Université de Toulouse

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :
Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)

Cotutelle internationale avec :

Présentée et soutenue par :
Aurélie Khimoun

Le vendredi 29 juin 2012

Titre :

Histoire évolutive, contexte spatial et écologique de la divergence de deux sous-espèces d'Antirrhinum majus

---

ED SEVAB : Écologie, biodiversité et évolution

Unité de recherche :
Laboratoire Evolution et Diversité Biologique, UMR 5174, CNRS/UPS

Directeur(s) de Thèse :

Dr Monique Burrus et Dr Christophe Andalo

Rapporteurs :

Pr Christophe Destombe et Dr John Thompson

Autre(s) membre(s) du jury :
Dr Nadir Alvarez (Examinateur)
Dr Sophie Gerber (Examinatrice)
Pr Sergine Ponsard (Examinatrice)
Dr Monique Burrus (Directrice de thèse)
Dr Benoit Pujol (Membre invité)

« On ne peut pas croire des choses impossibles [dit Alice].

- Je suppose que tu manques d'entrainement, dit la Reine… Il m'est arrivé quelquefois de croire jusqu'à six choses impossibles avant le petit déjeuner. »

Lewis Carroll :

De l'autre côté du miroir

À mes parents

Remerciements

Pour n'avoir lu que cette partie de bon nombre de thèses, me voici maintenant en position de l'écrire. Entre déballage sentimentaliste et pudeur frustratrice, je dois faire face à la tâche délicate de trouver les mots justes pour vous exprimer ma gratitude.

Mes premiers remerciements vont tout naturellement vers mes deux directeurs de thèse, Monique Burrus et Christophe Andalo. Merci de m'avoir accueillie si chaleureusement dans l'équipe *Antirrhinum,* dès le stage de M1, où tu m'apprenais la biomol, Monique. Christophe, on a découvert ensemble la modélisation de niche et cette période restera un des moments les plus agréables de cette thèse. Je vous suis reconnaissante de la confiance et de la liberté que vous m'avez accordées tout au long de cette thèse.

Benoit, bien que tu n'aies pas été officiellement mon encadrant, cette thèse te doit beaucoup et ma rédaction aussi… Merci pour toutes ces discussions passionnantes et ton positivisme à toute épreuve.

Christophe T, tu as été présent tout au long de cette thèse, tu l'as accompagnée de tes conseils toujours très éclairants et je t'en remercie.

Josselin, je suis ravie d'avoir travaillé avec toi sur le chapitre 3, notre collaboration a été très enrichissante et des plus agréables, merci copain !

Ferran, tu as participé à plusieurs projets de cette thèse, ton aide et tes compétences ont été très précieuses. Merci de m'avoir accueillie si chaleureusement d'abord à Vienne, puis chez toi à Valence. Grâce à toi j'ai moins peur du monde obscur des théoriciens.

De nombreuses personnes ont pris le temps de discuter avec moi lors de comités de thèse ou lors de visites dans leurs labos. Merci à Pauline Garnier-Géré, Olivier François, Ferran Palero et Lounes Chikhi.

Le génotypage des microsatellites était parfois délicat, merci Gabriel d'avoir été là pour me dire que les données microsats parfaites n'existent pas.

Joris et Boris, merci pour toutes nos discussions et merci d'avoir répondu présents, jusqu'à la fin pour la relecture du manuscrit.

Anaïs et Pierre-Jean, merci d'avoir répondu avec bienveillance et patience à mes questions paniquées de thésarde en fin de thèse.

J'ai été aidée dans les manips par des aides précieuses, merci à Clervie lecorre, Emma Trichet et Alice Martin.

Parce qu'enseigner a fait partie de ma thèse, merci à vous tous de m'avoir fait partager votre expérience : Pierrick, Emilie, Brigitte, Alexandre, Sergine, Jean-Baptiste, Nathalie, Monique, Christophe.

Au labo il n'y a pas que des collègues de travail, il y a aussi des amis…

Cette thèse m'a permis de vous rencontrer, Juju et Chlochlo, et de tisser des liens d'amitié avec vous. Nous avons vécu ces 3 années ensembles et ça a été un bonheur de partager votre quotidien. Même à distance votre soutien m'a accompagnée dans la dernière ligne droite. Merci les belettes !

Joss depuis notre M2, Jojo depuis ton M2 et Boris depuis peu, c'est un privilège de passer mes journées avec des gens comme vous. La fine équipe !

Céline et Sophie, votre bienveillance et votre amitié m'ont accompagnée tout au long de ces 3 ans. Merci pour votre soutien dans le sprint final. Céline, merci pour tes bons petits plats et ta gentillesse. Sophie, merci pour les lessives, pour m'avoir accueillie chez toi pour que je puisse travailler le weekend et merci pour les repas entre filles Manzi.

Amaia, tu croyais t'être débarrassée de moi au 4R3 ? et non ! Merci de m'avoir accueillie dans ton bureau, merci pour ta bonne humeur. Tout ce que je sais de l'Espagnol, je le tiens de toi !

Enfin, j'ai partagé une pause, un repas ou une bonne rigolade avec chacun d'entre vous. Ces moments de détente ont jalonné ces 3 ans et les ont rendus si agréables. Merci Léa (t'es trop cool !), Olivia (merci beaucoup pour le soutien de la fin et rappelle toi que c'est pas si mal les Pyrénées), Mathieu (Geeze), Noé (j'ai hâte de partager avec toi des jours de plus, en plus), Emeline, Maxime R.M. (merci d'avoir fait le prof à ma place !), Jérome M, M'sieur Besnard (c'est fini cette thèse !), mon copain Pierrick, Pascal, Poussin, Yann, Pauline, Pedro et Carlotta.

Merci aussi Charlotte et Nathalie P. de m'avoir fait une petite place dans votre bureau pour que je puisse rédiger au calme.

Comme il n'y a pas que la thèse durant une thèse,

Merci aux collocs et intermittents de la « Maison du bonheur » pour tous ces bons moments : Flo et Caro, Jano, Sophie, Etienne, Audrey, Mathieu, Alex et Ufo, Alex-Casquette, Scan, Babas, Antoine, Nico, Anaïs, Bratou et Marco. Enfin, merci à toi Nini, ma « Meilleure amie », une amie comme on n'en fait plus. Merci de toujours tout comprendre et de croire toujours en moi et surtout quand c'est pas gagné d'avance.

Enfin, le dernier, mais non des moindres, merci à vous, papa et maman d'avoir toujours eu suffisamment confiance en moi pour me soutenir inconditionnellement dans tous mes choix. Merci de m'avoir répété inlassablement que l'on peut faire si l'on s'en donne vraiment les moyens. Vous avez fini par me convaincre…

Avant propos

Cette thèse en Écologie, biodiversité et évolution de l'École doctorale des Sciences Écologiques, Vétérinaires, Agronomiques et Bioingénieries (SEVAB) de l'Université Toulouse III a été financée par les dotations récurrentes du CNRS et de l'Université. Ce travail s'est déroulé au sein de l'équipe Diversification Ecologique et Evolutive des Populations du laboratoire Évolution et Diversité Biologique (UMR 5174 CNRS-UPS).

Le manuscrit est composé de trois parties : (i) une introduction générale, (ii) cinq articles scientifiques (un publié, un accepté avec modifications mineures et trois en préparation) et (iii) une discussion générale faisant le lien entre les différents résultats obtenus.

Sommaire général

Introduction générale

Sommaire de l'introduction générale

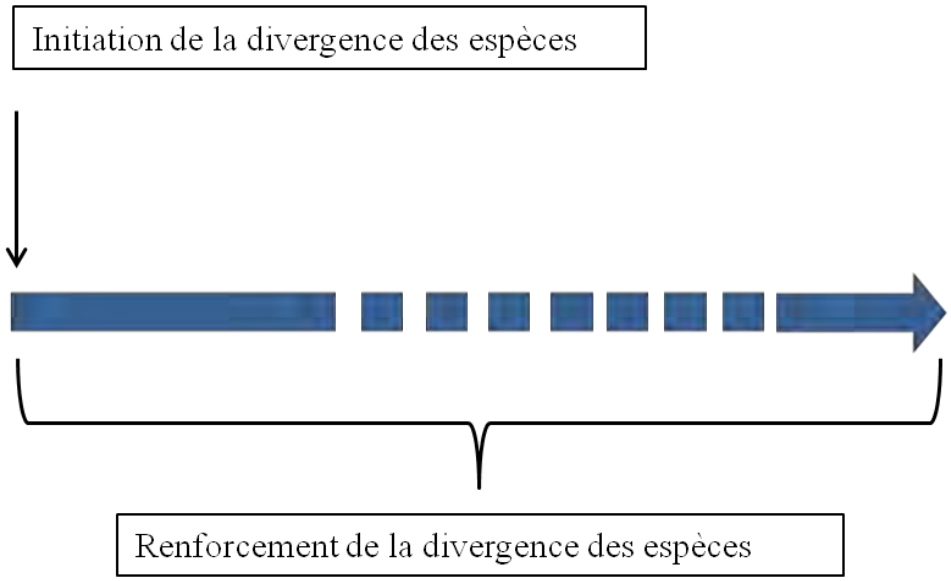
Introduction

Comprendre la mise en place et le maintien de la biodiversité (*i.e.* diversité des organismes vivants) sur Terre constitue, aujourd'hui encore, un des grands défis de la biologie. La biodiversité est observable à plusieurs niveaux (*e.g.* gènes, espèces, communautés, écosystèmes). L'enjeu de cette thèse est d'améliorer notre compréhension des processus impliqués dans la formation d'espèces (*i.e.* spéciation), ce qui requiert dans un premier temps de définir ce qu'est une espèce.

