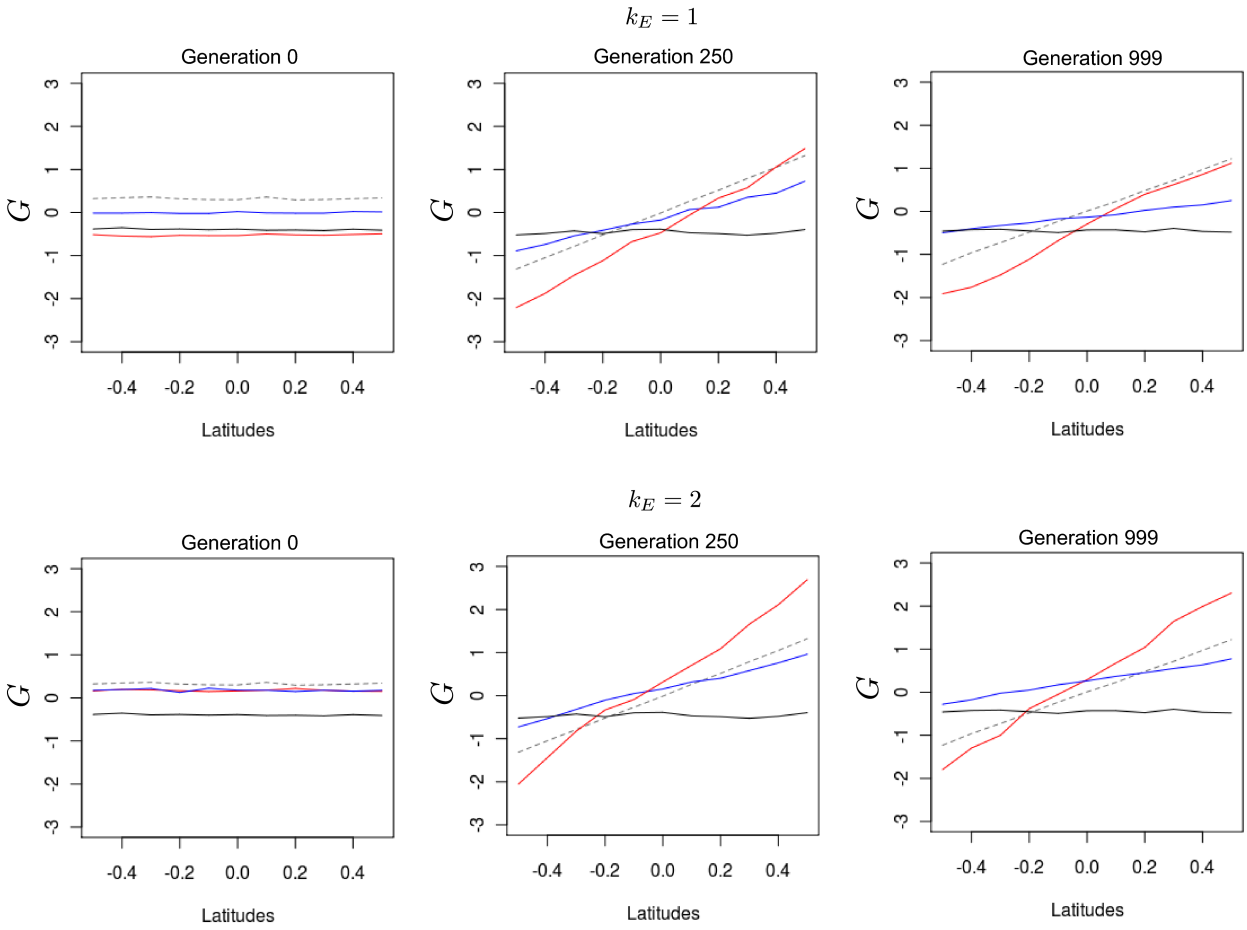
θ_w and V_w were monitored under three different strengths of assortative mating and two levels of environmental cline. All simulations were conducted under the island migration model with moderate gene flow (N_m=5.1). The red line indicates strong assortative mating (ρ=0.8), the blue line moderate assortative mating (ρ=0.3), and the black line random mating (ρ=0). Each line represents the mean of 50 independent replicates for each evolutionary scenario.

Figure 4. Variations in between-population allelic covariation (θ_B), between-population variation (V_B), and timing of bud burst (Q_{ST}).



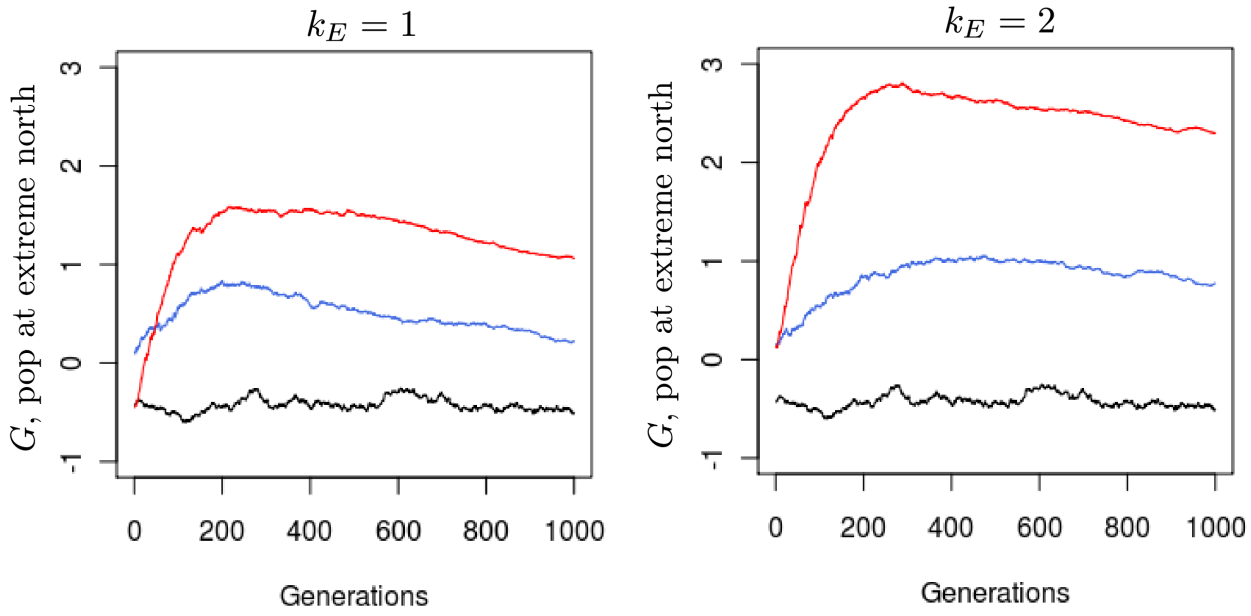
These measurements were monitored under three different strengths of assortative mating and two levels of environmental cline. All simulations were conducted under the island migration model with moderate gene flow (N_m=5.1). The red line indicates strong assortative mating (ρ=0.8), the blue line moderate assortative mating (ρ=0.3), and the black line random mating (ρ=0). Each line represents the mean of 50 independent replicates for each evolutionary scenario.

Figure 5. Variations in mean population genetic values at different latitudes and in different generations.



The value for each latitude is the average of the five mean genetic values for the populations concerned. Latitudinal means were computed and reported for two levels of environmental cline and three different strengths of assortative mating. All simulations were conducted under the island migration model with moderate gene flow (N_m=5.1). The red line indicates strong assortative mating (ρ=0.8), the blue line moderate assortative mating (ρ=0.3), and the black line random mating (ρ=0). The dashed line depicts the mean genetic value obtained under divergent selection modeled with ω^2=5 and $\sigma^2_{Z_{opt}}$=5 , without assortative mating. Each line represents the mean of 50 independent replicates for each evolutionary scenario.

Figure 6. Evolution of the mean genetic value of a population located at the extreme north of the landscape.



The mean genetic value of a population located at latitude +0.5 (dotted circle in figure 1) was monitored under two different levels of environmental cline and three different strengths of assortative mating. All simulations were conducted under the island migration model with moderate gene flow (N_m=5.1). The red line indicates strong assortative mating (ρ=0.8), the blue line moderate assortative mating (ρ=0.3), and the black line random mating (ρ=0). Each line represents the mean of 50 independent replicates for each evolutionary scenario.

Trait differentiation and genetic clines

Because assortative mating had strong consequences on within- and between-population genetic variances, it ultimately contributed to population differentiation of the trait.

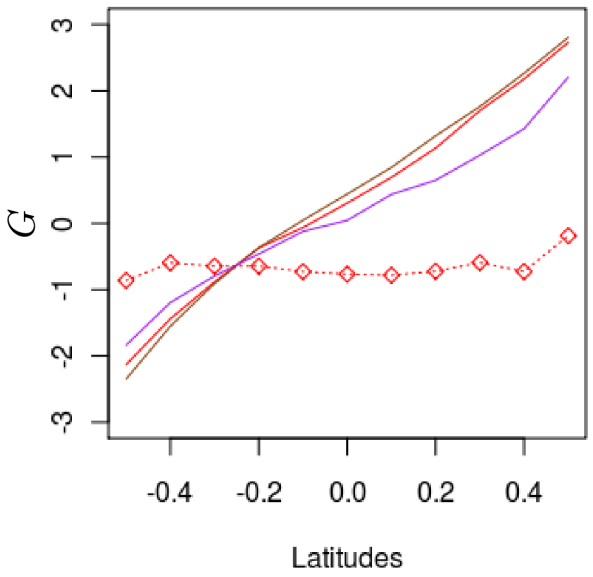
There were striking differences in the levels of differentiation observed under random and assortative mating. Q_{ST} values steadily increased under assortative mating and reached up to 0.7 when k_E=2 . There was only a slight effect of the steepness of the environmental cline on the level of differentiation: Q_{ST}=0.7 when k_E=2 *vs* 0.62 when k_E=1 .

This effect was due to the trade-off between variations in V_B and V_W in equation (1). The steepness of the environmental cline increased V_W (figure 3) and had a decreasing effect on Q_{ST}, but at the same time, it also increased V_B, increasing Q_{ST} (figure 4). As a result, Q_{ST} showed similar values at both levels of environmental cline.

These results suggested that assortative mating differentiated populations and shifted their mean genetic values. We consequently examined the spatial distribution of mean genetic values across the landscape; indeed, a cline of genetic values built up during the early generations following a south-north gradient (figure 5). The steepness of the genetic cline was stronger under assortative mating and under steep environmental clines resulting in a co-gradient variation with the environmental cline. The temporal dynamics of the cline could be illustrated by the changes in the genetic value of the population located at the extreme northern latitude (figure 6). This value reached a peak

between generation 200 and 400, depending on the steepness of the environmental cline and the level of assortative mating. No genetic cline developed under random mating. We also explored the clinal patterns resulting from a more extreme environmental cline, a higher migration rate, and the stepping-stone migration model (figure 7). Surprisingly the resulting genetic cline was less pronounced under $k_E=3$ than under $k_E=2$. When $k_E=3$, the environmental variance among populations was 3-fold larger than the within-phenotypic variance (equation (8)). Consequently, phenological matches between trees from different populations were limited, thus increasing the filtering of incoming genes to proximal populations (figures 9 and 10). Similarly, when the pollen dispersal distance was *a priori* reduced to the most proximal populations, as in the stepping-stone migration model, a very shallow genetic cline built up (figure 7). In this latter case, when $N_m=5.1$, $\rho=0.8$, and $k_E=2$, only populations at extreme latitudes became genetically differentiated. Despite this very shallow cline, Q_{ST} approached 0.45 at generation 1000 under the stepping-stone migration model; under the same simulations parameters, Q_{ST} values reached 0.7 under the island migration model. Finally, when pollen migration rates increased ($N_m=10.2$ *vs* $N_m=5.1$), no significant change was observed in the slopes of the clines. However, additional investigations indicated that lower migration rates decreased the slopes of the genetic clines and induced higher Q_{ST} values, owing to an important drift effect *[Le Corre and Kremer, 2003]* (supplemental figures 2 and 3). Overall large stochastic variations were associated with the genetic parameters that were monitored during the evolutionary scenarios (data not shown). We illustrate these variations only for Q_{ST} and V_W (figure 10). The trend among generations, *i.e.*, the form of the curve, was the same among the replicates.

Figure 7. Variations in mean population genetic values at different latitudes under multiple scenarios.



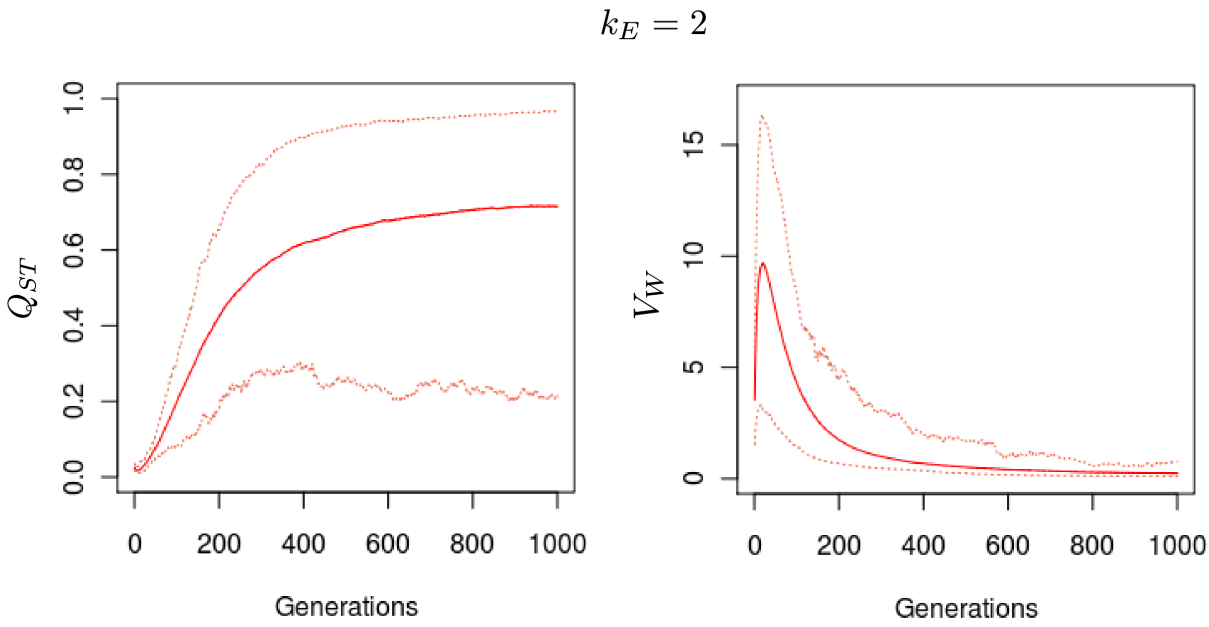
The value for each latitude is the average of the five mean genetic values for the populations concerned at generation 300. All scenarios (except the selection scenario, dashed line) were conducted under strong assortative mating ($\rho=0.8$). Red line: steep environmental cline ($k_E=2$), island migration model, moderate gene flow ($N_m=5.1$). Purple line: very steep environmental cline ($k_E=3$), island migration model, moderate gene flow ($N_m=5.1$). Brown line: steep environmental cline ($k_E=2$), island migration model, extensive gene flow ($N_m=10.2$). Red line with open circles: steep environmental cline ($k_E=2$), stepping stone migration model, high gene flow ($N_m=5.1$). Dashed line: random mating, divergent selection ($\sigma^2_{Z_{opt}}=1$), strong stabilizing selection ($\omega^2=5$), without assortative mating. Each line represents the mean of 50 independent replicates.

Pollen filtering by assortative mating

We monitored the incoming pollen composition in a population located at the extreme northern latitude. By doing so, we expected to predict the shift in genetic values that contributed to the development of the genetic cline under the island migration model (equation (16)). Figure 9 clearly shows that assortative mating filtered incoming alleles by geographic origin. Very rapidly, there was a preferential screening of incoming alleles from neighboring populations in the case of assortative mating, and the trend was more pronounced when the environmental cline grew steeper. The discrepancy between distant and proximal alleles was more pronounced with strong assortative mating.