La définition de l'espèce était déjà débattue à l'époque de Darwin qui énonçait la difficulté de trouver une définition unique qui satisferait tous les naturalistes : « *Nor shall I here discuss the various definitions which have been given to the term species. No one definition has as yet satisfied all naturalists ; yet, every naturalist knows vaguely what he means when he speaks of a species »;* extrait de *De l'origine des espèces*. Plus de 150 ans après, la multitude des concepts d'espèces (pas moins de 20 concepts différents) confirme l'absence de consensus au sein des biologistes. La confusion autour du concept d'espèce est due au fait qu'en réalité, ces « concepts » n'en sont pas mais constituent plutôt des critères d'identification des espèces (Cai *et al.*, 2011; de Queiroz, 2007; Hey, 2006; Taylor *et al.*, 2000). Par exemple, le « Concept Biologique des Espèces » met l'accent sur l'isolement reproducteur complet entre les espèces, le « Concept Ecologique des Espèces » considère que deux espèces sont adaptées à des niches écologiques différentes tandis que le « Concept Morphologique des Espèces » repose sur la différenciation morphologique des deux espèces (voir (de Queiroz, 2007) pour une revue plus complète). La dimension temporelle du processus de spéciation varie selon les organismes et peut faire intervenir différents processus à différents moments, entrainant l'apparition des critères d'identification à des moments différents. Chercher à utiliser de manière universelle ces critères pour définir toutes les espèces semble de fait impossible (Giraud *et al.*, 2008).

Identifier les mécanismes évolutifs impliqués dans le processus dynamique de divergence entre espèces (e.g. différenciation morphologique, écologique ...) est crucial dans une époque de crise de la biodiversité. Ces mécanismes sont dynamiques. Nous avons choisi de définir les espèces comme des « segments de lignées évolutives qui ont

évolué indépendamment les unes des autres » (de Queiroz, 1998). Cela me parait plus approprié car cette perspective nous force à considérer le cadre temporel de la spéciation comme continu. Comme tout processus, la spéciation a donc un début correspondant à l'initiation de l'évolution indépendante de deux populations lors de la mise en place d'un isolement reproducteur (Figure 1). S'en suit alors une phase de renforcement de l'indépendance évolutive des deux populations donnant naissance à des lignées séparées via leur isolement reproducteur. Ici, nous considérons que le processus de spéciation n'a pas de stade final, puisque, à moins que les lignées ne s'éteignent, elles évoluent perpétuellement, même si leurs trajectoires évolutives sont indépendantes l'une de l'autre. Le « concept » biologique des espèces a longtemps prévalu dans l'étude de la spéciation. De manière logique, comprendre le processus de spéciations était alors réduit à l'étude de la mise en place de l'isolement reproducteur entre populations. C'est donc très naturellement que l'attention s'est historiquement portée sur l'effet des barrières géographique à la reproduction, et plus généralement sur le contexte spatial des populations divergentes.



<u>Figure 1</u> : Dynamique temporelle de la formation et du maintien des espèces adapté d'après (Butlin *et al.*, 2008)

1. Classification traditionnelle des modes de spéciation : un paradigme à remettre en question

1.1. Une seule classification pour deux cadres conceptuels

Du en grande partie aux travaux d'Ernst Mayr sur le « Concept Biologique des Espèces » (*i.e.* BSC pour « Biological Species Concept » (Mayr, 1842)), au début du vingtième siècle, le contexte biogéographique a été le paradigme dominant de l'étude de la spéciation. Les mécanismes de la spéciation ont de ce fait, historiquement, été définis selon la structure spatiale des populations divergentes: i) la spéciation allopatrique se déroulant dans un contexte d'isolement géographique complet entre les populations divergentes, ii) la spéciation parapatrique pour laquelle les populations divergentes sont soumises à un isolement géographique partiel et enfin iii) la spéciation sympatrique qui décrit l'absence d'isolement géographique entre les populations divergentes. Aujourd'hui, au cadre conceptuel biogéographique initialement prépondérant s'est rajouté celui de la génétique des populations. Ces deux perspectives cohabitent bien dans les cas de spéciation allopatrique et parapatrique. En effet, le cadre biogéographique définit la spéciation allopatrique comme un isolement géographique complet, en accord avec l'absence totale de flux de gènes entre les populations divergentes définie par la génétique des populations (Coyne, Orr, 2004; Gavrilets, 2003). La même cohérence entre le cadre biogéographique et génétique existe pour la spéciation parapatrique au cours de laquelle l'échange de gènes est restreint géographiquement puisque les populations marginales se rencontrent seulement dans des zones de contact ((Endler, 1977; Futuyma, Mayer, 1980; Smith, 1955); cadre conceptuel biogéographique). Par contre, pour ce qui est de la spéciation sympatrique, deux définitions distinctes, biogéographique et génétique, sont répandues. La spéciation sympatrique décrit la formation d'une nouvelle espèce en l'absence de ségrégation spatiale dans la population ancestrale, c'est-à-dire sans barrière physique et sans séparation géographique, même partielle (Coyne, Orr, 2004; Endler, 1977; Ridley, 1993). Le cadre de la génétique des populations décrit, quant à lui, l'origine d'une nouvelle espèce à partir d'une seule

population panmictique, c'est-à-dire en l'absence d'une quelconque limitation du flux de gène (Coyne, Orr, 2004; Gavrilets, 2003). Dans ce cas, il est souvent nécessaire d'évoquer la sélection divergente ou des processus démographiques complexes pour expliquer la divergence en présence de flux de gènes non limité. Un exemple classique de spéciation sympatrique est celui des palmiers de l'île de Lord Howe (Figure 2) (Savolainen *et al.*, 2006).

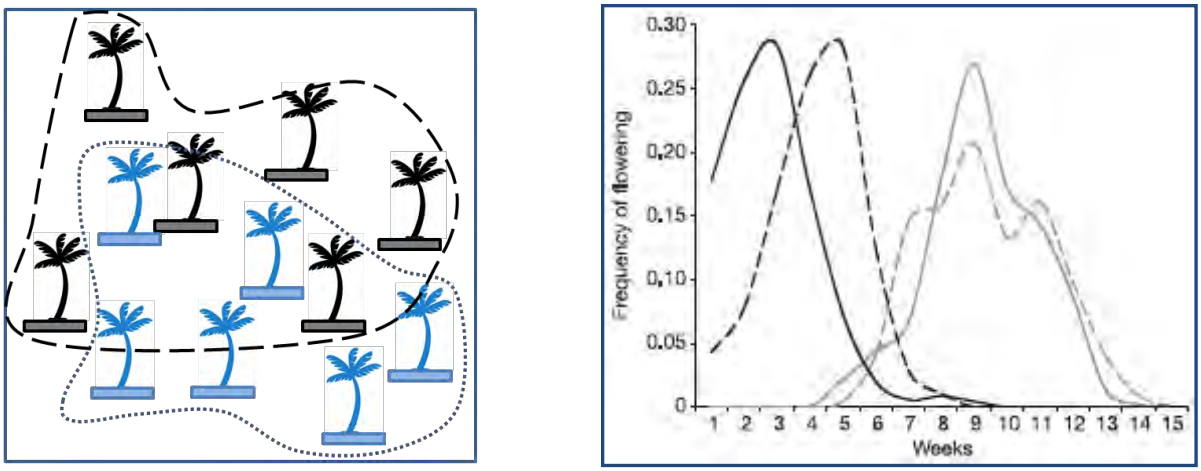


Figure 2: contexte géographique et écologique des populations de Palmier de l'île de Lord Howe (D'après (Savolainen *et al.*, 2006)). Malgré des aires de distribution sympatriques, les palmiers occupent des environnements différents (figure de gauche) ce qui conduit à au décalage des phénologies des palmiers (figure de droite).

Dans ce cas d'étude, bien que les deux espèces de palmier occupent des aires de distribution chevauchantes, leur préférence pour des types de sols différents (pH) a entrainé un isolement reproducteur temporel partiel car les dates des optima de floraison des deux espèces se sont décalées. Par conséquent, les deux espèces ont subi une limitation temporelle du flux de gènes. Les croisements entre populations divergentes n'ont donc pas eu lieu au hasard durant le processus de spéciation. La limite entre la définition stricte de la spéciation sympatrique et celle de la spéciation parapatrique (au sens génétique) est ici assez floue. Bien que de telles ambigüités n'entachent en rien l'importance de ces études pour la compréhension de la spéciation, elles ont néanmoins rendu la littérature sur ce sujet assez confuse. Il faut noter que ce n'est pas la perspective biogéographique qui a historiquement alimenté le débat autour

de l'existence de cas de spéciation sympatrique (voir Encadré 1) mais plutôt l'existence potentielle d'évènements de spéciation en présence de flux de gènes.

D'après la classification biogéographique de la spéciation, le contexte géographique des populations divergentes n'est considéré qu'en classes discrètes. Or, l'isolement géographique complet des populations (*i.e.* allopatrie), tout au long de la spéciation, semble rare parce que des migrants peuvent franchir des barrières parmi les plus extrêmes, comme illustré par la colonisation des îles océaniques (Butlin *et al.*, 2008). De la même manière, l'absence complète d'isolement géographique entre populations (*i.e.* sympatrie) est difficile à imaginer étant données les capacités de dispersion limitées et la distribution en patch des habitats favorables de la plupart des organismes (Endler, 1977). Ainsi, il semble peu probable que les distributions géographiques des populations divergentes soient maintenues dans des contextes aussi extrêmes (allopatrie et sympatrie) tout au long du processus de spéciation.

Encadré 1. Controverse autour de la spéciation sympatrique

La spéciation allopatrique a longtemps été considérée comme le seul mode de spéciation possible, car il est simple à appréhender. Partant d'une population ancestrale, la mise en place d'une barrière extrinsèque (principalement de nature physique) à la reproduction entraine l'isolement géographique de deux populations. Aucune migration n'est alors possible entre ces deux populations. En l'absence de l'effet homogénéisant du flux de gènes, les deux populations vont progressivement accumuler de manière neutre ou adaptative différentes mutations qui seront sources d'incompatibilités (Bateson, 1909; Dobzhansky, 1937; Muller, 1942) (Figure 3). Un isolement reproducteur intrinsèque va peu à peu s'installer entre ces deux populations qui vont évoluer de manière indépendante et diverger en deux espèces. Dans le modèle de spéciation allopatrique, l'isolement reproducteur intrinsèque entre deux populations est le sous-produit de leur différenciation génétique, induite par leur isolement géographique. En bref, après un temps suffisant, la spéciation est la conséquence inévitable de l'évolution de populations en allopatrie.

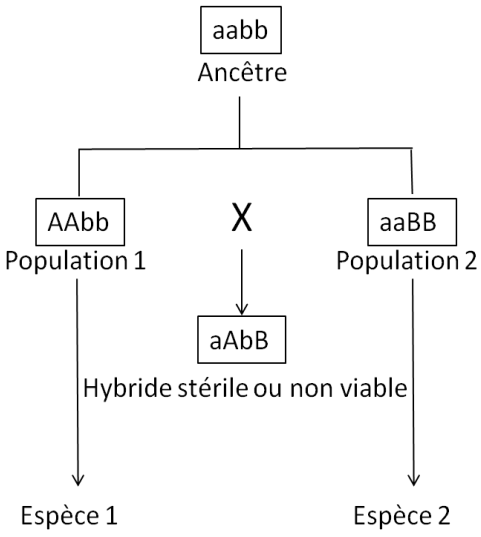


Figure 3: Modèle d'incompatibilité de Dobzhansky–Muller illustrant la diminution de la valeur sélective des hybrides entre deux espèces ayant divergé récemment.

Une espèce ancestrale portant à deux locus, les allèles *a* et *b* à l'état homozygote, donne naissance à deux lignées évoluant de manière indépendante. Dans la population 1, la mutation *A* apparait et est fixée alors que dans la population 2, une mutation *B* touchant l'autre locus apparait et est fixée. Les mutations *A* et *B*, n'ayant jusqu'à lors jamais coexisté au sein du même génome, peuvent ne pas être compatibles lorsqu'elles sont trouvées ensemble au sein d'un individu hybride.

Cependant, la raison pour laquelle la spéciation sympatrique a longtemps été controversée n'est pas la géographie (absence de ségrégation spatiale des populations divergentes) mais le flux de gènes. De nombreux biologistes (Theodosius Dobzhansky, Joseph Felsenstein, Ernst Mayr, *ect*…) ont longtemps été convaincus que des barrières intrinsèques à la reproduction ne pouvaient pas évoluer en l'absence de barrières extrinsèques à cause de l'action homogénéisante du flux de gène qui entrave l'effet diversifiant de la dérive. La formalisation de modèles théoriques a permis de mettre fin à cette controverse sur la vraisemblance

d'évènements de spéciation à partir d'une population ancestrale panmictique (flux de gène maximal). Il est maintenant couramment accepté que la sélection naturelle peut s'opposer à l'action homogénéisante du flux de gènes et de la recombinaison et ainsi conduire à la formation de populations distinctes qui seront à terme génétiquement isolées l'une de l'autre (Coyne, Orr, 2004; Gavrilets, 2003; Turelli *et al.*, 2001). Cette controverse a elle aussi entraîné un biais dans la littérature en faveur des études empiriques sur les cas de spéciation sympatrique, les auteurs ne voyant pas la nécessité de publier un énième cas de spéciation allopatrique. Si cette classification aurait pu permettre l'évaluation de la fréquence relative des différents modes de spéciation dans la nature, cela n'est même pas possible à partir de la littérature disponible.

1.2. Modification du contexte géographique des populations, rôle des dernières glaciations

Les aires de distribution des espèces ne sont pas immuables mais au contraire, sont très mobiles ; elles peuvent se déplacer, s'épandre ou au contraire se contracter (Brown *et al.*, 1996; Davis, Shaw, 2001; Gaston, 2003). Plusieurs études suggèrent que l'hypothèse de modification des aires de distribution en réponse à des changements environnementaux serait la plus fréquente (Davis, Shaw, 2001; Pease *et al.*, 1989). En réponse à ces variations géographiques d'aires de répartition régies par des changements environnementaux, la mise en jeu de processus de spéciation tels que ceux exposés ci-dessus est attendue.

La dernière période la plus marquante de changements environnementaux par bouleversement climatique date de l'époque Quaternaire (il y a 700 000 ans). Elle se caractérise par des oscillations climatiques où périodes glaciaires sèches et courtes périodes chaudes et humides se succèdent (Webb, Bartlein, 1992). Les registres fossiles témoignent des modifications d'aire de distribution de nombreuses espèces dues à ces oscillations (Tzedakis *et al.*, 2006). L'impact de telles oscillations sur la distribution des espèces est le mieux documenté pour le dernier cycle glaciaire (il y a environs 10 000 ans), avec la plupart des cas d'étude localisés en Europe ou en Amérique du Nord (Hewitt, 2003; Hewitt, 2004).

Des études comparatives ont révélé des impacts similaires au sein de certains groupes d'organismes. Les espèces de montagne adaptées au froid ont répondues par une expansion de leur distribution Une grande majorité d'espèces se sont déplacées et ont subi, au contraire, une réduction de leur aire de distribution durant la période glaciaire. Les espèces tempérées se sont réfugiées dans les régions de basses altitudes, plus au sud, où les conditions climatiques et les habitats étaient favorables à leur survie. C'est ensuite à partir de ces refuges que les espèces ont colonisé ou recolonisé de plus hautes latitudes et le nord de l'Europe. Ces cycles répétés de changement climatique ont façonné la distribution actuelle de la faune et de la flore.