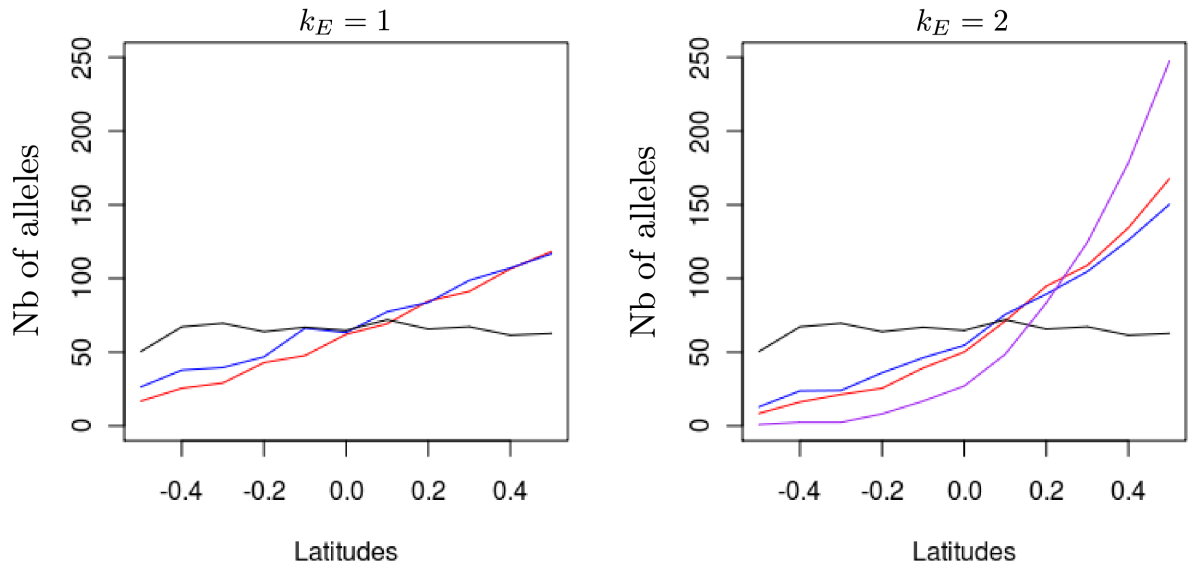
Furthermore, the level of filtering changed over generations. More alleles arrived from distant populations during the first 40 generations, especially when strong assortative mating was occurring (figure 10). These distant alleles would shift the genetic values of populations as predicted by Δ.

Figure 8. Stochastic variations in Q_{ST} and V_W among different simulations within a given scenario.



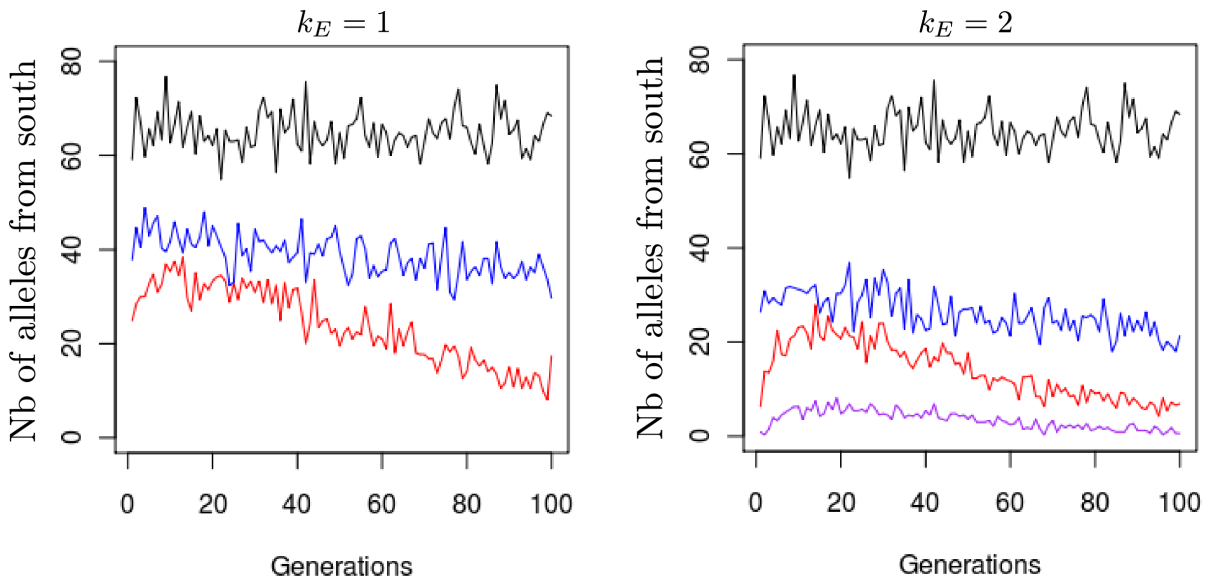
Upper and lower bounds of the 50 simulations conducted per scenario. ρ was set to 0.8 in all cases. k_E is the scaling factor of the environmental cline. Plain lines indicate mean values of the 50 simulations for each scenario and dotted lines represent the two simulations that gave the extreme results.

Figure 9. Amount of immigrant alleles received by a northern population.



Absolute number of immigrant alleles into a population located at the extreme northern latitude (+0.5 dotted circle in figure 1). Numbers on the y-axis are cumulative counts of alleles from generation 16 to 20. Counts of alleles were monitored at three strengths of assortative mating and three levels of the environmental cline. The red line indicates strong assortative mating ($\rho=0.8$), the blue line moderate assortative mating ($\rho=0.3$), the black line random mating ($\rho=0$), and the purple line strong assortative mating under an extreme environmental cline ($\rho=0.8$, $k_E=3$). Lines are mean values of 50 replicates for each evolutionary scenario.

Figure 10. Amount of southern immigrant alleles received by a northern population over generations.



Absolute number of immigrant alleles into a population located at the extreme northern latitude (+0.5 dotted circle in figure 1) and coming from southern latitudes (-0.5 to -0.1). Only gene flow between populations is represented here. Numbers on the y-axis are counts of alleles at a given generation (x-axis). Counts of alleles were monitored at three strengths of assortative mating and three levels of environmental cline. All simulations were conducted under the island migration model with moderate gene flow (N_m=5.1). The red lines indicate strong assortative mating (ρ=0.8), the blue line indicates moderate assortative mating (ρ=0.3), the black line random mating (ρ=0), and the purple line strong assortative mating under an extreme environmental cline (ρ=0.8 , k_E=3). Lines are mean values of 50 replicates for each evolutionary scenario.

Discussion

Our simulations demonstrated that genetic clines could be established in the absence of divergent selection. We showed that the combination of assortative mating and pre-existing environmental clines resulted in population genetic differentiation along the environmental cline. We also confirmed that assortative mating increased the within-population genetic variances in the early stages of the evolutionary scenarios. However, assortative mating was also responsible for the severe decline in genetic variation in later generations.

These patterns resulted in a positive covariance between genetic and environmental population values and corresponded to what has been called co-gradient variation *[Conover, 1995][Conover et al., 2009]*. We discuss here how such covariations may build up under assortative mating in the case of phenological traits in trees. Given the pre-existence of environmental clines, genetic clines are generated by the combined effects of assortative mating and gene flow. In particular, we examine how the interplay between assortative mating and gene flow will actually produce the genetic cline we observed. According to equation (15), the larger the physical distance between the mates associated by gene flow, the more different their genetic values. As a consequence, a larger shift in the mean genetic value should be expected at extreme latitudes in our grid settings (figure 1). In what follows, we illustrate this trend by providing values for the shift Δ obtained at the extreme northern latitude under the strongest assortative mating intensity and across the steepest environmental cline.

We can subdivide the evolutionary process into three main phases, illustrated in figures 4 to 8.

(1) In the very early generations (0–5), the mean genetic value is 0 for all populations, there is no within-population allelic covariance, and alleles are randomly spread over the landscape. During this period, assortative mating will generate phenotypes with extreme genetic values in each population. Hence the genetic variance within populations increases as predicted by previous analytical models *[Fox, 2003][Hayashi, 1998]* and numerical simulations *[Rosvall and Mullin, 2003][Devaux and Lande, 2008][Jorjani et al., 1997 (2)]*.

Gene flow during the early generations preferentially imports alleles from neighboring populations (figure 9), owing to the fact that populations at this stage are genetically undifferentiated over the whole grid and parents exhibiting similar phenotypes are more likely to be in neighboring populations.

As a result, the shift Δ remains limited: 0.0798 at the allelic level for northern populations.

(2) From generation 5 to about 30, because the increase in within-population genetic variance has now produced phenotypes with more extreme values, gene flow tends now to import alleles from more distant populations (figure 10). The fraction of imported alleles enriches the population gene pool and further facilitates an increase in genetic covariances θ_W. The genetic variance between populations continues to increase steadily. During the second phase, the Δ value tends to be larger (0.14) as a result of more divergent alleles imported by distant gene flow. A similar effect that symmetrically decreases the Δ value of incoming gene flow within southern populations is expected to take place at the same time. As a result, the mean genetic values of the population shift strongly, leading to the progressive formation of the genetic cline.

(3) After generation 30, most of the alleles have been spatially redistributed by gene flow constrained by assortative mating at the landscape level. Allelic covariations within populations have been exhausted and the genetic variance has now reached its maximum. Assortative mating within populations tends now to become a selective factor favoring phenotypes following the shift of the mean genetic values. Furthermore, gene flow again becomes strongly restricted to neighboring populations that share fewer divergent alleles than distant populations. Restricted gene flow therefore reinforces the decrease in the genetic variance. Overall, phase 3 is characterized by a

continuous decrease in genetic variance and the reaching of an asymptotic mean genetic value in populations; the genetic cline is establishing. We further advocate that restricted gene flow, together with within-population assortative mating, now constrains effective population sizes, accelerating the decrease in genetic variance due to drift. A similar decrease was observed by Devaux and Lande *[Devaux and Lande, 2008]* in a single population, despite a high mutation rate. Jorjani *et al*. also noticed a decreasing effect of negative assortative mating on the evolution of the genetic variance within a single population *[Jorjani et al., 1997 (2)]*.

These three phases were observed for all of the simulation settings we used. The lengths of the two first phases extended over longer periods, populations differentiated more rapidly, and genetic clines were shaped faster under strong assortative mating. By dissecting the evolutionary process, we showed that the screening of immigrant alleles due to assortative mating triggers shifts in the genetic values of populations (figures 6 and 10). Indeed, when assortative mating allows for long-distance filtered pollen flow, the shifts in the genetic values of recipient populations are strongly enhanced. Because moderate assortative mating generates less extreme genotypes over generations, distant gene flow is promoted less and the mean expected shift in the mean genetic values of populations remains limited. Consequently, under moderate assortative mating, the final steepness of genetic clines is less dependent on the steepness of environmental clines (figures 5 and 6).

Increasing the slope of the environmental cline generated more genetic variance and higher genetic differentiation as well. According to equation (15), each generation steeper environmental clines increase the expected divergence between mates from distinct populations. However, the divergence is constrained by the necessary overlap of parental flowering times. If long distance pollen flow is restricted by large phenological differences among populations, then assortative mating will favor matings between proximal populations, and the shift in genetic values will be limited. In our simulations, the latter case occurred with very large k_E values ($k_E=3$).

A similar outcome was observed under the stepping-stone migration model. In this case, populations do not differentiate except at the northern and southern edges of the landscape (figure 7). This result is only partly explained by the absence of distant gene flow. Indeed, according to the expression of Δ and considering the features of the stepping-stone migration model, limited Δ values are expected owing to pollen flow from adjacent latitudes. However, incoming alleles from neighboring northern populations balance with incoming alleles from neighboring southern populations. As a consequence, the shift induced within populations by southern gene flow is systematically canceled by the one caused by northern flow, resulting in a null contribution to the Δ values. Finally, because under the stepping-stone migration model, incoming gene flow is latitudinally unbalanced at the northern and southern margins of the grid, the genetic values of populations can be shifted by assortative mating at these latitudes. These results suggest that the spatial configuration of the populations in combination with the migration model may also contribute to the building of the genetic cline. Any combination that increases an asymmetry in gene flow between northern and southern populations will enhance the genetic cline, while symmetry will tend to even out the effects of northern and southern gene flow.

To summarize, the construction of a genetic cline as a result of the combined effects of gene flow and assortative mating can only be met under certain circumstances when there is a balance between the intra-population and between-population phenotypic variance (k_E varying between 1 and 3), when long distance pollen flow is possible, and when the patterns of incoming pollen flow at population level are unbalanced regarding the environmental cline. Interestingly these criteria are met under realistic situations. Taking oaks as an example, flushing dates may vary over 5 weeks from southwestern to central France *[Vitasse et al., 2011],* while the same range of variation may be observed between early and late flushing trees in a given forest stand. Viable pollen has also been shown to be dispersed over such distances *[Schueler2005]*.

Conclusion

Our simulations showed that interaction between assortative mating and gene flow across environmental clines may shape the genetic variability of phenologically-related traits and induce co-gradient variation without any divergent selection. We also demonstrated that the extent of genetic variability resulting from assortative mating was related to the patterns of incoming pollen flow at the population level. Because phenotypic clines have been very widely reported in forest trees *[Menzel et al., 2006][Vitasse et al., 2009][Wuelish et al., 1995]*, we suspect that assortative mating and gene flow could actually be responsible for the co-gradient variation observed in some species in common garden experiments *[Alberto et al., 2011][Vitasse, 2009 (CJFR)]*. However, most tree species actually exhibit counter-gradient variation *[Wright, 1976][Morgenstern, 1996]*, suggesting that other evolutionary forces, such as divergent selection, actually counteract the combined effects of assortative mating and gene flow. In a subsequent paper, we will explore how selection interacts with assortative mating and gene flow to generate counter-gradient variation. Finally, our simulations also indicated that very large levels of genetic variation should also be expected within populations, generated by genetic covariances in allelic effects due to assortative mating as predicted by other theories or simulations *[Fox, 2003][Fisher, 1918*]*[Devaux and Lande, 2008]*. Experimental data from progeny tests of forest trees indeed show that heritability values of phenologically-related traits can exceed 0.5, much larger than other phenotypic traits generally assessed in experimental plantations *[Cornelius, 1994]*. Furthermore, our simulations predict that the steep increase in genetic variation will be temporary and will be followed by a rapid decrease. Once all covariation has been exhausted, assortative mating will act as a selective force by constraining the synchronicity of male and female flowering periods. Given the large genetic variation still existing in extant forest stands, we suspect that the time of decrease has not yet been reached in natural populations, owing to the long generation times of trees. Finally, our simulations should be prolonged under more realistic ecological settings, including different patterns of gene flow and selection on multiple traits.

Acknowledgements

This research was supported by the EU project MOTIVE (project number: 226544). We acknowledge the help of Valérie Le Corre and Frédéric Raspail during the adaptation of Metapop to the study of phenological traits. We thank two anonymous reviewers for constructive comments.