Les dernières glaciations ont également eu un impact sur l'évolution de la diversité des espèces. Les cycles répétés de contraction/expansion des aires de distribution ont entrainé la perte de diversité génétique dans les régions nouvellement colonisées (Hewitt, 2000; Hewitt, 2003), de part les évènements répétés de goulot d'étranglement le long des routes de colonisation. Le retrait des espèces de certaines régions et leur survie dans des refuges méridionaux ont entrainé l'isolement géographique de certaines populations ce qui a favorisé des évènements de spéciation (Barton, Charlesworth, 1984; Hewitt, 1996). Enfin, les expansions post glaciaires et la recolonisation de certaines régions ont pu entrainer la remise en contact secondaire de lignées ayant commencé à diverger en allopatrie. Si les lignées en cours de divergence sont encore interfertiles, leur remise en contact se traduit par la formation de zones hybrides. L'existence d'un contact secondaire est souvent indiquée par la présence, dans la même zone géographique, de plusieurs zones hybrides entre différent couples d'espèces, appelées zones de sutures « suture zone », (Remington, 1968) (Avise, 2000; Hewitt, 2000; Hewitt, 1996; Hewitt, 1999; Kropf *et al.*, 2002; Redenbach, Taylor, 2002). Plusieurs hypothèses sont formulées pour expliquer les patrons communs de position de zones hybrides. Les zones de sutures jalonneraient les routes de colonisation, à mi-chemin entre deux zones de refuge glaciaire (Anderson, 1949; Avise, 2000; Remington, 1968). Elles pourraient aussi se rencontrer au pied de montagnes, suite à la dispersion des espèces le long de col, durant les périodes de réchauffement post-glaciaire (Remington, 1968; Swenson, Howard, 2005). Il faut tout de même noter ici que certaines zones de suture, notamment chez les plantes (Neuffer *et al.*, 1999), peuvent avoir une origine plus récente et résulter de perturbation anthropiques (« hybridization of the habitats », (Anderson, 1948)). En Europe, la position des zones de suture a souvent été

déduite suite à la caractérisation des routes de colonisation post-glaciaire reconstruites à partir des données phylogéographiques et paléobotaniques (Hewitt, 2000; Hewitt, 1999; Hewitt, 2001; Taberlet *et al.*, 1998). Deux zones de suture importantes, résultant de l'impact des dernières glaciations, ont ainsi été mises en évidence dans les massifs des Alpes (Taberlet *et al.*, 1998) et des Pyrénées (Guillaume *et al.*, 2000). Le devenir de ces zones hybrides et des espèces en formation dépend principalement des effets relatifs de la migration et de la sélection, laquelle peut être d'origine extrinsèque (dépendante des conditions écologiques) ou intrinsèque (dépendant uniquement des attributs spécifiques aux espèces) (Barton, Hewitt, 1985).

En bref : La classification biogéographique des modes de spéciation ne rend pas compte de l'aspect dynamique de la répartition géographique des populations qui influence la spéciation. Le contexte spatial des populations et celui du flux de gènes entre ces populations ne sont pas toujours évidents. La classification biogéographique de la spéciation permet de décrire le patron spatial des populations divergentes mais ne permet pas d'inférer les processus évolutifs impliqués dans la spéciation. Sans pour autant être ignoré, le contexte spatial ne devrait pas constituer, à lui seul, le critère majeur de classification des modèles de spéciation. L'étude du contexte géographique de la spéciation doit être accompagnée de l'étude des processus évolutifs dynamiques impliqués.

2. Les processus évolutifs impliqués dans la divergence des espèces

Le chapitre 1 illustre que les processus de spéciation peuvent se dérouler dans différents contextes spatiaux. De la même manière, différents mécanismes et forces évolutives peuvent conduire à la divergence de populations aboutissant à la formation de nouvelles espèces. Lors de la spéciation, les différents processus évolutifs à l'œuvre ne sont pas exclusifs. C'est ce que nous nous proposons d'aborder maintenant, en nous focalisant sur le rôle de l'écologie dans la divergence des espèces.

2.1. Lorsque l'écologie est le moteur de la divergence

La 'spéciation écologique' est définie comme « le processus au cours duquel l'isolement reproducteur entre populations évolue en réponse à une pression de sélection divergente exercée par l'écologie » (Dobzhansky, 1951; Funk, 1998; Rundle, Nosil, 2005; Schluter, 2001; Schluter, 2009). Cette définition souligne le rôle fondamental des facteurs écologiques à la base de pressions de sélection divergente. Il est important de noter que l'isolement reproducteur n'est pas forcément sélectionné directement, il peut être la conséquence de la différenciation génétique et phénotypique résultant de la sélection divergente. Il existe de multiples exemples de spéciation écologique dont de nombreuses études empiriques dans la nature (voir (Schluter, 2001)). Le processus de spéciation écologique s'appuie sur i) une pression de sélection divergente d'origine écologique, ii) l'évolution de l'isolement reproducteur et iii) un lien entre les deux.

2.1.1. Source écologique de sélection divergente

Les sources écologiques de sélection divergente ont ainsi été mises en évidence et répertoriées en trois catégories : i) différences environnementales, ii) sélection sexuelle et iii) interactions biotiques.

2.1.1.1. Différences environnementales à la base de la sélection divergente

L'adaptation locale des populations à leur environnement biotique (compétiteurs, prédateurs, etc.) ou abiotique (climat, ressources, etc.) peut être replacée dans un contexte de niche écologique divergente amenant à l'évolution de différences morphologiques, physiologiques ou comportementales entre individus de différentes populations. Le rôle de l'environnement abiotique, et plus particulièrement celui des ressources, est documenté. Il a notamment été au centre des premières études expérimentales de spéciation écologique (Rice, Hostert, 1993; Rice, Salt, 1988; Rice, Salt, 1990). Ces premières études démontraient que des colonies de bactéries se sont adaptées localement à des milieux de culture constitués de différentes sources de carbone. Le rôle de l'environnement abiotique comme initiateur de la sélection divergente a depuis été mis en évidence de multiples fois dans des populations naturelles (Schluter, 2000). Un exemple connu chez les plantes concerne les formes continentales et côtières de l'espèce *Mimulus guttatus* qui sont soumises à différentes conditions abiotiques favorisant des périodes de croissance et de floraison différentes (Hall, Willis, 2006; Lowry *et al.*, 2008). Leur adaptation locale a entrainé le décalage temporel de leur phénologie.

2.1.1.2 La sélection sexuelle comme source écologique de sélection divergente

Les animaux communiquent en émettant une grande variété de signaux et bon nombre d'études suggèrent que la divergence dans les signaux de communication est un facteur pouvant favoriser la spéciation (Boughman, 2002). L'hypothèse de l'entrainement sensoriel, « sensory drives hypothesis», a retenu une attention considérable en tant que mécanisme favorisant la diversification des signaux de communication entre espèces proches (Endler, 1992). Par exemple, les lézards du genre *Anolis* provenant de deux environnements distincts (xérique et mésique) possèdent un fanon gulaire (membrane de peau sous la gorge) pouvant être déployé par les mâles pour attirer les femelles. La coloration des fanons gulaires a divergé entre populations vers une augmentation de la détection du signal par les femelles spécifique à chaque habitat (Leal, Fleishman, 2002). Chez les plantes, la sélection sexuelle *via* le comportement des pollinisateurs peut aussi

entrainer la divergence morphologique des populations. Par exemple, les deux espèces vivant en sympatrie *Mimulus cardinalis* et *M. lewisii* présentent des morphologies florales différentes adaptées pour chacune à la morphologie de leurs différents pollinisateurs (Ramsey *et al.*, 2003; Schemske, Bradshaw, 1999).

2.1.1.3 Interactions écologiques à la base de la sélection divergente

Les interactions écologiques, telles que la compétition, sont distinguées des autres sources de sélection divergente car elles nécessitent un contact entre les populations divergentes. Des études menées en laboratoire sur des bactéries ont mis en évidence que la compétition fréquence-dépendante peut entrainer la différenciation écologique en sympatrie (Friesen *et al.*, 2004). En population naturelle, les tests directs du rôle de la compétition dans le processus de sélection divergente sont peu fréquents. Lorsque des populations partagent la même niche écologique et que les ressources y sont limitantes, ces populations vont entrer en compétition. Cette compétition peut s'effectuer sur l'acquisition des ressources et/ou la reproduction via l'accès aux partenaires reproducteurs chez les animaux, comme chez les plantes. Ce type de compétition peut favoriser le partitionnement ou le déplacement de la niche entre les deux populations et leur divergence phénotypique (Armbruster *et al.*, 1994; Bolnick, 2004; Miller, 1967).

2.1.2 L'isolement reproducteur

L'isolement reproducteur peut être classé selon qu'il prend place au stade pré-zygotique (*i.e.* avant la formation du zygote) ou post-zygotique (*i.e.* après la formation du zygote).

2.1.2.1 L'isolement reproducteur pré-zygotique

L'isolement pré-zygotique se produit lorsque les partenaires reproducteurs potentiels sont séparées dans l'espace (isolement des habitats, (Dres, Mallet, 2002; Funk *et al.*, 2002)) ou dans le temps (isolement temporel (Lamont *et al.*, 2003; Wood, Keese, 1990)). Les populations géographiquement ou temporellement isolées vont se reproduire peu entre elles (Funk *et al.*, 2002; Johnson *et al.*, 1996) C'est le cas, par exemple, des

populations d'insectes herbivores, comme le puceron du pois, qui ont des préférences différentes vis-à-vis de la plante hôte dont ils se nourrissent. Chaque population va se reproduire sur la plante hôte dont elle se nourrit, les deux populations vont donc être isolées géographiquement sur des plantes hôtes différentes (Via, 1999). L'isolement temporel se produit lorsque les populations présentent des adaptations physiologiques différentielles qui vont conduire à des phénologies de reproduction décalées dans le temps. C'est le cas de l'exemple des populations de l'espèce *Mimulus guttatus,* présenté plus haut, qui ont évolué des périodes de floraison partiellement chevauchantes en réponse aux conditions abiotiques différentes. Il s'agit d'un isolement reproducteur partiel dans le temps (Lowry *et al.*, 2008).

Lorsque l'adaptation locale réduit la probabilité de rencontre des partenaires reproducteurs, elle peut favoriser l'isolement pré-zygotique. Dans le cas d'espèces interfertiles de papillon du genre *Heliconius* (*H. melpomene* et *H. cydno)*, les migrants ayant des patrons de coloration inhabituels en comparaison de la population native ne sont pas reconnus par les prédateurs (vraisemblablement des oiseaux) comme un signal de danger et subissent donc une plus forte pression de prédation que dans leur milieu natif (Mallet, 1989; Mallet, Barton, 1989). En conséquence, la probabilité d'évènements de reproduction inter-spécifique est réduite (Funk, 1998; Nosil, 2004; Nosil *et al.*, 2005; Via *et al.*, 2000).

2.1.2.2 L'isolement reproducteur post-zygotique

L'isolement reproducteur post-zygotique évolue lorsque la valeur sélective des individus issus de croisements entre groupes identifiés est diminuée (en comparaison à celle des individus parentaux) parce que leur phénotype n'est pas adapté. A l'issue de croisements hybrides, les hybrides peuvent présenter dès les premiers stades d'hybridation (F1, F2) des phénotypes qui ne sont pas vraiment adaptés aux niches environnementales de leurs parents (Coyne, Orr, 2004; Rice, Hostert, 1993). Si ces phénotypes sont intermédiaires et en l'absence de niche environnementale intermédiaire, les hybrides seront contre-sélectionnés ce qui réduit les échanges de gènes entre populations parentales. Ce mécanisme de sélection environnementale à l'encontre des hybrides est équivalent au mécanisme de contre-sélection des migrants, à

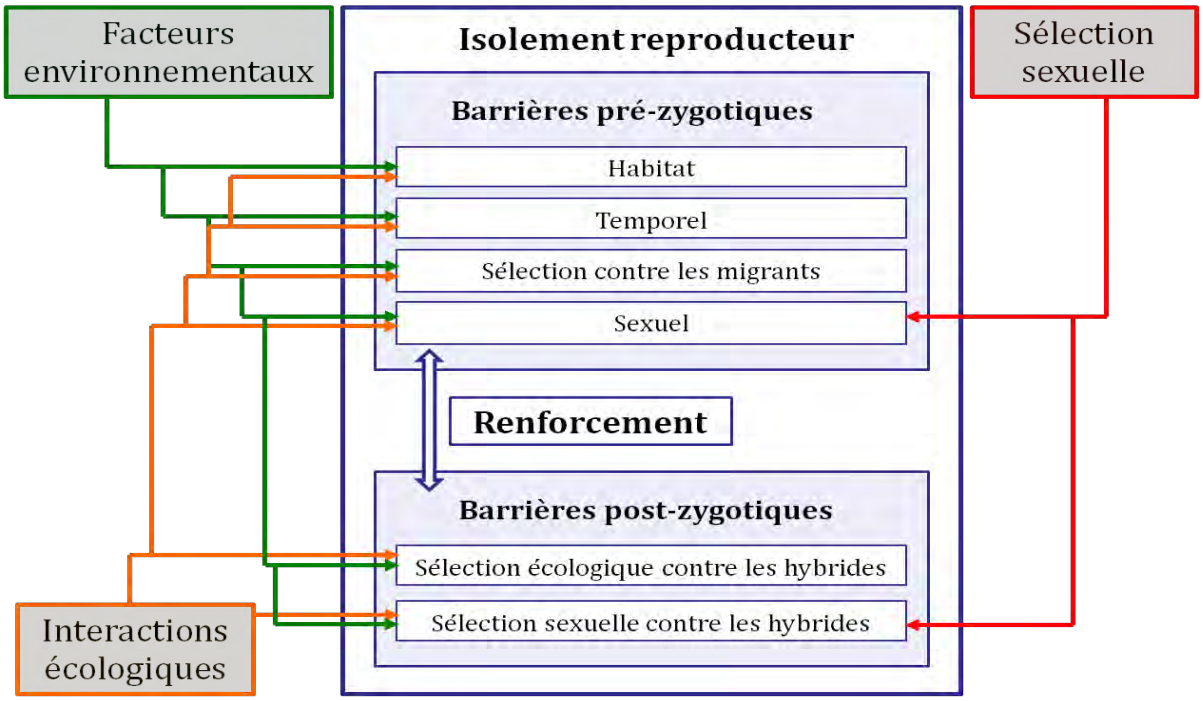
la différence qu'il n'entraine pas un isolement pré-zygotique, mais post-zygotique. Chez les plantes, la discrimination exercée par les pollinisateurs à l'encontre des phénotypes intermédiaires des hybrides peut avoir les mêmes conséquences d'isolement post-zygotique (Emms, Arnold, 2000; Schemske, Bradshaw, 1999).

2.1.2.3 La combinaison des deux mécanismes

Il est important de noter que les deux mécanismes d'isolement—pré-zygotique et post-zygotique—ne sont pas exclusifs. Chez *Heliconius* par exemple, la contre-sélection des migrants due à des patrons de coloration défavorables dans des environnements de prédation non-natifs s'ajoute à la contre-sélection des hybrides de par leurs patrons de coloration intermédiaires (Mallet, 1989; Mallet, Barton, 1989).

Parce que l'isolement postzygotique va contre-sélectionner les individus issus de croisements interspécifiques, il va réciproquement sélectionner positivement les individus qui se croisent au sein de la population native, et ainsi favoriser les mécanismes d'isolement prézygotique. Ce processus de renforcement de l'isolement pré-reproducteur entre populations divergentes (Servedio, Noor, 2003) est connu au sens large sous le nom de renforcement (Dobzhansky, 1937). Les caractères phénotypiques peuvent à fois être impliqués dans l'isolement post-reproducteur en étant soumis à des pressions de sélection divergentes (contre-sélection des hybrides et des migrants) et dans l'isolement pré-reproducteur (choix du partenaire) en favorisant les appariements intra-spécifiques (i.e. homogames) au détriment des appariements interspécifiques (i.e. hétérogames) (Jiggins *et al.*, 2001).

La figure ci-dessous synthétise les relations entre sources écologiques de sélection divergente et nature de l'isolement reproducteur engendré. Des pressions de sélection divergente indépendantes peuvent être à l'origine d'un même type d'isolement reproducteur. Inversement, un type de sélection divergente peut être impliqué dans l'évolution de différents types d'isolement reproducteur.



2.1.3 Lien génétique entre sélection divergente et isolement reproducteur

La dernière composante du processus de spéciation écologique à considérer est le mécanisme génétique par lequel la sélection divergente modèle l'évolution d'un trait écologique tout en ayant un impact sur les gènes impliqués dans l'isolement reproducteur. Un tel lien peut être direct ou indirect.

2.1.3.1 Sélection directe, épistasie et pléiotropie

Dans un contexte d'adaptation locale des populations ou des espèces, c'est-à-dire lorsqu'un phénotype a une bonne valeur sélective dans son environnement natif alors

qu'il a une mauvaise valeur sélective dans un environnement non natif (Figure 4), il est attendu que la valeur sélective des migrants soit de fait relativement moins bonne.

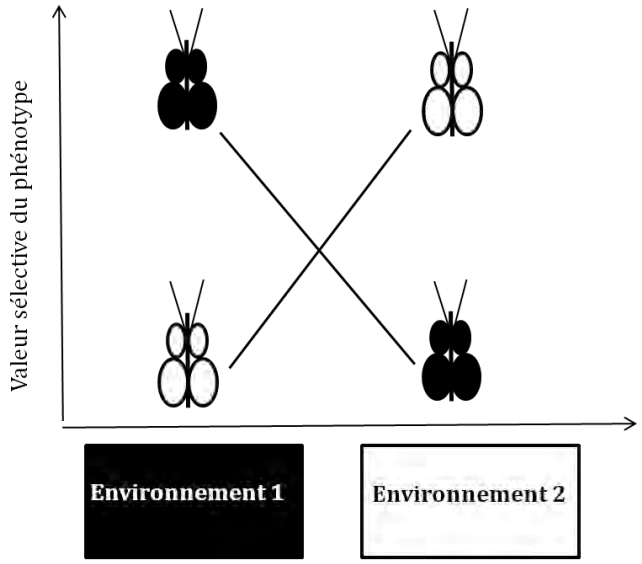


Figure 4 : Contre sélection environnementale des migrants.

Dans l'environnement le phénotype noir est adapté alors qu'il est mal adapté dans l'environnement 2. La relation est inverse pour le phénotype blanc .

La moins bonne viabilité des hybrides peut être due à la rupture par la recombinaison de la relation épistatique liant un complexe de gènes coadaptés à l'intérieur du génome (Presgraves *et al.*, 2003).