References

Alberto F, Bouer L, Louvet JM, Lamy JB, Delzon S, Kremer A: **Adaptive responses for seed and leaf phenology in natural populations of sessile oak along an altitudinal gradient.** *J Evol Biol* 2011, 24(7):1442:1454, [http://dx.doi.org/10.1111/j.1420-9101.2011.02277.x]

Asch MV, Tienderen P, Holleman L, Visser M: **Predicting adaptation of phenology in response to climate change, an insect herbivore example**. *Global Change Biology* 2007, 13:1596:1604

Bennie J, Kubin E, Wiltshire A, Huntley B, Baxter R: **Predicting spatial and temporal patterns of bud-burst and spring frost risk in north-west Europe: the implications of local adaptation to climate**. *Global change biology* 2010, 16:1503:1514

Bertin RI: **Plant phenology and distribution in relation to recent climate change**. *J. Torrey Bot. Soc.* 2008, 135:126:146

Caisse M, Antonovics J: **Evolution in closely adjacent plant populations. IX. Evolution of reproductive isolation in clinal populations**. *Heredity* 1978, 40:371:384

Chmura D, Rozkowski R: **Variability of beech provenances in spring and autumn phenology**. *Silvae Genetica* 2002, 51:123:127

Chuine I, Cour P: **Climatic determinants of budburst seasonality in four temperate-zone tree species**, *New Phytol.* 1999, 143:339:349

Conover DO, Duy TA, Hice LA: **The covariance between genetic and environmental inuences across ecological gradients: reassessing the evolutionary signicance of counter-gradient and co-gradient variation**. *Ann N Y Acad Sc*i 2009, 1168:100:129, [http://dx.doi.org/10.1111/j.1749-6632.2009.04575.x]

Conover DO, Schultz ET: **Phenotypic similarity and the evolutionary significance of counter-gradient variation.** *Tree* 1995, 10:248:252

Cornelius J: **Heritabilities and additive genetic coecients of variation in forest trees**. *Canadian Journal of Forest Research,* 1994*,* 24*:*372-379

Derory J, Scotti-Saintagne C, Bertocchi E, Dantec LL, Graignic N, Jaures A, Casasoli M, Chancerel E, Bodenes C, Alberto F, Kremer A: **Contrasting relationships between the diversity of candidate genes and ariation of bud burst in natural and segregating populations of European oaks**. *Heredity* 2010, 104:438:448

Devaux C, Lande R: **Incipient allochronic speciation due to non-selective assortative mating by owering time, mutation and genetic drift**. *Proc Biol Sci* 2008, 275(1652):2723:2732, [http://dx.doi.org/10. 1098/rspb.2008.0882]

Ekberg I, Eriksson G, Namkoong G, Nilsson C, Norell L: **Genetic correlations for growth rhythm and growth capacity at ages 3-8 years in provenance hybrids of picea-abies**. *Scandinavian journal of forest research 1994, 9:25:33*

Fisher RA: **The correlation between relatives on the supposition of Mendelian inheritance**. *Trans. Roz. Soc. Edinb.* 1918, 52:339:433

Fox GA: **Assortative mating and plant phenology: evolutionnary and pratical consequences**. *Evolutionnary Ecology Research* 2003, 5:1:18

Franjic J, Sever K, Bogdan S, Skvorc Z, Krstonosic D, Aleskovic I: **Phenological asynchronization as a restrictive factor of ecient pollination in clonal seed orchards of Pedunculate Oak (Quercus robur L.)**. *Croat. j. for. eng.* 2011, 31:141:156

Ghelardini L, Santini A: **Avoidance by early ushing: a new perspective on Dutch elm disease research**. *iForest* 2009, 2:143:153

Gianola D: **Assortative mating and the genetic correlation**. *Theor Appl Genet* 1982, 62:225:231

Hayashi T: **Genetic variance under assortative mating in the innitesimal model**. *Genes Genet. Syst.* 1998, 73:397:405

Howe G, Aitken S, Neale D, Jermstad K, Wheeler N, Chen T: **From genotype to phenotype: unravellingthe complexities of cold adaptation in forest trees**. *Can. J. Bot.* 2003, 81:1247:1266

Jensen JS, Hansen J: **Geographical variation in phenology of Quercus petraea (Matt.) Liebl and Quercus robur L. oak grown in a greenhouse**. *Scandinavian journal of forest research* 2008, 23:179:188

Jorjani H, Engström G, Strandberg E, Liljedahl LE: **Genetic studies of assortative mating - a simulation study. II. Assortative mating in unselected populations**. *Acta Agric. Scand.* 1997, 47:74:81

Jorjani H, Engstrøom G, Strandberg E, Liljedahl LE: **Genetic studies of assortative mating - a simulation study. III. Assortative mating in selected populations**. *Acta Agric. Scand.* 1997, 47:129:137.

Kirkpatrick M: **Reinforcement and divergence under assortative mating**. *Proc Biol Sci* 2000, 267(1453):1649:1655, [http://dx.doi.org/10.1098/rspb.2000.1191].

Kremer A, Le Corre V, Petit R, Ducousso A: **Molecular Approaches in Natural Resource Conservation and Management**. *Cambridge University Press* 2010

Kremer A, Le Corre V: **Decoupling of differentiation between traits and their underlying genes in response to divergent selection**. *Heredity* 2011, doi: 10.1038/hdy.2011.81

Lande R: **The influence of the mating system on the maintenance of genetic variability in polygenic characters**. *Genetics* 1977, 86:485:498

Le Corre V, Kremer A: **Genetic Variability at Neutral Markers, Quantitative Trait Loci and Trait in a Subdivided Population Under Selection**. *Genetics* 2003, 164:1205:1219

Le Corre V, Kremer A: **The genetic differentiation at quantitative trait loci under local adaptation**. *Molecular Ecology* 2011, doi: 10.1111/j.1365-294X.2012.05479.x

Le Corre V, Machon N, Petit RJ, Kremer A: **Colonization with ong-distance seed dispersal and genetic structure of maternally inherited genes in forest trees: a simulation study**. *Genet. Res.* 1997, 69:117:125

Machon N, Bardin P, Mazer S, Moret J, Godelle B, Austerlitz F: **Relationship between genetic structure and seed and pollen dispersal in the endangered orchid Spiranthes spiralis**. *New Phytologist* 2003,157:677:687

Menzel A, Sparks TH, Estrella N, Koch E, Aasa A, Ahas R, Alm-Kubler K, Bissolli P: **European phenological response to climate change matches the warming pattern**. *Global Change Biology* 2006, 12:1969:1976

Morgenstern EK: **Geographic variation in forest trees: genetic basis and application of knowledge in silviculture**. *UBC press* 1996

Nordli O, Wielgolaski F, Bakken A, Hjeltnes S, Mage F, Sivle A, Skre O: **Regional trends for bud burst and owering of woody plants in Norway as related to climate change**. *International Journal of Biometeorology* 2008, 52:625:639

Rosvall O, Mullin TJ: **Positive assortative mating with selection restrictions on group coancestry enhances gain while conserving genetic diversity in long-term forest tree breeding**. *Theor Appl Genet 2003*, 107(4):629:642, [http://dx.doi.org/10.1007/s00122-003-1318-9]

Saintagne C, Bodenes C, Barreneche T, Pot D, Plomion C, Kremer A: **Distribution of genomic regions differentiating oak species assessed by QTL detection**. *Heredity* 2004, 92:20:30

Schueler S, Schlünzen K, Scholz F: **Viability and sensitivity of oak pollen and its implications for pollen-mediated gene flow**. *Trends Ecol. Evol* 2005, 19:154:161

Turelli M: **Heritable genetic variation via mutation-selection balance: Lerch's zeta meets the abdominal bristle**. *Theor. Popul. Biol.* 1984, 25:138:193

Vitasse Y, Delzon S, Bresson C, Michalet R, Kremer A: **Altitudinal differentiation in growth and phenology among populations of temperate tree species in a common garden**. *Canadian Journal of Forest Research* 2009, 39:1259:1269

Vitasse Y, Delzon S, Dufrene E, Pontailler JY, Louvet JM, Kremer A, Michalet R: **Leaf phenology sensitivity to temperature in European trees: Do within-species populations exhibit similar responses?** *Agricultural and forest meteorology* 2009, 149:735:744

Vitasse Y, Francois C, Delpierre N, Dufrêne E, Kremer A, Chuinee I, Delzon S: **Assessing the eects of climate change on the phenology of European temperate trees**. *Agricultural and Forest Meteorology 2011*, 151 (7):969:980

Von Wuehlisch G, Krusche D, Muhs H: V**ariation in temperature sum requirement for ushing of beech provenances**. *Sylvae Genetica* 1995, 44:343:346

Weis AE, Kossler TM: **Genetic variation in owering time induces phenological assortative mating: quantitative genetic methods applied to Brassica rapa**. *American Journal of Botany* 2004, 91:825:836

Weis AE: **Direct and indirect assortative mating: a multivariate approach to plant owering schedules**. J Evol Biol 2005, 18(3):536:546, [http://dx.doi.org/10.1111/j.1420-9101.2005.00891.x]

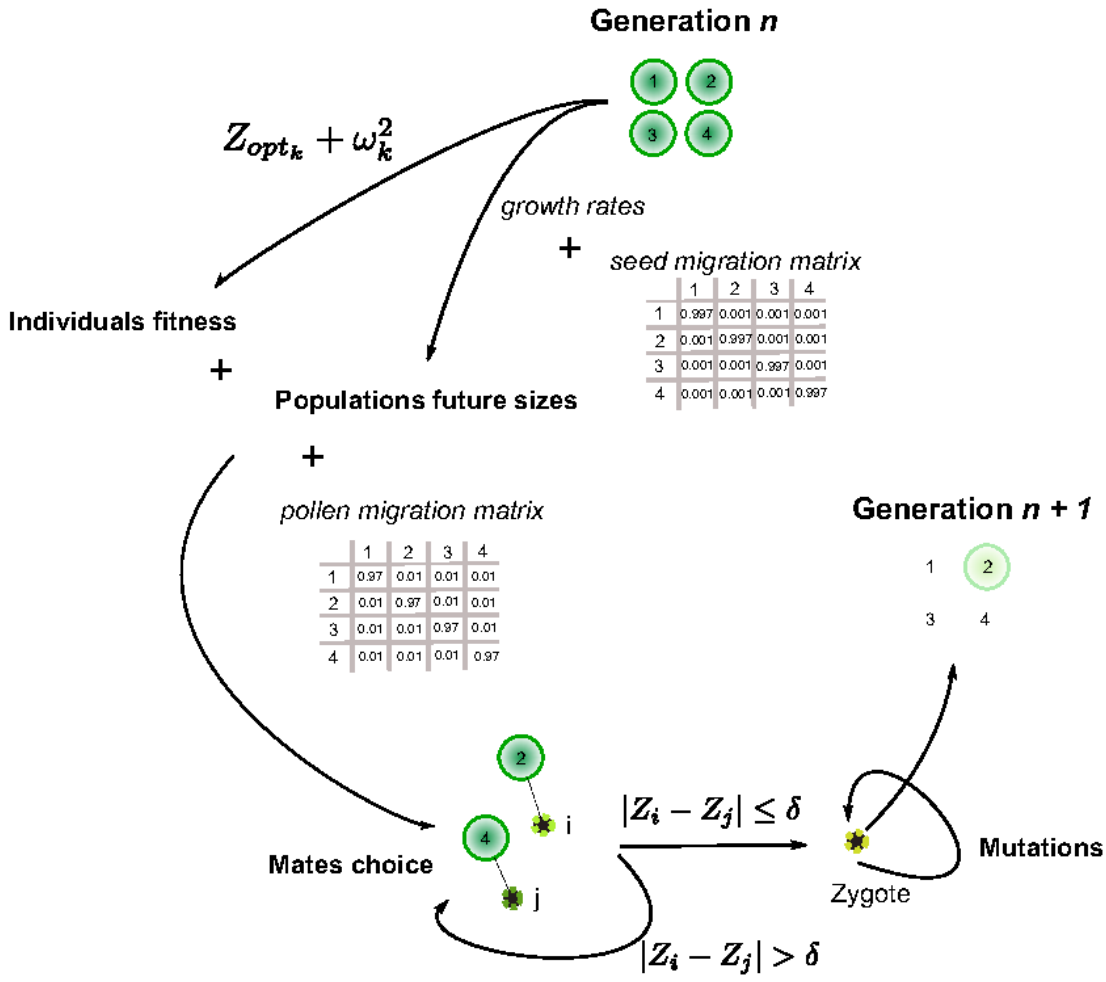
Worrall J: **Temperature - bud-burst relationships in amabilis and subalpine for provenance tests replicated at dierent elevations**. *Silvae genetica* 1983, 32:203:209

Wright JW: **Introduction to forest genetics**. *Academic Press, INC.* 1976

Wright S: Systems of mating. III. **Assortative mating based on somatic resemblance**. *Genetics* 1921, 6:144:161

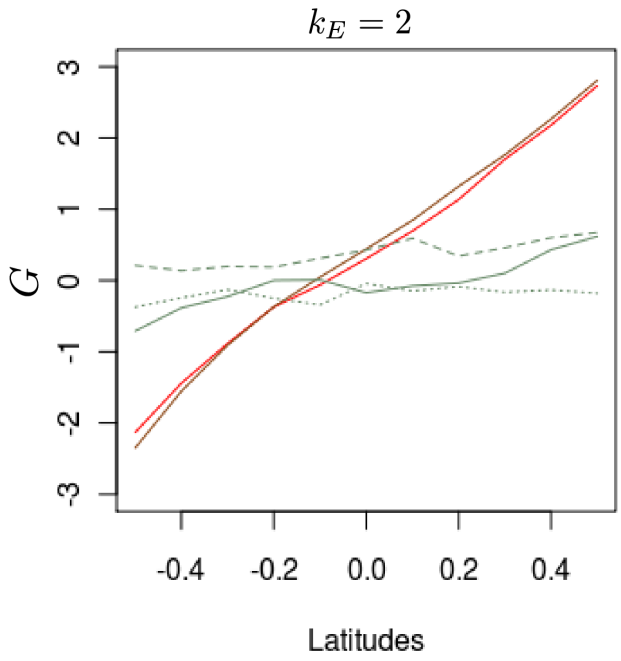
Supplemental figures

Supplemental figure 1. Summary of the evolutionary processes within a generation.



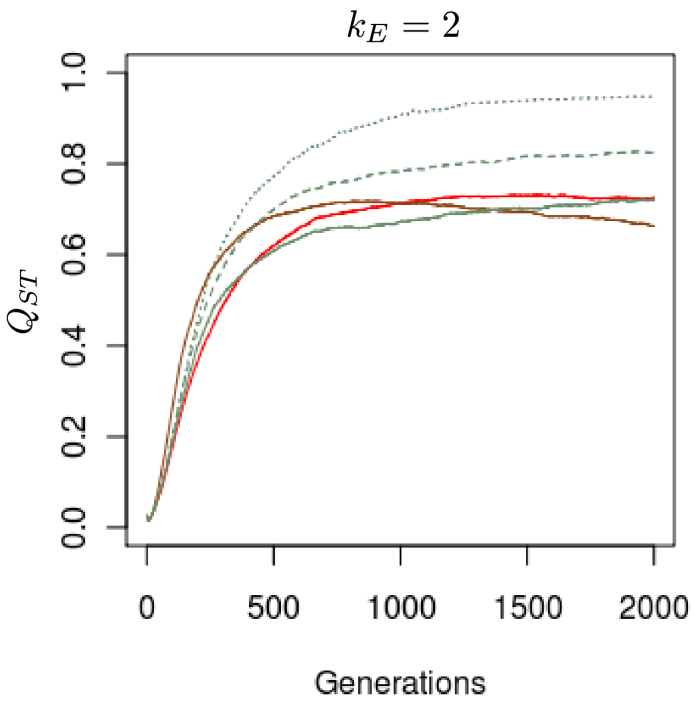
Fitness values and sizes of populations are first computed according to selection settings, demographic settings, and the seed migration matrix. Reproduction takes place between mates paired according to fitness, seed migration settings, and pollen migration settings. Assortative mating may bear additional iterations for the choice of male and female parents because mates must share close phenotypic values. Mutations may occur.

Supplemental figure 2. Variations in mean population genetic values at different latitudes under a range of migration rates.



The value for each latitude is the average of the five mean genetic values for the populations concerned at generation 300. All scenarios were conducted under strong assortative mating ($\rho=0.8$), island migration model and steep environmental cline ($k_E=2$). Brown line: $N_m=10.2$, red line: $N_m=5.1$, green line: $N_m=1$, green dashed line: $N_m=0.5$ and green dotted line: $N_m=0.1$. Each line represents the mean of 50 independent replicates for each evolutionary scenario.

Supplemental figure 3. Q_{ST} **values after 1000 generations under a range of migration rates.**



All simulations were conducted under under strong assortative mating (ρ=0.8), island migration model and steep environmental cline (k_E=2). Brown line: N_m=10.2 , red line: N_m=5.1 , green line: N_m=1 , green dashed line: N_m=0.5 and green dotted line: N_m=0.1 . Each line represents the mean of 50 independent replicates for each evolutionary scenario.

Chapter 3

Evolutionary responses of tree phenology to the combined effects of assortative mating, gene flow and divergent selection

Jean-Paul Soularue and Antoine Kremer*

INRA, UMR 1202 BIOGECO, Cestas F-33610, France

Univ. Bordeaux, BIOGECO, UMR 1202, Talence F-33400, France

***Corresponding author**: antoine.kremer@pierroton.inra.fr, phone: +33 5 57 12 28 32, fax:+33 5 57 12 28 81

Abstract

The timing of bud burst (TBB) in temperate trees is a key adaptive trait, the expression of which is triggered by temperature gradients across the landscape. TBB is strongly correlated with flowering time and is therefore probably mediated by assortative mating. We derived theoretical predictions and realized numerical simulations of evolutionary changes in TBB in response to divergent selection and gene flow in a meta-population. We showed that the combination of the environmental gradient of TBB and assortative mating creates contrasting genetic clines, depending on the direction of divergent selection. If divergent selection acts in the same direction as the environmental gradient (co-gradient settings), genetic clines are established and inflated by assortative mating. Conversely, under divergent selection of the same strength but acting in the opposite direction (counter-gradient selection), genetic clines are slightly constrained and have shallower slopes. We explored the consequences of these dynamics for population maladaptation, by monitoring pollen swamping. Depending on the direction of divergent selection with respect to the environmental gradient, pollen filtering due to assortative mating either facilitates or impedes adaptation in peripheral populations.

Keywords: phenology, assortative mating, adaptation, gene flow.