Les gènes sous sélection divergente peuvent aussi être impliqués dans l'isolement reproducteur de manière pléiotropique. Par exemple, chez les plantes du genre *Mimulus*, l'isolement reproducteur est une conséquence directe de l'adaptation à différents pollinisateurs, avec *Mimulus lewisi* majoritairement pollinisée par des bourdons et *Mimulus cardinalis* majoritairement pollinisée par des colibris (Schemske, Bradshaw, 1999). La couleur des fleurs, impliquée dans l'attraction des pollinisateurs, est principalement contrôlée par un seul locus (*YUP*) dont les différents allèles attirent différemment les deux types de pollinisateurs. De la même manière, l'isolement temporel engendré par le décalage des phénologies peut résulter d'effets pléiotropiques de l'adaptation à certaines conditions environnementales (Macnair, Christie, 1983). Ici le trait sélectionné et le trait impliqué dans l'isolement reproducteur impliquent le même gène. Cependant, des gènes différents peuvent aussi être impliqués par sélection indirecte.

2.1.3.2 Sélection indirecte et déséquilibre de liaison

Deux gènes différents peuvent être impliqués, le premier soumis à la sélection divergente étant directement impliqué et le deuxième impliqué dans l'isolement reproducteur étant indirectement sélectionné car étant physiquement lié au premier. L'intensité de la liaison entre les deux gènes dépend de la distance physique qui les sépare sur le chromosome, puisque la probabilité de recombinaison entre les deux gènes est d'autant plus faible que les deux gènes sont physiquement proches. Parce que l'association entre le gène sous sélection et le gène causant l'isolement reproducteur n'est pas parfaite, la sélection sera en conséquence, allégée sur ce dernier. La sélection indirecte est donc moins effective que la sélection directe dans la mise en place de l'isolement reproducteur (Kirkpatrick, Barton, 1997).

2.2 Lorsque l'écologie n'est pas le moteur de la divergence

De nombreux modèles dits « non-écologique » font intervenir des mécanismes et processus autres que ceux apparentés à la spéciation écologique et ont le potentiel de mettre en place un isolement reproducteur majeur : la polyploïdisation, la sélection uniforme, etc. sont discutés ci dessous.

2.2.1 Contexte de polyploïdisation

La polyploïdisation (i.e. duplication du génome entier), qu'elle soit autopolyploïde (au sein de la même espèce) ou allopolyploïde (entre deux espèces différentes), résulte de la fusion de gamètes n'ayant accidentellement pas subi de méiose et contenant le même nombre de chromosomes que les cellules somatiques. Les croisements entre individus de ploïdie différentes produisent des hybrides de valeur sélective moindre comparée aux hybrides issus de croisements entre individus de même ploïdie, probablement à cause de discordances entre la ploïdie de l'embryon et celle de l'endosperme (Ramsey, Schemske, 1998). La polyploïdisation peut entrainer la mise en place instantanée de l'isolement reproducteur post-zygotique, sans intervention de l'écologie. De nombreux auteurs considèrent la spéciation par polyploïdisation comme un mode de spéciation « non-écologique » car l'isolement reproducteur n'évolue pas en réponse à des pressions

de sélection divergentes (Rundle, Nosil, 2005; Schluter, 2000; Schluter, 2001). A l'inverse, certain auteurs (Sobel *et al.*, 2010) considèrent que la spéciation par polyploïdie est « non-écologique » seulement dans le cas où l'on considère l'émergence d'une nouvelle espèce de suite après la formation du néo-polyploïde (dont la ploïdie diffère de celles de ses progéniteurs) sans se soucier de sa capacité à persister. Les croisements entre différents niveaux de ploïdie conduisent souvent à la formation d'hybrides stériles. Le succès reproducteur du néo-polyploïde sera faible tant qu'il demeure peu fréquent par rapport aux cytotype majoritaire parental. Dans ces systèmes polyploïdes, le succès reproducteur fréquence-dépendant a été décrit comme le désavantage du cytotype minoritaire (Levin, , 1975). Selon *Ramsey et al.* (2002), le désavantage du cytotype minoritaire qui touche les néo-polyploïdes peut être compensé par des processus neutres tels que i) la stockasticité démographique qui conduirait le néo-polyploïde a devenir le cytotype majoritaire, ii) la migration qui permettrait au néo-polyploïde de coloniser une nouvelle région et de rompre le flux de gène avec les individus parentaux ou iii) l'augmentation de l'autofécondation. Il peut arriver que les néo-polyploïdes aient une meilleure valeur sélective dans l'environnement parental et co-existent voire remplacent le cytotype parental mais dans ce cas, ce type de spéciation s'apparente à un cas de spéciation écologique.

2.2.2. Contexte de sélection uniforme : spéciation par accumulation de mutations

En réponse à des pressions de sélection similaires (i.e. sélection uniforme), des populations géographiquement isolées peuvent acquérir des phénotypes similaires en fixant des mutations différentes. Au gré du hasard, les populations peuvent fixer des mutations adaptatives différentes à un même locus ou sur des loci différents et dans un ordre différent (d'où le nom de « mutation order speciation »). Ces mutations différentes peuvent causer des interactions négatives entre ces allèles dans les génomes des hybrides, entrainant la non-viabilité ou la stérilité des hybrides (modèle d'incompatibilités de Dobzhansky-Muller, voir Figure 3) et conduire à la mise en place d'un isolement post-zygotique entre les deux populations.

La réponse adaptative différentielle des populations face à une pression de sélection uniforme a pu être observée lors d'expériences de sélection artificielle en

laboratoire. Des lignées identiques d'*Escherichia coli* ont été contraintes à évoluer sur un milieu limité en glucose et contenant du citrate que ces bactéries ne sont pas capables d'utiliser comme source de carbone, dans les conditions de l'expérience (Blount *et al.*, 2008). Durant 31 500 générations et malgré une pression de sélection uniforme, les lignées répliquées d'*E. coli* ont développé des mutations différentes dont certaines ont permis l'utilisation du citrate comme source de carbone. En conditions naturelles, les cas de spéciation résultants de conflits intragénomiques, tels que la stérilité cytoplasmique mâle, sont vraisemblablement des cas de « spéciation due à l'ordre des mutations » (Case, Willis, 2008). La spéciation par sélection sexuelle est aussi considérée comme un cas de « spéciation due à l'ordre des mutations » dans le cas où la divergence des préférences dans le choix du partenaire se produisent par la fixation de différentes mutations avantageuses dans différentes populations, comme c'est le cas dans le cadre de conflits sexuels au cours desquels les intérêts évolutifs des mâles diffèrent de ceux des femelles (Rice, 1998). La « spéciation due à l'ordre des mutations » est difficilement envisageable en présence de flux de gènes entre les populations divergentes. En effet, si l'on imagine deux allèles incompatibles mais fournissant tous deux un avantage sélectif dans chaque population : En présence de flux de gènes ces deux allèles vont diffuser librement entre les deux populations et parce qu'un des deux allèles fournira un avantage sélectif légèrement supérieur à l'autre il sera, à terme, fixé dans les deux populations. Ce modèle de spéciation sous-tend donc, au moins durant la phase d'initiation de l'isolement reproducteur, un contexte d'isolement géographique complet des populations (Nosil, Flaxman).

Bien que l'écologie puisse être la source de la sélection uniforme, la spéciation due à l'ordre des mutations n'est couramment pas considérée comme un cas de spéciation écologique car l'écologie n'est pas, en elle-même, la source de la divergence ((Schluter, 2009), mais voir (Sobel *et al.*, 2010)).

2.3. Processus de divergence non-écologique : spéciation par dérive génétique

Même si les facteurs susceptibles de modifier l'impact de la dérive, et donc la fixation de mutations tels que la variation de la taille des populations, peuvent être influencés par des facteurs écologiques, la spéciation par dérive n'est pas considérée comme spéciation

écologique. Le modèle de spéciation par dérive est facile à appréhender : par le simple fait du hasard, des mutations différentes sont fixées dans les populations allopatriques et peuvent par la suite entrainer l'isolement reproducteur complet des populations. Néanmoins, une série d'arguments a été développée contre la vraisemblance de processus de spéciation sous l'action unique de la dérive (Coyne, Orr, 2004; Turelli *et al.*, 2001). Parmi ces arguments, on peut relever que la plupart des traits potentiellement impliqués dans l'isolement reproducteur des populations sont vraisemblablement aussi sous sélection naturelle (e.g. période de floraison pour des plantes, patrons de coloration chez les animaux …). De ce fait, il semble peu probable que la dérive génétique puisse remanier ces traits sans causer de dommage sérieux à la survie des individus. Un autre argument contre l'action de la dérive génétique seule dans les processus de spéciation est celui du temps nécessaire pour que, par dérive exclusivement, les populations deviennent isolées reproductivement (Nei, Chesser, 1983). En lien avec cette question se pose le problème du contact secondaire entre populations pour lesquelles l'isolement reproducteur n'est pas complet. Ces arguments ont été documentés graduellement dans la littérature scientifique et ont peu à peu jeté le doute sur la significativité d'un impact possible de la dérive seule dans la spéciation. De plus, les études menées en laboratoires tentant de reconstruire des évènements de spéciation par effet fondateur ont, à ce jour, échoué dans la mise en place d'un isolement reproducteur (Rundle, 2003; Rundle *et al.*, 1998).

En bref, les facteurs écologiques ne sont pas forcément la source de la sélection divergente dans les cas de spéciation mais ils semblent intervenir dans de nombreux processus mettant en place l'isolement reproducteur, même lorsque ceux-ci ne sont pas par définition considérés comme des cas de spéciation écologique. La classification dichotomique sur le modèle processus écologique versus processus non-écologique semble assez peu fidèle à la réalité des processus dynamiques et rarement indépendants mis en jeu lors de la spéciation. Identifier la spéciation par un seul type de mécanisme revient à considérer qu'un seul type mécanisme est impliqué dans l'ensemble du processus de spéciation. Or, les mécanismes de mise en place de l'isolement reproducteur sont rarement exclusifs. De plus, différents processus ou combinaisons de processus peuvent être impliqués à différents stades de la spéciation.

Le processus de spéciation est dynamique et ne peut que difficilement être associé à une définition figée, définie sur la base d'un mécanisme ou processus évolutif ou d'un cadre géographique unique. Ces paramètres ne sont ni exclusifs, ni figés dans le temps comme le laisse entrevoir ces définitions. La spéciation passe par une phase dite de « renforcement de la divergence » qui est longue et dont la progression n'est pas linéaire. Dans ce cadre dynamique, le contexte géographique tout comme les mécanismes et les processus évolutifs mis en jeu peuvent changer au cours du temps et venir entraver ou au contraire renforcer la mise en place de l'isolement reproducteur.

Les dichotomies traditionnellement appliquées pour décrire les processus de spéciation (i.e. allopatrie versus sympatrie, écologique versus non-écologique …) sont restrictives et ne permettent pas de révéler la réalité complexe du processus. Les mécanismes impliqués dans la mise en place de l'isolement reproducteur et le contexte géographique dans lequel ils se déroulent sont susceptibles d'interagir les uns avec les autres et de varier au cours du processus de spéciation (i.e. au cours des phases d'initiation de l'isolement reproduction, de renforcement de l'isolement reproducteur). Afin de garantir une meilleure compréhension de la spéciation, il est indispensable d'intégrer les différentes composantes de ce processus – isolement spatial, écologique et reproducteur – dans un cadre dynamique.

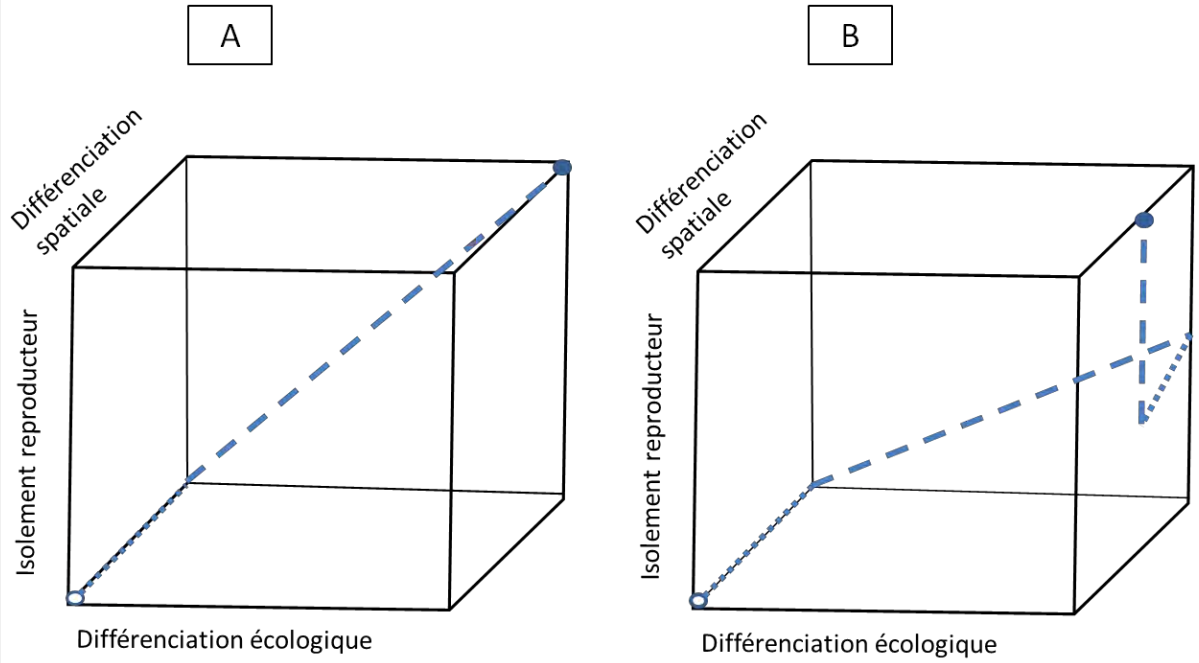
3. Objectifs de la thèse

Afin de mieux comprendre la spéciation, il faut prendre en compte la dynamique temporelle de ce processus et étudier l'interaction des différentes forces évolutives à l'œuvre. Dans ce but, un nouveau cadre conceptuel a été proposé par (Dieckmann *et al.*, 2004) qui décrit le processus de spéciation suivant un volume à trois dimensions orthogonales : i) le contexte spatial des populations, ii) leur contexte écologique et iii) leur isolement reproducteur (Encadré 2).Cette approche permet de reconstruire les « routes de spéciation » empruntées par les différents organismes.

La route de spéciation commence au coin inférieur gauche du cube, où les populations ne sont pas isolées spatialement ou écologiquement et se reproduisent au hasard (Encadré 2). La route de spéciation se construit au cours du temps par un déplacement en réponse à une modification extérieure (e.g. vicariance, contact secondaire…), par l'action de la dérive génétique, de la sélection naturelle ou sexuelle. A terme, la route de spéciation aboutit au plateau supérieur du cube qui correspond à un isolement reproducteur complet entre les populations (*i.e.* espèces) auquel est associé un certain degré d'isolement spatial et écologique. L'aspect intéressant de cette approche réside dans le fait qu'elle n'essaie pas de réduire la complexité des processus et de leurs interactions en les classant de manière exclusive et dichotomique. Au contraire, cette approche vise à intégrer les différents processus impliqués dans la formation des espèces.

Tout en s'inscrivant dans cette perspective, l'objectif de ce travail de thèse à été de déterminer les processus impliqués dans la divergence en cours de deux sous-espèces de plantes à fleurs, *Antirrhinum majus pseudomajus* et *Antirrhinum majus striatum.*

ENCADRÉ 2 : Exemples de routes de spéciation



La figure A représente un scénario classique de spéciation allopatrique dans lequel un évènement extérieur entraine l'isolement géographique des deux populations divergentes, entrainant un niveau élevé d'isolement spatial (ligne en pointillés). L'adaptation locale des populations à leurs environnements respectifs va entrainer, par des effets pléiotropiques sur le choix du partenaire, l'augmentation concomitante de l'isolement écologique et reproducteur.

La figure B représente un scénario de spéciation à deux phases où la divergence en allopatrie sera suivie par une phase de renforcement lors d'un contact secondaire. L'isolement géographique et l'adaptation locale entrainent la différenciation écologique et l'isolement reproducteur partiel des populations divergentes (comme dans le scénario A). S'ensuit une phase de contact secondaire (ligne en pointillés à droite) entrainant un renforcement de l'isolement reproducteur (ligne en tirets verticale).

4. Modèle de l'étude

Le genre *Antirrhinum* compte une vingtaine d'espèces et sous-espèces étroitement apparentées (Rothmaler, 1956; Sutton, 1988) distribuées de l'ouest de l'Europe au nord de l'Afrique avec une diversité maximale concentrée dans la péninsule Ibérique (Figure 5).