Introduction

Ongoing climate change and climatic predictions have sparked serious concerns about the persistence of tree species in many temperate and boreal areas. As forests play a crucial role in human societies through the maintenance of sustainable ecosystems services, considerable effort is currently being devoted to assessments of their possible evolutionary responses. A number of predictions based on climatic change and greenhouse gas emission models have forecasted major shifts in the bioclimatic envelopes of temperate tree species *[Thuiller et al., 2003][Cheaib et al., 2012]*, however these predictions did not take genetic variation into account. Common garden experiments have highlighted high levels of genetic diversity within and among tree populations *[Wright, 1976][Hamrick et al., 1992]*, reporting strong clinal genetic variations along environmental gradients *[Savolainen et al. 2007][Mimura and Aitken, 2007][Alberto et al., 2013]* which suggests strong adaptive responses to past environmental changes. Genetic clines vary both

in magnitude and orientation, and may vary either with or against environmental gradients *[Conover and Schultz, 1995][Conover et al., 2009]*. It is now necessary to determine which combinations of evolutionary factors were responsible for generating these differentiation patterns when trying to predict how a species might evolve in response to current and future rapid changes in climate.

Timing of bud burst (TBB) is a key adaptive trait of trees in the context of climate change, for three main reasons: (1) TBB varies with temperature, (2) TBB is related to the fitness of trees owing to its influence on growing season *[Chuine and Beaubien, 2001]* and (3) TBB is strongly correlated with flowering time *[Franjic et al., 2010]* suggesting that this trait may be subject to assortative mating. Clinal patterns of differentiation along environmental gradients have mostly been interpreted as resulting from divergent selection, but we have shown that the interaction between assortative mating and gene flow may cause TBB to vary with the environmental gradient (variation in the same direction, co-gradient variation) without divergent selection *[Soularue and Kremer, 2012]* which confirms earlier results reported by Stam (1983) *[Stam, 1983]*. We showed that the synchronicity constraint imposed by assortative mating reversed the mixing effect of pollen flow among populations, generating a divergent effect. Indeed, across environmental gradients, gene flow over large distances, with filtering by assortative mating, leads to the import of alleles with genetic values very different from the mean genetic values of the recipient populations. Larger distances between the two mating parents are associated with larger differences in their genetic values, due to the environmental gradient. Thus, the net result of intra- and inter-population matings is a shift in the genetic values of the populations along the environmental gradient, creating a co-gradient genetic cline *[Soularue and Kremer, 2012]*. Additionally, we showed that the scale of the final differentiation was related to the slope of the environmental cline and the patterns of pollen dispersal. Here, we extended this approach to the context of divergent selection. The combined effects of assortative mating and natural selection in plant species have been investigated, theoretically and experimentally, several times, but usually in single populations. Analytical predictions and numerical simulations have mostly predicted a stronger response to selection in the presence of positive assortative mating than in conditions in which selection is applied to randomly mated populations, through an increase in genetic variation *[Gianola, 1982][Fox, 2003][Jorjani, 1997]*. A few studies have also indicated that negative assortative mating may soften the impact of selection *[De Lange, 1974][Jorjani, 1997]*. All theoretical predictions and observations to date have considered single isolated populations and none has yet explored the balance between assortative mating and the various patterns of selection occurring in populations positioned along environmental gradients and interconnected by gene flow. As gene flow may either prevent adaptation *[Kirkpatrick and Barton, 1997][Yeaman and Guillaume, 2009]* or enhance evolutionary change *[Bridle et al., 2009][Kremer et al., 2012]*, we also showed that the degree of TBB differentiation generated by assortative mating in the absence of selection depends strongly on the pollen dispersal model applied *[Soularue and Kremer, 2012]*. We thereby also investigated here the interaction between gene flow and assortative mating in the presence of selection.
In summary, we investigated the interactions between three major evolutionary drivers of the timing of bud burst: divergent selection, assortative mating and gene flow. It is essential to consider the directional nature of these drivers when trying to understand their joint effects. Indeed, the environmental gradient triggers assortative mating for TBB by constraining preferential matings between neighboring populations in a given direction. Similarly, divergent selection may also act in the same direction as the environmental gradient or in the opposite direction, generating co- or counter-gradient variation. Finally, gene flow also has a directional component: in the case of latitudinal gradients, populations from northern latitudes can for example only receive pollen from more southerly latitudes. Here, we compared the evolutionary response of TBB for various evolutionary scenarios within numerical simulations. Our main objectives was to monitor the combined impact of the various forces on the evolutionary response and to determine which conditions and scenarios are associated with an enhancement or restriction of adaptation. We

predict that assortative mating may either amplify or constrain the response to selection, depending on the combination of the directions and the extent of the environmental and selection gradients considered.

Methods

Evolutionary change in populations under hypotheses of assortative mating, divergent selection and gene flow

We considered a set of populations connected by gene flow, with each population undergoing stabilizing selection. Within each population, the fitness of an individual k was given by the Gaussian function *[Turelli, 1984]*:

$$W(Z_k)=\exp[\frac{-(Z_k-Z_{opt})}{2\omega^2}]$$

where Z_k is the phenotypic value of individual k, Z_{opt} is the optimum value of the given population and ω^2 is the strength of stabilizing selection. We considered here that divergent selection in the landscape is driving populations toward different Z_{opt} values and assumed that the strength of within-population selection is identical in different populations. We assumed that gene flow occurs principally via pollen. We were interested in the evolutionary change, over successive generations, within a given population, as a result of the various evolutionary forces. On the basis of standard quantitative genetics principles, we subdivided the within-population phenotypic value of each tree in the population of interest into a genetic and a random environmental component: $Z=g+\varepsilon$, where g is the sum of the additive effects of the alleles carried by the tree and contributing to the trait. We assumed that there is no dominance or epistasis. The mean genetic value of the population at generation *t,* before selection, is G_t, whereas that after selection is G_t^s. As ε has a mean value of 0 in each population, $G_t=Z_t$. The mean genetic value of a population at the generation immediately following syngamy and preceding selection can thus be written as:

$$G_{t+1}=(1-m_p)(\frac{1}{2}G_t^{\female s}+\frac{1}{2}G_t^{\male s})+m_p(\frac{1}{2}G_t^{\female s}+\frac{1}{2}G_t^{s^*}) \qquad (1)$$

where m_p is the rate of pollen immigration, $G_t^{\female s}$ and $G_t^{\male s}$ are the mean genetic values for the female and male parents within the population after selection, and $G_t^{s^*}$ is the mean genetic value of male parents outside the population after selection. We assume here that $G_t^{\female s}=G_t^{\male s}=G_t^s$. In addition, $(1-m_p)(\frac{1}{2}G_t^{\female s}+\frac{1}{2}G_t^{\male s})$ represents the proportion of the genetic value owing to intra-population matings and $m_p(\frac{1}{2}G_t^{\female s}+\frac{1}{2}G_t^{s^*})$ represents the proportion of the genetic value owing to inter-population matings.

We accounted for assortative crosses involving parents from different populations, by also defining the overall phenotypic value of a tree as $z'=g+E+\varepsilon$, making it possible to compare phenotypic values across populations. E is a macro-environmental component of identical magnitude for all trees in a given population and ϵ a micro-environmental component assigned independently to each tree. Within a population, the mean phenotypic value of the male parents corresponding to the immigrant alleles can be expressed as

$$Z'^{s^*}_t=G_t^{s^*}+E^*+\epsilon^* \qquad (2)$$

whereas the mean phenotypic value of the female parents is

$$Z'^{\female s}_t=G_t^{\female s}+E+\epsilon \qquad (3)$$

where E and E^* are the macro-environmental components of the phenotypes of the female and male parents while ϵ and ϵ^* stand for the micro-environmental components of their phenotypes, respectively. Assuming full assortative mating, the phenotypic values of both parents should be equal, and the mean genetic value of the external male parents can then be written as

$$G_t^{s^*} = G_t^{\varphi^s} + E - E^* + \epsilon - \epsilon^*$$

Assuming $\epsilon - \epsilon^*$ is negligible compared to $E - E^*$, we have

$$G_t^{s^*} = G_t^{\varphi^s} + E - E^* \qquad (4)$$

From equation (1), we can express the shift Δ_t of the genetic value of a population $G_{t+1} - G_t$ as

$$\Delta_t = (1 - \frac{1}{2} m_p) G_t^s + \frac{1}{2} m_p G_t^{s^*} - G_t \qquad (5)$$

The mean genetic value of the population after selection can be written as *[Bulmer, 1980][Lopez et al., 2008]*:

$$G_t^s = \frac{\sigma_s^2}{\sigma_s^2 + V_w} G_t + \frac{V_w}{\sigma_s^2 + V_w} Z_{opt}$$

Where σ_s^2 is the sum of the variance of the random environmental deviation σ_ε^2 and the selection intensity ω^2, V_w stands for the genetic variance observed within the population of interest and Z_{opt} is its phenotypic optimum. Replacing G_t^s and $G_t^{s^*}$ in equation (5), we obtain:

$$\Delta_t = (1 - \frac{1}{2} m_p)[\frac{\sigma_s^2}{\sigma_s^2 + V_W} G_t + \frac{V_W}{\sigma_s^2 + V_W} Z_{opt}] + \frac{1}{2} m_p [\frac{\sigma_s^2}{\sigma_s^2 + V_W^*} G_t^* + \frac{V_W^*}{\sigma_s^2 + V_W^*} Z_{opt}^*] - G_t$$

If we assume that V_w is the same for all populations, then:

$$\Delta_t = \frac{V_W}{\sigma_s^2 + V_W} (Z_{opt} - G_t) + \frac{1}{2} m_p [\frac{\sigma_s^2}{\sigma_s^2 + V_W} (G_t^* - G_t)] + \frac{1}{2} m_p [\frac{V_W}{\sigma_s^2 + V_W} (Z_{opt}^* - Z_{opt})] \qquad (6)$$

Equation (6) partitions the evolutionary change within populations into three terms: $\Delta_t = A + B + C$. The *A* term accounts for evolutionary change due to stabilizing selection within the population of interest, it corresponds to the analytical prediction of evolutionary change for a single population under stabilizing selection, as suggested in previous studies *[Lande, 1976][Gomulkiewicz, 2009]*. It depends mainly on the strength of selection (σ^2_s) and the degree of maladaptation $Z_{opt} - G_t$. The *B* and *C* terms account for evolutionary change due to cross-pollination with trees from other populations. Since the genetic values of the mating partners are affected by the macro-environmental component under assortative mating, the *B* term expresses indirectly the influence of the environmental gradient. Finally, the *C* term expresses the influence of the selection gradient.
Clearly, the evolutionary shift of populations will depend on the selective pressures, the macro-environmental component of the phenotypic values and the dispersal of pollen. Besides, from equations (2), (3) and (4), varying adaptive responses can be expected in the case of selection and environmental gradients. Assuming selective pressures sufficiently strong to drive the genetic values of individual towards a phenotypic optimum, we can assume that G_t^s and $G_t^{s^*}$ will rapidly approach the Z_{opt} values of populations. Thereby, owing to the opposition between the selection and the environmental gradients, the phenotypic difference between distant individuals $Z'^{\varphi^s}_t - Z'^{s^*}_t$ will be less important in the case of counter-gradient settings, which will promote matings among

distant individuals (see *[Soularue and Kremer, 2012]* for more details about matings between individuals from distinct populations across environmental gradients). Distant pollen flow will thus import alleles with additive contribution strongly diverging from the phenotypic optimum of the recipient populations, which will constrain the adaptive response of populations over the landscape. Conversely, in the case of co-gradient settings, the mean phenotypic values of distant populations will be so important that distant mating will be impossible. As a result of limited distant pollen flow, assortative mating will in this case rapidly inflate the adaptive response of populations to selective pressures as indicated in previous theory *[Fox, 2003]*. Additionally, we expect the extent of the selection and environmental gradients to mediate the effects of assortative mating. But very steep gradient might totally even out the evolutionary impact of assortative mating. Indeed, we pointed out in *[Soularue and Kremer, 2012]* that too important slopes of environmental gradient disconnect the distributions of the phenotypic values of distant populations and even out the divergent effect of assortative mating in absence of selection.
Finally, the evolutionary responses of populations may be amplified by the dispersal of pollen occurring in the direction of the environmental gradient under assortative mating. By contrast, pollen dispersal against the environmental gradient will tend to reduce the genetic cline. Finally, directional pollen flow would be expected to have a significant impact on the evolution of populations in conditions of uniform selection.

Simulations

We compared our expectations with the results of forward and individual-based simulations. To this end, we simulated the evolution of a subdivided population over successive generations with Metapop, a tractable numerical simulating framework allowing the implementation of evolutionary processes. Descriptions of this software have been provided elsewhere *[Austerlitz et al., 2000]*, *[Le Corre and Kremer, 2003]* and *[Kremer and Le Corre, 2011]*. We recently described the way in which the environmental gradient and assortative mating were introduced into the simulations *[Soularue and Kremer, 2012]*.

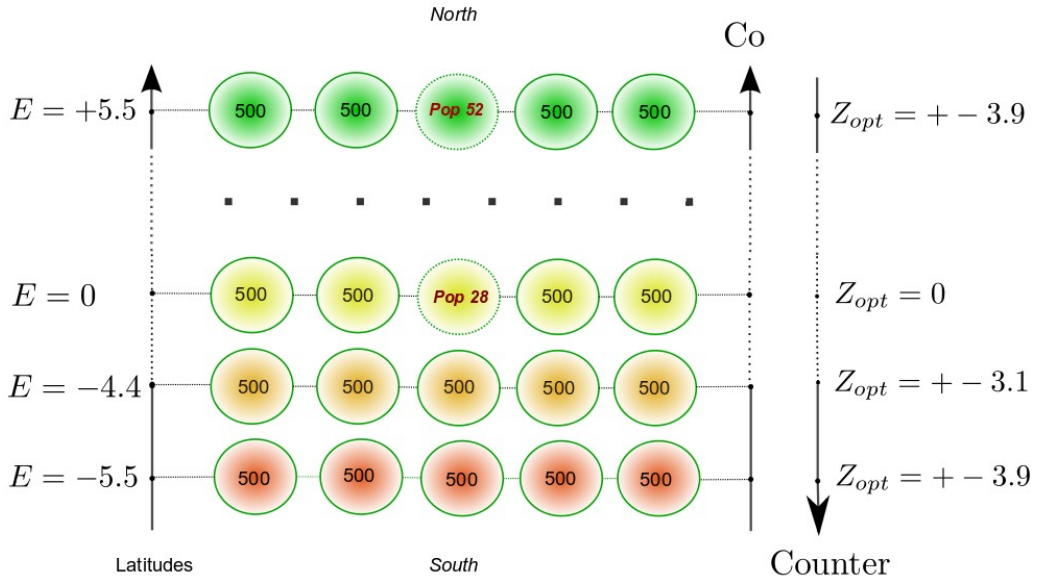
Environmental and genetic gradients
We considered a set of 55 populations spatially distributed over a two-dimensional grid, reflecting real settings observed in natural conditions (figure 1). For the sake of simplicity, we use here “latitude” and “longitude” to label the two dimensions. We assumed that the macro-environmental values are distributed along a latitudinal gradient in the landscape such that all populations at a given latitude have the same E value (see *[Soularue and Kremer, 2012]* for further details). These settings for E values mimic the patterns reported *in situ* for latitudinal gradients of flushing date in trees *[Alberto et al., 2011][Vitasse et al., 2009][Chuine et al., 1999]*. Each population was composed of 500 individuals, this number remaining constant among populations and over successive generations.
Within Metapop, selection is implemented via Turelli's model of stabilizing selection *[Turelli, 1984]*:

$$W(Z_k)=\exp[\frac{-(Z_k-Z_{opt})}{2\omega^2}]$$

$W(Z_k)$ being the fitness value assigned to the phenotype Z_k given ω, the intensity of selection, and Z_{opt}, the optimal trait value in the population of interest. We simulated divergent selection among populations by assigning different Z_{opt} values to populations along a one-dimensional latitudinal gradient. All populations occupying the same latitude were given the same Z_{opt}. We also considered the gradients of E and Z_{opt} values to be running in the same direction (co-gradient settings) or in opposite directions (counter-gradient settings) (figure 1). We also simulated uniform selection, by applying a single Z_{opt} value ($Z_{opt}=0$) to the entire landscape. The environmental cline

was scaled by its slope k_E *[Soularue and Kremer, 2012]*, and the genetic cline by the variance of optimal values σ^2_{Zopt}. We mainly simulated steep environmental clines ($k_E=2$) that provided the most significant results in our earlier paper. In addition, we considered two steepness of selection gradient, $\sigma^2_{Z_{opt}}=6$ and $\sigma^2_{Z_{opt}}=1$, in order to further explore the interplay between the magnitudes of the environmental and selection gradients. The strength of selection was mostly moderate ($\omega^2=50$). Assortative mating was scaled by the correlation of the phenotypic values ρ of mating parents (see *[Soularue and Kremer, 2012]* for further details). In the simulations, k_E, $\sigma^2_{Z_{opt}}$ and ρ were set at generation 0 and kept constant over successive generations. Finally, the starting meta-population was at mutation-migration-drift equilibrium and all populations were considered to have a genetic value of 0 at generation 0.

Figure 1. Spatial settings of populations, environmental and selection gradients.



Fifty-five populations of 500 individuals each were spread homogeneously on a 5 x 11 grid along 11 latitudinal positions. *E* represents the environmental effect at a given latitude (see *[Soularue and Kremer, 2012]* for more details). Co-gradient and counter-gradient divergent selection were considered mostly with $\omega^2=50$ and $\sigma^2_{Z_{opt}}=6$. Some cases were explored with $\omega^2=5$ or $\sigma^2_{Z_{opt}}=1$. Finally, uniform selection was simulated with $\sigma^2_{Z_{opt}}=0$.

Gene flow: directional and long-distance dispersal

Interest in the dispersal of tree pollen by wind has recently increased, with the confirmation that viable pollen can be dispersed over very long distances *[Schueler, 2005][Nathan et al., 2008]* (*[Kremer et al., 2012]* for a review). In many regions, winds tend to blow in one particular direction, due to the rotation of the Earth and pressure gradients (http://www.wcc.nrcs.usda.gov/), with major potential consequences for pollen dispersal, as observed in a previous experimental study *[Bohrerova et al., 2009]*. We included in the simulations the preferential dispersal of pollen in one particular direction, such that most dispersal among population was latitudinal and occurs either towards north or towards south. Only a small proportion dispersed longitudinally. Moreover, latitudinal pollen dispersal occurs over long distances, with realistic migration rates, following a fat-tailed leptokurtic distribution *[Austerlitz et al., 2004]*. Under the long-distance directional migration model, the exchange of pollen among populations was scaled by calculating the ratio of latitudinal to longitudinal dispersal. Ninety-five percent of the pollen shed from each population was

disseminated latitudinally, with migration rates drawn at random from the Weibull distribution with the most leptokurtic shape *[Austerlitz et al., 2004]*, with parameters (3,1). The remaining 5% of the pollen was disseminated longitudinally, with similar proportions reaching the two closest neighboring populations. Below, long-distance dispersal models oriented towards the north are referred to as LDN, whereas those oriented towards the south are referred to as LDS. In addition to the unidirectional pollen migration model, we also considered pollen migration according to an island model (IMM). Seed migration was systematically considered to follow the island migration model. We considered a single level of gene flow $Nm = 10.2$.

We also assessed incoming pollen flow, with filtering due to the combined effects of assortative mating and pollen dispersal, *a posteriori*. We monitored exhaustively, over a number of generations, the origin of the external pollen reaching the female parents of a population located at the northern edge of the landscape. We then determined numerically the distribution of the origin of the alleles received by the population of interest. All simulation settings are summarized in table 1.

Table 1. Initial simulation settings.

heritability	0,83
selfing rate	0,02
nb. of populations	55
nb. of individual per population	500
pollen migration models	IMM, LDN, LDS
seed migration model	IMM
level of gene flow N_m	10,2
nb. of QTL n	10
mutation rate μ	10^{-5}
nb. of latitude levels Y	11
interval of latitudes Y	[-0.5, +0.5]
steepness of environmental cline E scaled by k_E	2
variance of Z_{opt}	0, 1, 6
intensity of stabilizing selection ω^2	5, 50
assortative mating ρ	0, 0.8

Results

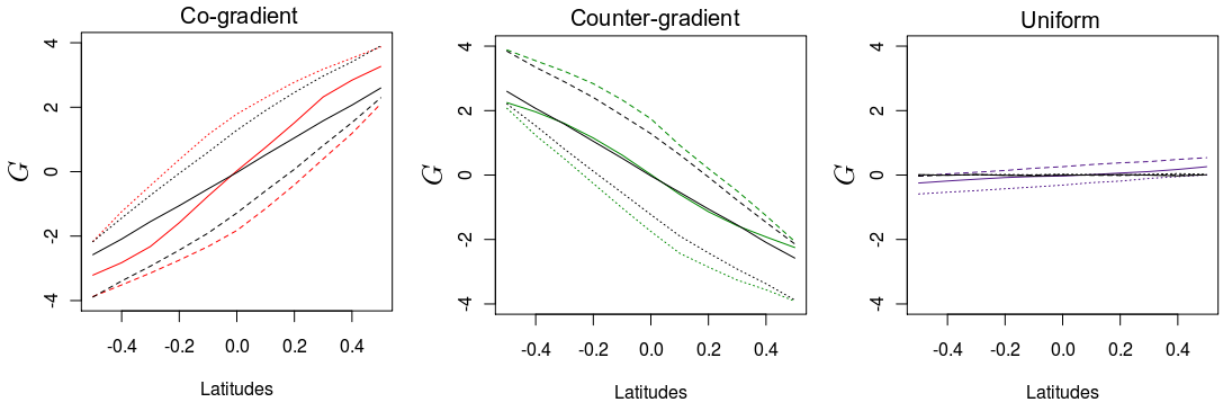
Genetic differentiation

We first examined the shift in the mean genetic value of populations at the landscape scale when steep selection gradients ($\sigma^2_{Z_{opt}}=6$) were simulated (figure 2). We focused here principally on the genetic values stabilized at the end of the simulation process, at generation 1000. Because of the joint influence of natural selection and assortative mating across the selection and the environmental gradients, the mean genetic values of populations were shifted rapidly and genetic clines built up. As expected, the extent and the shape of the differentiation established varied among the scenarios. The effects of assortative mating were particularly visible near the two edges of the landscape owing to the environmental gradient and long-distance pollen flow, which resulted in differentiation roughly clinal under IMM (figure 2). Overall, assortative mating clearly increased the level of differentiation under co-gradient divergent selection. Conversely counter-gradient settings slightly decreased the differentiation observed at generation 1000. Under IMM, the slope of the genetic cline generated averaged 6.2 in co-gradient conditions, greater than the slope of 4.7 obtained without assortative mating. Counter-gradient selection induced the shallowest cline with a slope of -4.2. Interestingly, assortative mating combined with uniform selection generated clinal differentiation as well (0.8 under IMM), even though populations shared similar optimal trait's values. As expected, under random mating, without filter of pollen, the mean genetic value of populations remained close to the Z_{opt} values assigned.

Directional migration models shifted uniformly the genetic values of the populations but had few consequences on the magnitude of the differentiation among populations at generation 1000 in presence of steep selection gradients. In this case, they resulted mainly in slightly more important shifts in the center of the landscape. However more pronounced effects were observed in earlier generations when selection gradients with lower slopes were simulated (supplemental figure 1). In conditions of uniform selection, directional pollen flow increased substantially the slope of the genetic clines induced by assortative mating (0.7 under LDN or LDS) (figure 2).

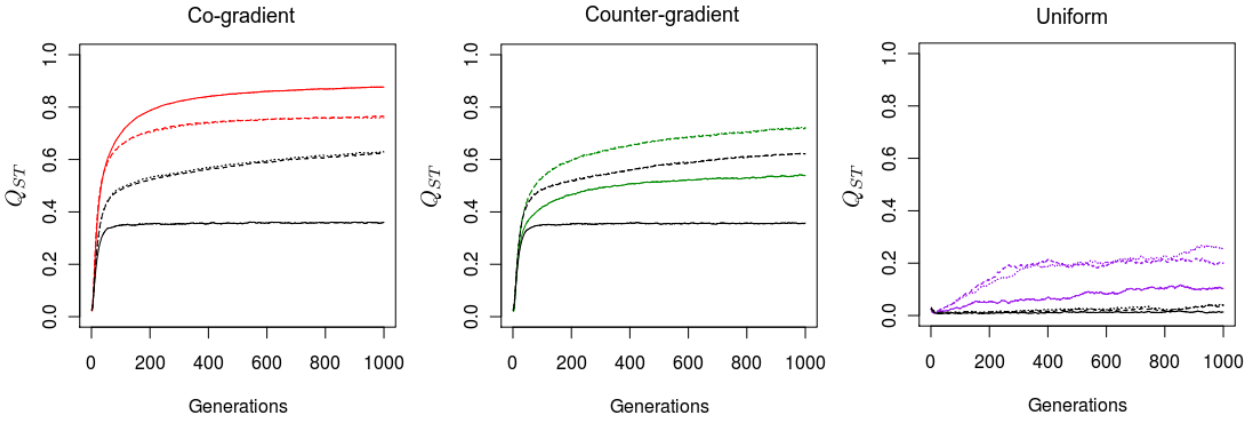
Q_{st} variation was driven by the establishment of clines: the highest Q_{st} value was obtained across co-gradient landscapes (0.85), while assortative mating across counter-gradient landscapes resulted in Q_{st} values of 0.5 and random mating yielded the lowest Q_{st} value, 0.35. When combined with uniform selection, assortative mating resulted in a significant differentiation of 0.09 under IMM and 0.2 under directional dispersal models (figure 3).

Figure 2. Variations in mean population genetic values at different latitudes.



The value for each latitude is the average of the five mean genetic values for the populations concerned at generation 1000 under steep selection and environmental gradients ($\sigma^2_{Z_{opt}}=6$, $k_E=2$, respectively). Latitudinal means were computed and reported for co-gradient (red), counter-gradient (green) and uniform selection (purple). The color lines indicate assortative mating ($\rho=0.8$), the black lines indicate random mating ($\rho=0$). All simulations were conducted under three migration models with moderate gene flow ($N_m=10.2$): the plain lines indicate IMM, the dashed lines LDN and the dotted lines LDS. Each line represents the mean of 30 independent replicates.

Figure 3. Variations in Qst under different evolutionary scenarios.



Variations monitored under assortative mating combined with co-gradient (red), counter-gradient (green) and uniform (purple) selection. In these scenarios, the black lines represent the values generated under random mating, the colored lines represents the QST values obtained under assortative mating. Three dispersal models were used here: IMM (plain line), LDN (dashed-line), LDS (dotted-line). Each line represents the mean of 30 independent replicates for each evolutionary scenario.

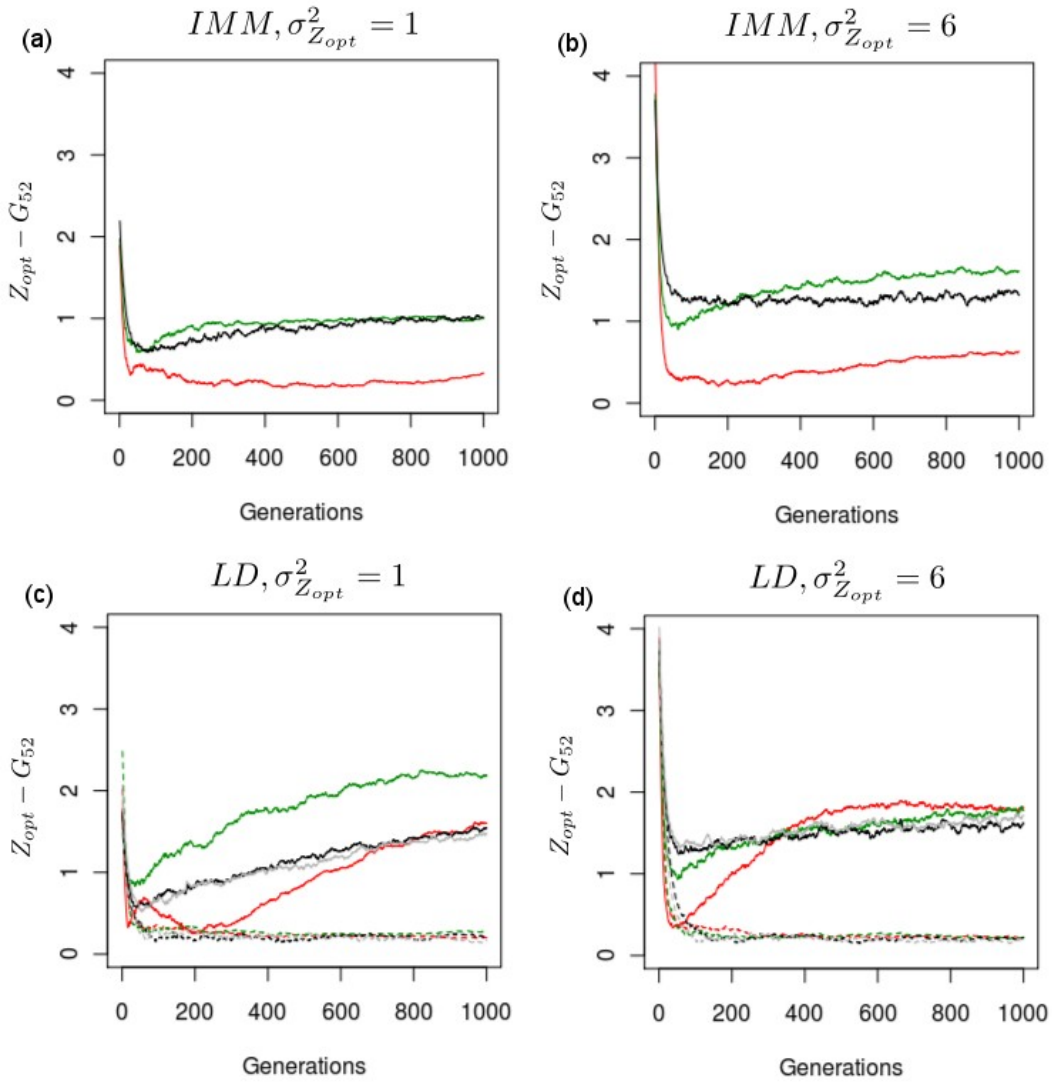
Maladaptation

We then monitored changes in the mean level of adaptation of population 52, located at the northern edge of the landscape, and population 28, located in the center of the landscape. While the examination of genetic clines yields a spatial and static information, the monitoring of evolutionary change of single populations gives insights into the evolutionary and adaptive dynamics. The level of adaptation of populations can be assessed by calculating the difference between their phenotypic optima and their current mean genetic values ($|Z_{opt} - G_t|$), larger differences indicating a greater degree of maladaptation.

Genetic values were shifted very rapidly towards their optimal values in the first generations, and levels of maladaptation therefore rapidly decreased (figure 4). At the northern edge, maladaptation decreased most strongly in co-gradient conditions: a mean maladaptation rate of 0.3 was observed at generation 50, *vs* 1.0 in counter-gradient conditions and 1.3 under random mating. Since the meta-population was at the mutation-migration-drift equilibrium at the beginning of the simulation process, alleles were randomly spread over the landscape and populations had similar mean genetic values. Thus, during the early generations, the adaptative response of populations was inflated as long as alleles were not significantly sorted across the environmental gradient through long-distance pollen flow filtered by assortative mating *[Soularue and Kremer, 2012]*. However, after generation 50, pollen flow from distant sources decreased rapidly the preliminary adaptation, especially in counter-gradient settings (figure 4). Indeed, the import of alleles with additive effects strongly diverging from the local optimum of recipient populations gradually restored and amplified maladaptation that reached a level significantly more important than under random mating (1.5 *vs* 1.2 under IMM, respectively, at generation 1000). However, this swamping effect was considerably reduced under co-gradient settings which induced a much lower maladaptation level of 0.5 at generation 1000 for population 52, as a result of more stringent pollen filtering severely constraining long-distance pollen flow (figure 5). Besides, under IMM, the differences among the adaptive responses were maximized when moderate selection across steep selection gradient was simulated (figure 4 (b)). On the other hand, weaker selection gradients ($\sigma^2_{Z_{opt}} = 1$) and higher stabilizing selection intensity ($\omega^2 = 5$) reduced and canceled the differences observed among the scenarios, respectively (figure 4 (a) and supplemental figure 2).

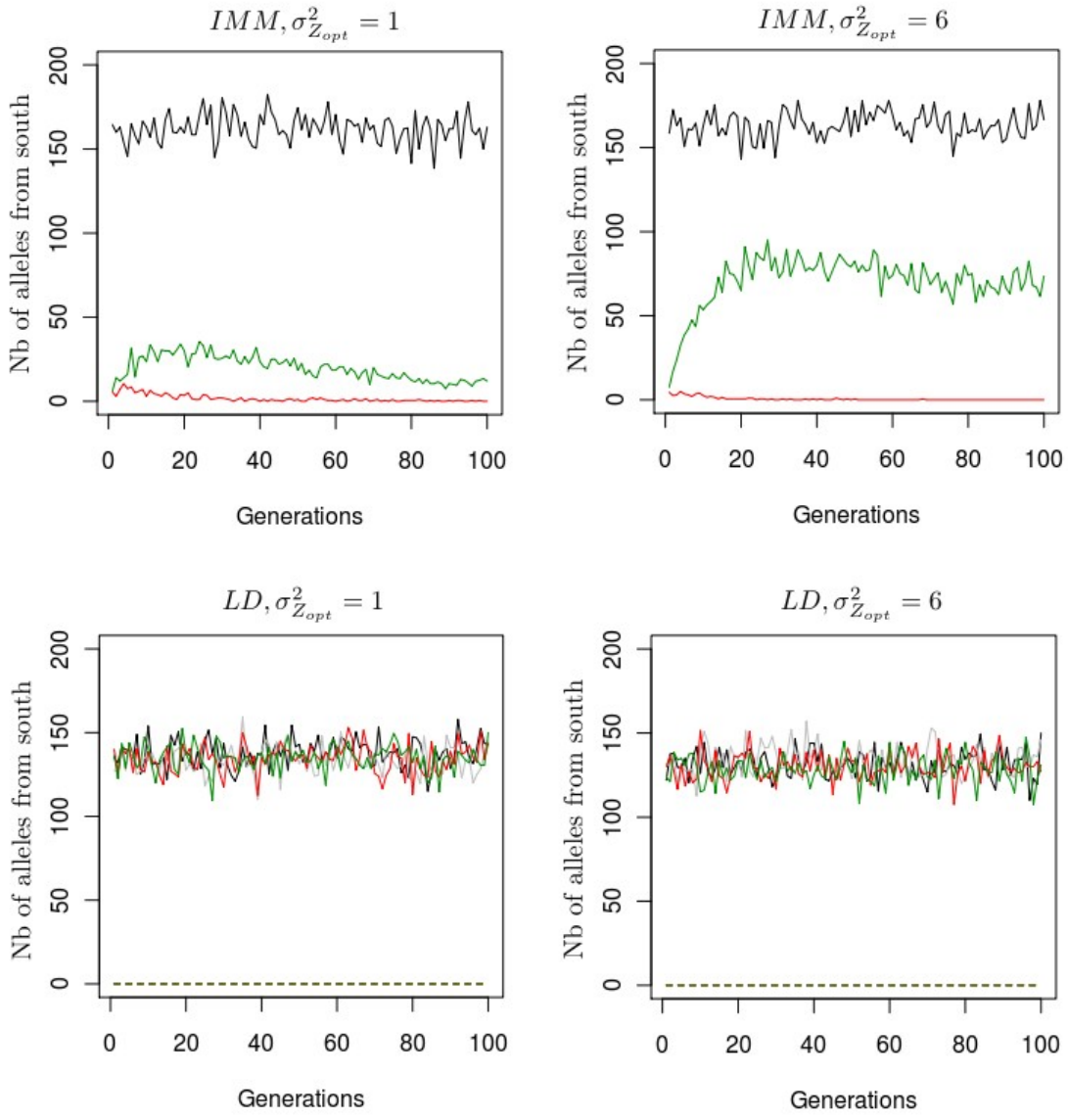
Under LDS, because the northern population of interest received no pollen from distant sources, maladaptation remained close to 0 throughout the simulation process, in all landscape configurations. By contrast, LDN was the migration model for which the number of distant alleles was the largest in population 52 (figure 6). Indeed, LDN swamped the pollen filtering occurring under assortative mating with distant pollen originating from southerly populations, hence promoting even more matings among distant individuals. As a result, overall LDN induced the highest maladaptation levels in population 52 (figure 4c, 4d). Nonetheless, in this case, contrasts among co, counter-gradient and random mating settings were particularly visible when selection gradients with low slope were simulated ($\sigma^2_{Z_{opt}} = 1$) (figure 4 (c)). In this scenario, LDN thus amplified the trend observed under IMM, the level of maladaptation reaching 1.3 at generation 200 and 2.1 at generation 1000 in conditions of counter-gradient variation. In the absence of assortative mating, the maladaptation level reached 0.8 at generation 200 and 1.3 at generation 1000 under LDN. Interestingly, the swamping effect of distant gene flow was considerably limited by co-gradient selection: 0.3 at generation 200 and 1.4 at generation 1000. Steeper selection gradients ($\sigma^2_{Z_{opt}} = 6$) resulted in initial adaptive responses strongly amplified by co-gradient selection, however, after 300 generations, the maladaptation levels observed were similar among the configurations simulated (figure 4 (d)). Indeed, in this latter case, LDN combined with the reduced differences among southerly and northern individuals canceled the filtering effect of assortative mating and lead to the highest maladaptation levels observed.

Figure 4. Variation in maladaptation level of a northern population.



The difference between the phenotypic optima and the mean genetic values of population 52 was monitored within 4 scenarios involving moderate ($\sigma^2_{Z_{opt}}=1$) and steep selection gradients ($\sigma^2_{Z_{opt}}=6$). Overall, the intensity of selection was moderate ($\omega^2=50$). Assortative mating ($\rho=0.8$) was simulated and combined either with co-gradient (red) or counter-gradient (green) selection. Alternatively, random mating was simulated: black lines represent the maladaptation level monitored when the selection gradient co-varied positively with the latitudinal gradient. Gray lines, only represented in (c) and (d), stand for the maladaptation level observed under random mating when directional dispersal models were associated with a reversed selection gradient. Because the direction of the selection gradient had no importance under IMM and random mating, this latter case was only explored when directional dispersal models were simulated. Three dispersal models were simulated: IMM (plain lines in (a) and (b)), LDN (plain lines in (c) and (d)) and LDS (dashed lines in (c) and (d)). LD mentioned in the titles of the sub-figures (c) and (d) means "long-distance and directional dispersal", it refers both to LDN and LDS. Each line represents the mean of 30 independent replicates for each evolutionary scenario.

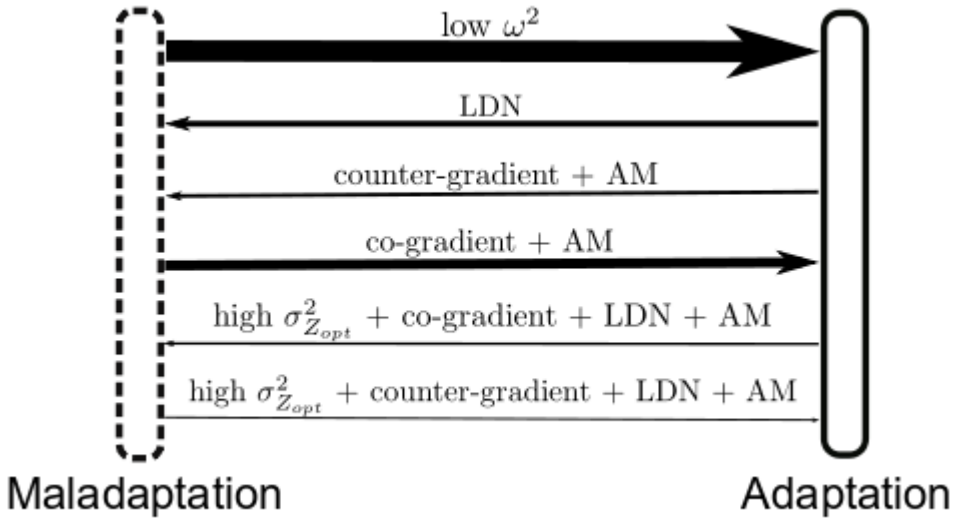
Figure 5. Amount of southern immigrant alleles received by a northern population over generations.



Absolute number of immigrant alleles into a population located at the extreme northern latitude (pop 52) and coming from southern latitudes (-0.5 to -0.1). Only gene flow between populations is represented here. Numbers on the y-axis are counts of alleles at a given generation (x-axis). Counts of alleles were monitored for two distinct selection gradients ($\sigma^2_{Z_{opt}}=1$, $\sigma^2_{Z_{opt}}=6$), three models of pollen dispersal (IMM, LDS, LDN) and two mating systems (assortative or random mating). Title of the figures indicates the steepness of the selection gradient and the dispersal model simulated. LD means that the pollen dispersal in the scenario is of "long-distance and directional" type, it refers both to LDN (plain line) and LDS (dashed-line). The color lines indicate assortative mating ($\rho=0.8$), red standing for co-gradient and green for counter-gradient directional selection. The black and gray lines stand for random mating with the selection gradient co-varying positively or negatively with latitudes, respectively. However, because under random mating the direction of the selection gradient has no impact when combined with IMM, the latter case was only considered when directional dispersal models were simulated (LD with $\sigma^2_{Z_{opt}}=1$ and LD with $\sigma^2_{Z_{opt}}=6$). Each line represents the mean of 30 independent replicates for each evolutionary scenario.

Similar trends were observed in the middle of the landscape (population 28) when directional pollen migration models were simulated (supplemental figure 3). The level of maladaptation increased particularly rapidly under counter-gradient settings (1.0 at generation 200), while it was limited under co-gradient settings during the first 300 generations (0.6 at generation 200). In comparison, the level of maladaptation of population 28 equaled 0.8 at generation 200 under random mating. However, the final level of maladaptation observed at generation 1000 was similar under co- and counter-gradient settings (1.8) and exceeded the value observed under random mating (1.2). When IMM was simulated, incoming pollen flow was balanced in relation to the environmental gradient in population 28, as a result, no significant effect on maladaptation was induced by assortative mating in this latter case.
At last, under conditions of uniform selection, the Z_{opt} value of all populations equaled 0. The level of maladaptation corresponded thus to the mean genetic values of the populations (figure 2). In this configuration, assortative mating generated significant levels of maladaptation, which nevertheless remained below those obtained with divergent selection. For example, population 52 showed a level of maladaptation of 0.3 at generation 1000 under IMM and 0.6 under LDN.

Figure 6. Drivers of adaptation.



Relative influence of the main determinants of the adaptive trajectories of northern populations. The direction and the thickness of each arrow represent the overall effect and its magnitude, respectively. LDN indicates long-distance pollen dispersal in the direction of north. AM stands for assortative mating. ω^2 represents the strengh of selection. Increase in the steepness of the selection gradient is designated by "high $\sigma^2_{Z_{opt}}$ ".

Discussion

Assortative mating and spatial subdivision of genetic variation
In silico simulations were consistent with our expectations, suggesting that the combination of assortative mating with gene flow and pre-existing environmental clines can generate contrasting distributions of genetic variation for traits across the landscape. Under clinal divergent selection acting in the same direction as the environmental cline (co-gradient conditions), genetic clines and differentiation are inflated by assortative mating. Conversely, under divergent selection of the same

strength but acting in the opposite direction (counter-gradient selection), the slope of the genetic clines is decreased. Finally, even under uniform selection, genetic clines are established that mimic co-gradient variation (figure 2). The theoretical settings considered here are entirely appropriate for TBB, as TBB is triggered by temperature and is strongly correlated with flowering time in trees. Interestingly, trees display strong clinal genetic variation in provenance tests also known as common garden experiments *[Wright, 1976][Morgenstern, 1996][Savolainen et al., 2007]*. Our results suggest that co-gradient genetic clines should be steeper than counter-gradient clines. Unfortunately, few reviews of experimental results in diverse species are available, making comparisons with theoretical predictions difficult. However, in a recent comparative study conducted along the same gradient of elevation in European oaks and beech, which display co- and counter-gradient variations, respectively, genetic clines were found to be steeper for the oaks than for beech *[Vitasse, 2009][Vitasse, 2011]*.

Assortative mating and temporal dynamics of differentiation and maladaptation

In addition to the overall effects on clines, we were also able to monitor the temporal dynamics of adaptive trajectories of populations. The combined effects of assortative mating and pollen flow serve to filter external male parents for synchrony to female parents in the recipient populations. This filtering effect changes over generations, with the spatial dispersal of alleles across the landscape (figure 5). The overall pattern of the evolution of maladaptation showed an initial adaptation state followed by an increase in the maladaptation rate whose extent strongly varied among scenarios when assortative mating was simulated. In co-gradient conditions, assortative mating significantly inflated the initial responses of recipient populations to selection, by filtering distant pollen flow and favoring the import of alleles helping in the tracking of optimal values. This finding confirms and extends previous results *[Fox, 2003][Jorjani et al., 1997]*, indicating that positive assortative mating induces the strongest response to selection in single populations. Additionally, in combination with co-gradient divergent selection, assortative mating tends to reduce the time to adaptation (figure 4). However, our models and simulations revealed that, when combined with counter-gradient selection, assortative mating decreases the ability of populations to track the shift of optimal values (figure 4). Moreover, in such a configuration, less stringent pollen filtering results in assortative mating increasing the time to adaptation, such that it may become critical for the population concerned *[Burger and Lynch, 1995]*. Overall, the contrast observed in the early generations between co-and counter-gradient settings was particularly visible when steep selection and environmental gradients were simulated (figures 4 (b) and 4 (c)). These findings are of prime importance in the context of climate change. Global warming will probably shift the optimal values of populations *[Lynch and Lande, 1993]*, which we considered to remain constant over successive generations in our simulations, thereby increasing the time to adaptation. Under such circumstances, in case of moderate selection intensity, assortative mating in co-gradient settings will probably favor adaptation and recovery, whereas, in counter-gradient settings, it may greatly delay adaptation and rescue.

Interestingly, phenological assortative mating gave rise to a striking difference at later time points. Indeed, in our simulations, the initial adaptation phase was short (30 to 50 generations) and was followed by the maintenance of adaptation or by a gradual loss of the state of adaptation initially attained by populations (figure 4). Despite a slight loss of adaptation level, especially present in case of counter-gradient conditions, the initial adaptation state attained by the populations was pretty well conserved under IMM, owing to the efficient filter of pollen realized by assortative mating (figure 5). However, when we modified the content of the pollen cloud by simulating directional dispersal models allowing for long-distance pollen flow, large backward shifts of the genetic values of the populations occurred, steadily increasing their degree of maladaptation (figure 4 (b) and (c), supplemental figure 3). This swamping effect was strongest at the edge of the landscape, where pollen from distant sources was more abundant. These findings are reminiscent of those of previous studies showing that peripheral populations are more likely to be swamped by

maladapted gene flow originating principally from central populations *[Garcia-Ramos and KirckPatrick, 1997][Kirckpatrick and Barton, 1997][Holliday et al., 2007]*. This swamping effect was most readily triggered in counter-gradient landscapes under conditions of extensive long-distance gene flow. By contrast, it was greatly delayed and constrained by assortative mating, especially across co-gradient landscapes exhibiting selection gradients of low slope. Interestingly, maladaptation of population 52 remained lower under co-gradient selection than under random mating during 800 generations. Nonetheless, the long-term filtering effect of assortative mating was reduced by steep selection gradients that increased even more the swamping effect of directional dispersal models (figure 4 (d)). Besides, the simulation of a very strong selection intensity evened out the effects of assortative mating (supplemental figure 2). Figure 6 resumes the global effects of the main drivers of adaptation observed in our simulations.

Maladaptation is difficult to assess in long-lived species, as it is difficult to record optimal population values and selection intensity. We are therefore unable to make comparisons with observed values for maladaptation, for the confirmation or rejection of our theoretical predictions. However, overall, this suggests that, below a very high threshold of long-distance pollen flow stemming from dispersal patterns, the filtering effect of assortative mating in the co-gradient configuration may be strong enough to transform the swamping effect of distant pollen flow into a favourable effect, promoting durable adaptation within peripheral populations. In this interesting configuration, our results are consistent with those of Davis and Shaw (2001), indicating that pre-adapted alleles from central populations may promote adaptation within recipient peripheral populations *[Davis and Shaw, 2001]*. On the whole, our results fall into the antagonist effects of gene flow on local adaptation in forest trees described by Kremer *et al.* (2012) *[Kremer et al., 2012]*.

We considered constant environmental variation over generations within our simulations, but field observations have shown that the slope of the environmental gradient may differ among tree species *[Vitasse et al., 2009]*. For example, Vitasse *et al*. (2009) showed that environmental gradients were more marked in oak and ash than in beech. We predicted that the impact of assortative mating would be dependent on the slope of the environmental gradient (equation 6) and we confirmed these predictions by numerical simulation *[Soularue and Kremer, 2012]*. Empirical and theoretical temperature-phenology relationships have been well described in a number of temperate tree species *[Vitasse et al. 2011]*, whereas comparisons with common garden observations are less frequent but can identify co- and counter-gradient settings *[Wright, 1976][Morgenstern, 1996] [Savolainen et al., 2007]*. This obtainable information should make it possible to predict the evolutionary responses of various temperate tree species in the context of future climate change.

Acknowledgement

We thank Sally Aitken, Susannah Tysor and Sam Yeaman for helpful discussions during model development. Thanks to Ophelie Ronce for her constructive comments on an earlier version of the manuscript. This research was supported by the EU project MOTIVE (project number: 226544).

Conflict of interest

The authors declare that they have no conflict of interest.

References

Aitken SN, Yeaman S, Holliday JA, Wang T, Curtis-McLane S: **Adaptation, migration or extirpation: Climate change outcomes for tree populations.** *Evolutionary Applications* 2008, 1:95–111

Alberto F, Bouffier L, Louvet J-M, Lamy J.-B, Delzon S., Kremer A: **Adaptive responses for seed and leaf phenology in natural populations of sessile oak along an altitudinal gradient**. *J Evol Biol* 2011, 24:1442-1454

Alberto F, Aitken S N, Alía R, González-Martínez SC, Hänninen H, Kremer A, Lefèvre F, Lenormand T, Yeaman S, Whetten R, Savolainen O: **Potential for evolutionary responses to climate change - evidence from tree populations.** *Global Change Biology* 2013, DOI: 10.1111/gcb.12181

Austerlitz F, Dick CW, Dutech C, Klein EK, Oddou-Muratorio S, Smouse PE, SorkVL: **Using genetic markers to estimate the pollen dispersal curve**. *Molecular Ecology* 2004*,* 13:937–954

Austerlitz F, Mariette S, Machon N, Gouyon PH, Godelle B: **Effects of colonization processes on genetic diversity: differences between annual plants and tree species**. *Genetics* 2000, 154.3: 1309-1321

Bohrerova Z, Bohrer G, Cho KD, Bolch MA, Linden KG: **Determining the viability response of Pinus taeda pollen under abiotic stresses typical to atmospheric conditions during long distance dispersal**. *Ecol. Appl.* 2009, 19:656–667

Bridle JR, Polechova J, Vines TH: **Limits to adaptation and patterns of diversity**. In: Speciation and Patterns of Diversity (eds Butlin RK, Bridle JR,Schluter D). *Cambridge University Press* 2009, 77–102

Burger R, Lynch M: **Evolution and Extinction in a Changing Environment: A Quantitative-Genetic Analysis.** *Evolution* 1995, 49:151-163

Cheaib A, Badeau V, Boe J, Chuine I, Delire C, Dufrêne E, François C, Gritti ES, Legay M, Pagé C, Thuiller W, Viovy N, Leadley P*:* **Climate change impacts on tree ranges: model intercomparison facilitates understanding and quantification of uncertainty**. *Ecology Letters* 2012, *15:*533-544

Chuine I and Beaubien E: **Phenology is a major determinant of tree species range**. *Ecology Letters* 2001, 4:500-510

Chuine I and Cour P: **Climatic determinants of budburst seasonality in four temperate-zone tree species**. *New Phytol.* 1999*,* 143*:*339-349

Conover DO and Schultz ET: **Phenotypic similarity and the evolutionary significance of counter-gradient variation**. *Tree* 1995*,* 10*:*248-252

Conover DO Duffy TA, Hice LA:**The covariance between genetic and environmental influences across ecological gradients: reassessing the evolutionary significance of counter-gradient and co-gradient variation**. *Ann N Y Acad Sci* 2009*,* 1168*:*100-129

Davis MB and Shaw RG*:* **Range shifts and adaptive response to quaternary climate change**. *Science* 2001*, 292:673-679*

De Lange AO: **A simulation study of the effects of assortative mating on the response to selection**. *Proceedings, 1st World Congress on Genetics Applied to Livestock Production* 1974, III: 421-425

Fox GA: **Assortative mating and plant phenology: evolutionnary and pratical consequences**. *Evolutionnary Ecology Research 2003, 5:*1-18

Franjic J, Sever K, Bogdan S, Skvorc Z, Krstonosic D, Aleskovic I: **Phenological asynchronization as a restrictive factor of ecient pollination in clonal seed orchards of Pedunculate Oak (Quercus robur L.)**. *Croat. j. for. eng.* 2011, 31:141:156

García-Ramos G and Kirkpatrick M: **Genetic models of adaptation and gene flow in peripheral populations**. *Evolution* 1997 51:21-28

Gianola D: **Assortative mating and the genetic correlation**.*Theor Appl Genet* 1982*, 62:* 225-231
Hamrick JL, Godt MJ, Sherman-Broyles SL: **Factors influencing levels of genetic diversity in woody plant species**. New For. 1992, 6: 95–124

Gomulkiewicz R, Houle D: **Demographic constraints on evolution**. Am Nat 2009, 174(6):E218-29

Holliday JA, Suktan H, Aitken SN: **Divergent selection and heterogeneous migration rates across the range of Sitka spruce (Picea sitchensis)**. *Proceedings of the Royal Society Series B* 2012, 279: 1675-1683.

Jorjani H, Engstrøom G, Strandberg E, Liljedahl LE: **Genetic studies of assortative mating - a simulation study. III. Assortative mating in selected populations**. *Acta Agric. Scand.* 1997, 47:129:137

Kirkpatrick M and Barton, NH: **Evolution of a species range**. *Am. Nat. 1997.* 150:1–23

Kremer A, Le Corre V: **Decoupling of differentiation between traits and their underlying genes in response to divergent selection**. *Heredity* 2011, doi: 10.1038/hdy.2011.81

Kremer A, Ronce O, Robledo-Arnuncio JJ, Guillaume F, Bohrer G, Nathan R, Bridle JR, Gomulkiewicz R, Klein EK, Ritland K, Kuparinen A, Gerber S, Schueler S: **Long-distance gene flow and adaptation of forest trees to rapid climate change**. *Ecology Letters* 2012, 15: 378–392, doi: 10.1111/j.1461-0248.2012.01746.x

Lande R: **Natural Selection and Random Genetic Drift in Phenotypic Evolution**. *Evolution* 1976*,* 30 (2):314-334

Lopez S, Rousset F, Shaw FH, Shaw RG, Ronce O: **Migration load in plants: role of pollen and seed dispersal in heterogeneous landscapes**. *J Evol Biol.* 2008, 21(1):294-309

Le Corre V and Kremer A: **Genetic Variability at Neutral Markers, Quantitative Trait Loci and Trait in a Subdivided Population Under Selection**. *Genetics* 2003, 164(3):1205-1219

Lynch M and Lande R: **Evolution and extinction in response to environmental change**. In P.

Kareiva, J. Kingsolver, and R. Huey (eds.) *Biotic Interactions and Global Change. Sinauer Assocs., Inc. Sunderland, MA*, 1993, pp. 234-250

Mimura M and Aitken SN: **Increased selfing and decreased effective pollen donor number in peripheral relative to central populations in Picea sitchensis (Pinaceae)**. *American Journal of Botany* 2007, 94:991-998

Morgenstern EK: **Geographic variation in forest trees: genetic basis and application of knowledge in silviculture**. *UBC press* 1996

Nathan R, Schurr FM, Spiegel O, Steinitz O, Trakhtenbrot A, Tsoar A: **Mechanisms of long-distance seed dispersal**. *Trends Ecol. Evol.* 2008, 23:638-647

Savolainen O, Pyhäjärvi T, Knürr T: **Gene flow and local adaptation in trees**. *Annu Rev Ecol Evol Syst* 2007, 38:595-619

Stam P: **The evolution of reproductive isolation in closely adjacent plant populations through differential flowering time**. *Heredity* 1983, 50 (2), 105-118

Schueler S, Schlünzen K, Scholz F: **Viability and sensitivity of oak pollen and its implications for pollen-mediated gene flow**. *Trends Ecol. Evol* 2005, 19:154:161

Soularue JP and Kremer A: **Assortative mating and gene flow generate clinal phenological variation in trees**. *BMC Evolutionary Biology* 2012, 12:79

Thuiller, W: **BIOMOD: Optimising predictions of species distributions and projecting potential future shifts under global change**. *Global Change Biology* 2003 9:1353-1362

Turelli M: **Heritable genetic variation via mutation-selection balance: Lerch's zeta meets the abdominal bristle**. *Theor. Popul. Biol.* 1984, 25:138:193

Vitasse Y, Delzon S, Dufrene E, Pontailler JY, Louvet JM, Kremer A, Michalet R: **Leaf phenology sensitivity to temperature in European trees: Do within-species populations exhibit similar responses?** *Agricultural and forest meteorology* 2009, 149:735:744

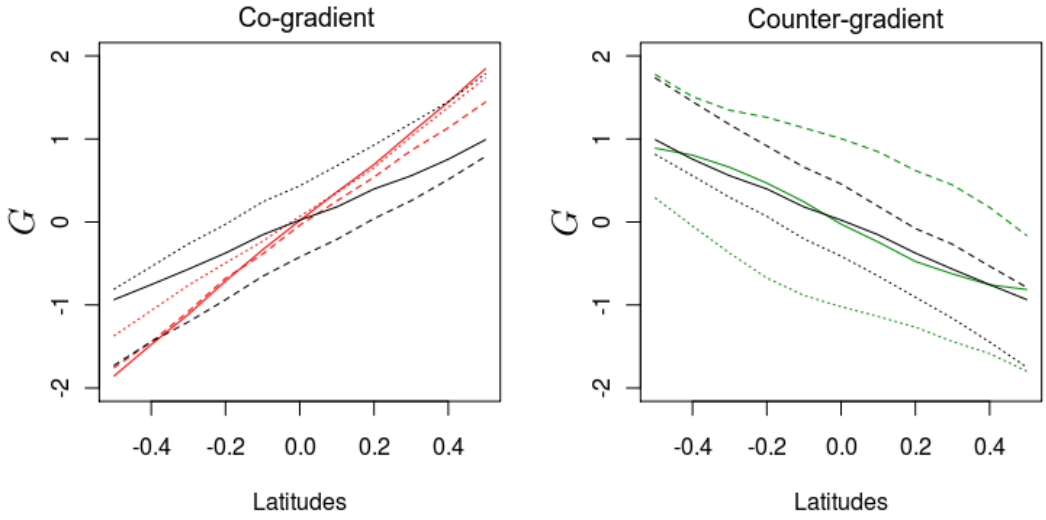
Vitasse Y, Francois C, Delpierre N, Dufrêne E, Kremer A, Chuine I, Delzon S: **Assessing the eects of climate change on the phenology of European temperate trees**. *Agricultural and Forest Meteorology* 2011, 151 (7):969:980

Wright JW: **Introduction to forest genetics**. *Academic Press, INC.* 1976

Yeaman S, Guillaume F: **Predicting adaptation under migration load: the role of genetic skew** *Evolution* 2009 *63:2926-2938*

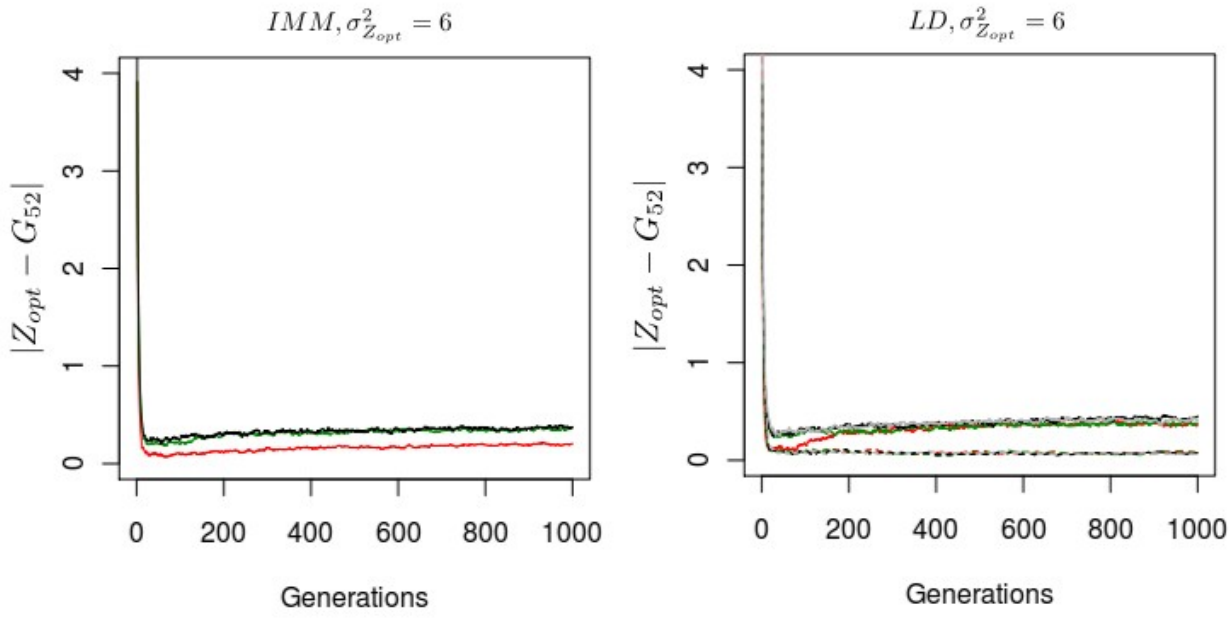
Supplemental figures

Supplemental figure 1. Variations in mean population genetic values at different latitudes.



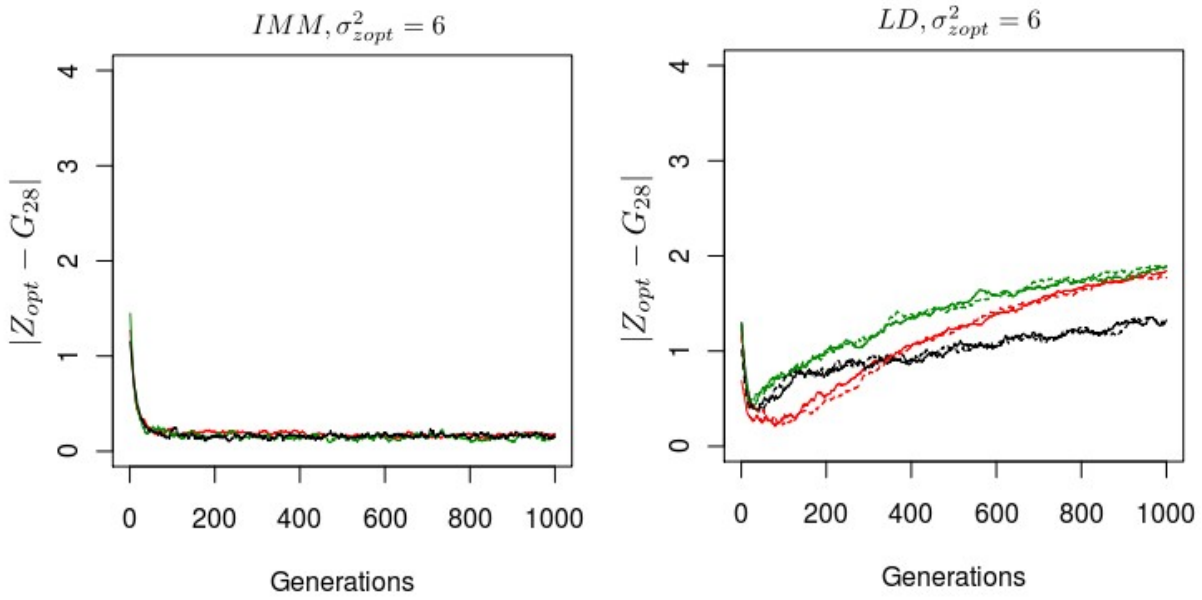
The value for each latitude is the average of the five mean genetic values for the populations concerned at generation 300 under limited selection and steep environmental gradients ($\sigma^2_{Z_{opt}}=1$, $k_E=2$). Latitudinal means were computed and reported for co-gradient (red), counter-gradient (green) and uniform selection (purple). Limited selection gradients were simulated here ($\sigma^2_{Z_{opt}}=1$) The color lines indicate assortative mating ($\rho=0.8$), the black lines indicate random mating ($\rho=0$). All simulations were conducted under three migration models with moderate gene flow ($N_m=10.2$): the plain lines indicate IMM, the dashed lines LDN and the dotted lines LDS. Each line represents the mean of 30 independent replicates.

Supplemental figure 2. Variation in maladaptation level of a northern population under a stronger selection intensity.



The difference between the phenotypic optima and the mean genetic values of a northern population (population 52) was monitored under two additional scenarios involving a stronger selection intensity associated with a steep selection gradient ($\omega^2=5$, $\sigma^2_{Z_{opt}}=6$). Title of the figures indicates the steepness of the selection gradient and the dispersal model simulated, LD meaning "long-distance and directional dispersal", which includes both LDN (plain lines) and LDS (dashed lines). Assortative mating ($\rho=0.8$) was combined with co-gradient (red) and counter-gradient (green) selection. Random mating was also simulated; black lines represent random mating when the selection gradient co-varied positively with the latitudinal gradient, gray lines represent random mating when the selection gradient is opposed to the latitudinal gradient. The reversion of the selection gradient cannot have any effect under randoim mating when IMM is simulated, that is why it was only considered with directional dispersal models: LDN and LDS. Each line represents the mean of 30 independent replicates for each evolutionary scenario.

Supplemental figure 3. Variation in maladaptation level of a population located in the central part of the landscape.



The difference between the phenotypic optima and the mean genetic values of a population located in the central part of the landscape (latitude 0, population 28) was monitored under moderate selection intensity associated with a steep selection gradient ($\omega^2 = 50$, $\sigma^2_{Z_{opt}} = 6$). Assortative mating ($\rho = 0.8$) was combined with co-gradient (red) and counter-gradient (green) selection. In these scenarios, the black lines stand for random mating. Three models of pollen dispersal were simulated: IMM, LDN and LDS. Title of the figures indicates the steepness of the selection gradient and the dispersal model simulated. LD means "long-distance and directional dispersal", it refers both to LDN and LDS. Each line represents the mean of 30 independent replicates for each evolutionary scenario.

Discussion

Synthesis

This work demonstrated in two steps the evolutionary impact of assortative mating on phenological traits in landscape settings comprising multiple populations and environmental gradients. First, in a fictive context excluding natural selection, we showed that the sole interaction between assortative mating and gene flow is sufficient to shift the genetic values of populations in the presence of environmental gradients. In this context, the resulting differentiation was particularly important when long-distance and non-isotropic pollen flow were simulated. However, though steep environmental gradients amplify the resulting differentiation, too steep environmental gradients tend to even it out, because they generate too much differentiation among distant phenotypes and limit long distance dispersal events. Along with the differentiation generated, we observed a drastic and definitive decrease of the genetic variance within populations occurring after a sharp initial increase. This decrease suggests that assortative mating could amplify a drift effect normally weak in large populations, it is also a consequence of the increase in variance among populations observed.
In a second step we considered a more realistic context, including divergent selection and pollen dispersals occurring over long distance at rates following realistic fat-tailed leptokurtic distributions. This second study had also a strong emphasis on the directionality of the processes interacting. It indicated that assortative mating clearly promotes the adaptive response of populations to selection across co-gradient landscapes, while slightly constraining it across counter-gradient landscapes, particularly in the presence of long-distance gene flow. Interestingly, assortative mating resulted in genetic differentiation even when homogeneous selection was simulated. Our simulations showed also that assortative mating combined with co-gradient selection strongly constrains the migration load at the margins of the landscape, even under extensive pollen dispersal exclusively originating from central populations. On the other hand, assortative mating in combination with counter-gradient selection greatly amplifies and accelerates this swamping effect.
Since considerable attention has been drawn to the shift of phenological traits induced by shifts in environmental conditions observed in the frame of climate changes, this work suggests that the antagonistic effects of assortative mating should be accounted for, both in the interpretation of genetic variability characterized within provenance tests and in future theoretical frameworks aiming at predicting the evolution of the distributions of tree species over large scales. It could also be interesting to conduct a meta-analysis of the data produced within provenance tests along environmental gradients comparing the magnitude of the differentiation at TBB in several tree species. If the slopes of the clines characterized in *co-gradient species* showed on average steeper slopes than *counter-gradient species* relatively to the environmental gradients, one of the underlying causes could be assortative mating. However, this thesis has to be prolonged in order to further disentangle the evolutionary dynamics shaping genetic variation in multiple populations. We propose in what follows some additional elements and perspectives that deserve to be explored.

Outlook: future developments

More complex patterns of variations in environmental conditions
The patterns of genetic differentiation we aimed to explain here varied according to single dimensions following environmental gradients. Consequently, the influence of the evolutionary factors we implemented here varied according to single directions. Nonetheless, mostly, environmental influence does not vary linearly according to single dimensions, across the whole distributions of species. As an example, in British Columbia, the distribution of Sitka Spruce ranges

over thousands kilometers both latitudinally and longitudinally, the presence of mountains and altitudinal clines together with strong latitudinal variations in temperature regimes suggest to integrate environmental variations following two dimensions. In collaboration with Sally Aitken we realized a small study case exploring the effects of phenological assortative mating on selection across landscapes showing 2-dimensional variations. The environmental component of the phenotypes simulated was composed of E_{lat}, varying clinally along latitudes (equivalent to E in chapter 2) and E_{long} varying either clinally along longitudes or heterogeneously. The resulting phenotype Z' expressed in chapter 2 became

$$Z' = G + E_{lat} + E_{long}$$

In addition to longitudinal environmental variations, we defined a second axis of selection, defining longitudinal variations of Z_{opt}. The objective was to verify the very simple hypothesis that the low level of differentiation for TBB in Sitka Spruce within provenance tests could stem from the interplay between assortative mating and divergent selection along landscapes prone to environmental fluctuations along two axis. This is of course what we observed, the simulations involving distinct pollen dispersal models and different associations of E_{lat} and E_{long} showed that the longitudinal and heterogeneous variations shrinked the latitudinal cline, which resulted in reduction of the final differentiation observed. Overall, this small study case that confirms our intuition, stresses the fact that the environmental variations accounted for have to stick with the intrinsic properties of the landscape considered.
Furthermore, the landscapes are also subject to environmental variations in time that have to be combined with variations in space, noticeably in the respect of simulating variations of selection pressures and environmental influence associated to climate change. Because in this thesis we mainly tried to explain static patterns of variations, our models and simulations were primarily based on selective forces constant in time, from the assumption that the existing clinal patterns of genetic variation at TBB resulted from divergent selection. Considering existing evolutionary theory, and since it is very difficult to determine exactly the nature of selection forces that natural populations have underwent over such long period, this is an acceptable assumption. But climate change is inducing new selection pressures that arise after long period of pretty stable environmental conditions. It might therefore be interesting to extend our work and study more deeply evolutionary dynamics of populations by shifting uniformly Z_{opt} values. The magnitude of the shift in phenotypic optimum values could be scaled via functions relating environmental changes and Z_{opt} changes. To be realistic, this function should take several input related to climatic predictions, the species considered and other characteristics of the area simulated. As indicated in *[Kremer et al., 2012]*, environmental variations are not limited to change in single factors such as temperatures. They more likely involve changes in several factors: temperature, precipitation, CO2 concentrations and other biotic and abiotic conditions. They should be considered together to improve the realism of the studies realized.

Contrasted densities of populations
Up to now we assumed that the number of individuals was constant and identical among populations, which is far from being the case in real settings. Indeed, densities of populations at the edge of distribution resulting from recent colonization events are often lower than the densities observed in well established central populations. Since the number of individuals of populations is tightly correlated to pollen dispersal and their overall adaptation ability, it might be relevant to account for variation in number of individuals both in space and time. We could start from the simple model of populations growth in discrete generations proposed by Holt and Gomulkievicz (2004) *[Holt and Gomulkiewicz, 2004]*:

$$N_{i_{t+1}} = \bar{W}_{i_t} N_{i_t}$$

where N_{i_t} denotes the number of individual of population *i* at generation *t* and $\bar{W}_{i_t}$ the mean fitness value of population *i* at generation *t*. The implementation of this model would allow to simulate colonization and establishment process at the edges of the distribution. Interestingly this model would allow to assess accurately the evolution of small populations at the edge, undergoing environmental change and receiving extensive gene flow originating from large central populations when assortative mating occurs. The goal would be to explore how the filtering of assortative mating balances with the direction of the selection gradient and the demography of populations. Will the effects of assortative mating identified in this thesis be decreased or totally canceled, depending on the disequilibrium in individual numbers among central and peripheral populations ?

Phenotypic plasticity

As pointed out in the introduction part, local adaptation involving phenotypic change can rely both upon genetic evolution and phenoypic plasticity. According to theoretical works *[Price et al., 2003] [Chevin et al., 2010]* and provenance tests results *[Alberto et al., 2011]*, much of the phenotypic change observed within populations facing environmental change is caused by phenotypic plasticity. The phenotypic decomposition implemented within our simulations accounted here for constant plasticity; change in environmental conditions resulted in similar linear norms of reaction for any possible genotypes. But plasticity is often genetically variable, heritable and evolvable *[Scheiner and Lyman, 1989]*.

Starting from the model proposed by Lande in which plasticity results from linear reaction norms *[Lande, 2009]*, we could implement genetic plasticity within our simulation framework according to

$$Z = a + bE + \varepsilon$$

where *Z* represents the phenotypic value of individuals, *a* the additive value of the genes at the non plastic loci, *b* the additive value at the plastic loci, *E* the macro-environmental effect introduced in chapter 2 and ϵ, the random environmental variation introduced in chapter 2 as well. It would be then possible to combine plasticity with assortative mating, which would result in

$$Z' = a + bE + E + \varepsilon$$

where *E* stands for constant plasticity and *bE* stands for heritable plasticity. By considering landscapes showing spatial variations of environmental conditions and selection pressures, heterogeneity in densities among populations, and occurrence of abrupt shift in environmental conditions, it would hence be possible to mimic realistic situations relevant with ongoing climate change. This extension allows to investigate new avenues regarding population responses to environmental changes. Overall, could plasticity in interaction with assortative mating and distant gene flow help populations in withstanding steeper shift in environmental conditions? Will plasticity mediate the final genetic differentiation observed ? Will the results described in *[Lande, 2009][Chevin and Lande, 2009]* be confirmed in our context ? The models of Lande and Chevin indicated that, compared to the standard darwinian adaptation, plasticity accelerates phenotypic change through an interesting 2-phase adaptation process. Just after a sudden abrupt change in environmental conditions, plasticity first accelerates change in phenotypic values, helping individuals to track their phenotypic optima and thus increasing the mean fitness of populations. This first phase noticeably comes along with an increase in genetic additive variance at the plastic loci (*b* in the model presented above). This first phase is then systematically followed by a further genetic assimilation of the phenotypic change. Genetic assimilation is defined by Lande as a "potentially important process following the evolution of increased plasticity during adaptation to environmental change". The predictions provided by his model suggest genetic assimilation to cause a fitness decrease called "cost of plasticity". The implementation of plasticity within our framework will allow to extend our work into more complex settings by introducing changes of

populations densities and different genetic architectures of adaptive traits. It will make possible to monitor these two phases noticeably by following the evolution of the covariance between the plastic and the non plastic additive values ($cov(a,b)$) both over time and space. Finally it will be possible to further investigate accurately how assortative mating and gene flow will balance with phenotypic plasticity and, whether in our context, plasticity does increase the magnitude of the shift in Z_{opt} values that can be tolerated by populations.

Epistasis

Going further in the investigations of the influence of genetic architecture supporting the phenotypes under selection leads then to the study of the effects of interactions among loci: dominance and epistasis. Because dominance can be seen as a special case of epistasis, we will only use the term epistasis in what follows. Though epistatic interactions are often neglected within the interpretation of wide genetic data set, it has been suggested and demonstrated that such interactions might considerably affect the fate of populations undergoing selection pressures, either accelerating or constraining adaptation *[Carter et al., 2005][Lamy et al., 2011]*. Hansen *et al.* (2006) modeled directional epistatic interactions through a multilinear model, assigning shift in additive effect at loci involved in epistatic interactions. They showed theoretically that epistasis based on complex patterns of interactions greatly affect evolutionary responses. Moreover, the overall effect induced depend strongly on the degree of complexity of the overall patterns of interaction and the additive effects of each interaction *[Hansen, 2006]*. In the light of these results, it seems necessary to account for epistatic interactions, and preferentially, interactions of order *n*. In addition, beyond the order of epistatic interactions, a point rarely considered concerns the heritability and the evolvability of epistatic interaction networks. Integrating such aspects within existing pure analytical frameworks can be very challenging. A first step could be to extend existing frameworks such as *Forsim* and *Quantinemo* (see chapter 1), which propose to simulate a fixed number of interaction patterns with associated additive effect. However that would not be sufficient for the simulations of evolvability through the quantitative genetics paradigm. It would rather require to initially generate randomly networks of interactions that will be inherited generations after generations. Using simple aspects of graph theory implemented with matrix, in combination with quantitative genetics, it might be possible to describe additively general properties of graphs and make them heritable by setting locus describing additively each feature of the networks: number of edges, vertex connected, epistatic effects... However, the implementation and the exploitation of such networks over thousands generations and individuals will be CPU-time consuming; much attention will have to be paid on the complexity of the resulting program and the use of heuristics could be necessary.

Towards realistic and large-scale predictions

The implementation of the various suggested extensions in a unique inferential framework would enable to deeply analyze the evolution of quantitative characters relying on complex architecture over complex environments. It would also enable to explore many distinct hypothesis, hence facilitating the exploitation and the interpretation of the data obtained from provenance and progeny tests. This would also be a first step toward realistic predictions conducted over wide areas.

However, the prerequisite for such an inferential framework, would be to allow for accurate descriptions of complex interactions among realistic phenotypes and environmental conditions. Because realistic phenotypes undergoing selection are likely to be composed of multiple characters, each one undergoing distinct selection pressures, there is a clear need for extending the framework to a multi-trait context. In addition, the QTLs associated to each character might be in pleiotropic interactions.

As argued in the introduction and in chapter one, it might be difficult to set up pure analytical frameworks for the description of such complex genetic architectures and environment. Though analytical models constitute valuable resources to disentangle general evolutionary dynamics and trends, simulations frameworks constitute ideal tools owing to higher descriptive and integrative

potentials. However, the review of the quantitative simulation frameworks proposed in chapter 1 indicates that the existing tools propose useful features, but none of them is sufficiently integrative yet. Generally speaking, the way models and features are programmed is tightly related on the overall representations and the expectations of the investigators. On the other hand, generic frameworks like *Simupop* require the user to spend much time in designing programs.

In a context in which the computer power is steadily increasing, it is possible to extend existing individual-based simulation packages to larger genetic and environmental contexts. The future simulating frameworks should allow to match any kind of real configurations which will considerably increase the relevance of predictions. Many questions will have to be answered noticeably to find the equilibrium among genericity, need for details and friendliness for the future frameworks. At the same time, isolated theoretical investigations of typical cases, based on analytical models and simpler software, will still be necessary to further enhance our overall understanding of evolutionary dynamics. Such investigations will further help in identifying the processes that must be included in integrative inferential platforms because of their important evolutionary impact.

References

Alberto F, Bouffier L, Louvet J-M, Lamy J.-B, Delzon S., Kremer A: **Adaptive responses for seed and leaf phenology in natural populations of sessile oak along an altitudinal gradient**. *J Evol Biol* 2011, 24:1442-1454

Carter AJR, Hermisson J, Hansen TF: **The role of epistatic gene interactions in the response to selection and the evolution of evolvability**. *Theoretical Population Biology* 2005, 68:179–196

Chevin L-M, Lande R, Mace GM: **Adaptation, Plasticity, and Extinction in a Changing Environment: Towards a Predictive Theory**. *PLoS Biol* 2010, *8(4): e1000357, doi:10.1371/journal.pbio.1000357*

Chevin LM and Lande R. **When do adaptive plasticity and genetic evolution prevent extinction of a density-regulated population?**. *Evolution,* 2009, 64:1143-1150

Hansen TF, Álvarez-Castro JM, Carter AJR, Hermisson J, Wagner GP: **Evolution of the genetic architecture under directional selection**. *Evolution* 2006, 60: 1523–1536. doi: 10.1111/j.0014-3820.2006.tb00498.x

Holt RD and Gomulkiewicz R: **Conservation Implications of Niche Conservatism and**

Kremer A, Ronce O, Robledo-Arnuncio JJ, Guillaume F, Bohrer G, Nathan R, Bridle JR, Gomulkiewicz R, Klein EK, Ritland K, Kuparinen A, Gerber S, Schueler S: **Long-distance gene flow and adaptation of forest trees to rapid climate change**. *Ecology Letters* 2012, 15: 378–392, doi: 10.1111/j.1461-0248.2012.01746.x

Lamy JB, Plomion C, Kremer A, Delzon S: **QST < FST As a signature of canalization**. *Molecular Ecology* 2012

Lande R: **Adaptation to an extraordinary environment by evolution of phenotypic plasticity and genetic assimilation**. *J. Evol. Biol.* 2009, 22:1435–1446

Price TD, Qvarnstrom A, Irwin DE: **The role of phenotypic plasticity in driving genetic evolution**. *Proc. R. Soc. Lond. B Biol. Sci. 2003, 270:1433 -1440*

Scheiner SM and Lyman RF: **The genetics of phenotypic plasticity I. Heritability**. *J. evol. Biol. 2*: 95- 107 (1989)