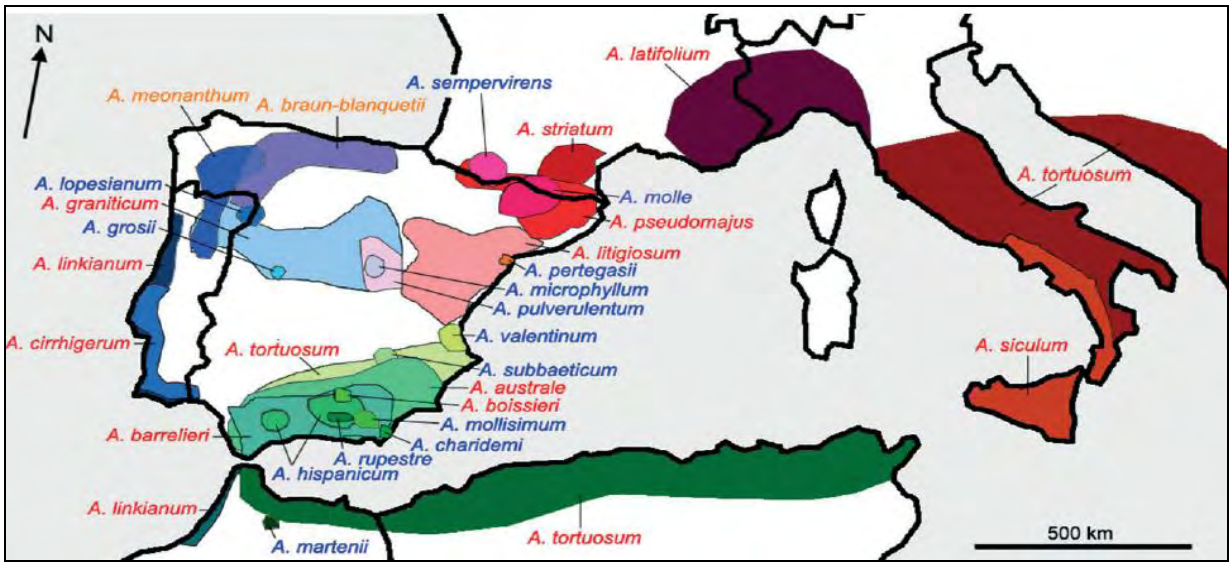


Figure 5: Distribution des espèces du genre *Antirrhinum*

Extraite de l'article de (Wilson, Hudson)

Les espèces du genre *Antirrhinum* sont des herbacées semi-pérennes arborant des fleurs zygomorphes arrangées en racèmes. Les cinq pétales qui constituent la corolle sont soudés à leur extrémité pour former un tube à cinq lobes (Figure 6). Malgré sa grande diversité phénotypique et écologique (Langlade *et al.*, 2005), le genre *Antirrhinum* semble avoir une origine relativement récente. La divergence des espèces du genre *Antirrhinum* pourrait avoir débuté il y a 5.3 à 3.7 millions d'années, avec la plupart des espèces étant apparues il y a moins d'1 million d'années (Gubitz *et al.*, 2003). L'hypothèse de cette radiation récente est confirmée par l'absence d'isolement reproducteur complet entre la majorité des espèces (excepté *A. siculum*) qui donnent naissance à des hybrides fertiles à la suite de croisements artificiels (Wilson, Hudson,

2011). La possibilité d'hybridation entre ces espèces a entrainé un débat sur le rôle de l'hybridation et de l'introgression dans la diversification du genre et a partiellement contribué à l'histoire évolutive complexe des taxons de ce genre (Sutton, 1988).

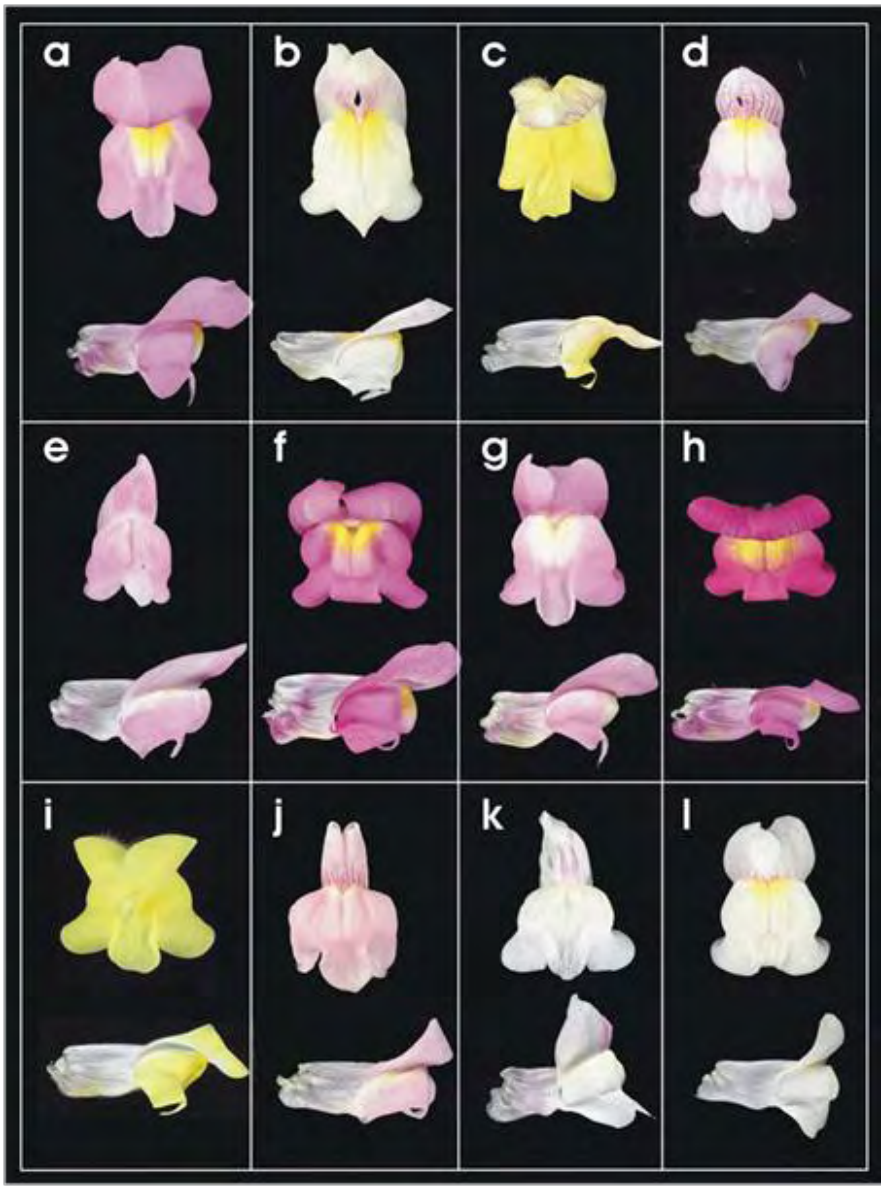


Figure 6: Diversité des phénotypes floraux de quelques espèces du genre *Antirrhinum*

a) *Antirrhinum australe;*

b*) A. barrelieri;*

c*) A. braun-blanquetti;*

d) *A. charidemi;*

e*) A. hispanicum;*

f*) A. majus ssp. Linkianum*

g*) A. majus ssp. litigiosum;*

h*) A. majus ssp. pseudomajus;*

i*) A. majus ssp. striatum;*

j*) A. mollissimum;*

k*) A. pertegassi;*

l*) A. pulverentum.*

Source Whibley 2004

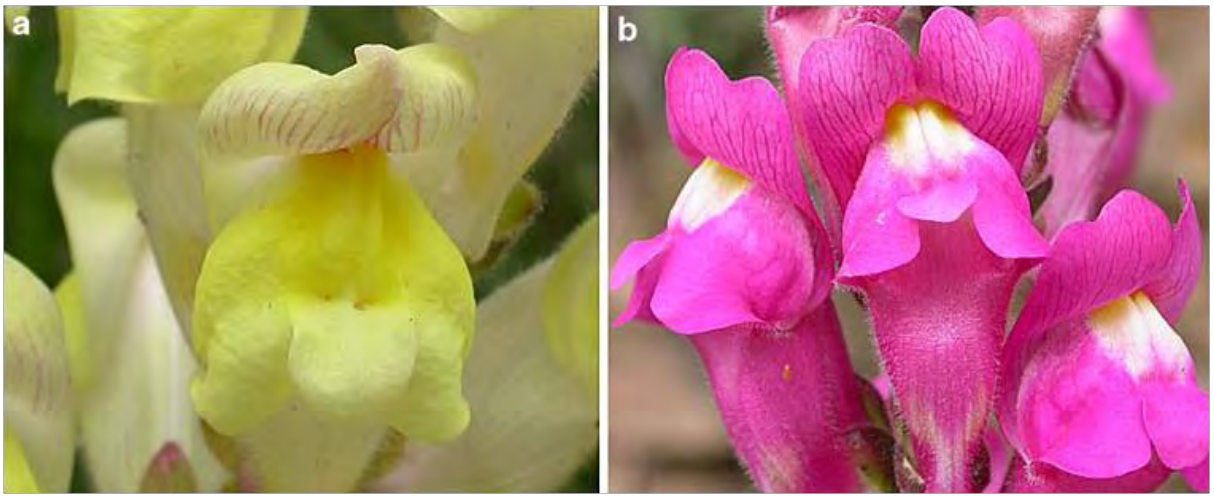


Figure 7 : Phénotypes floraux des deux sous-espèces d'*Antirrhinum majus*

a) *A. m. striatum* à fleurs jaunes et b) *A. m. pseudomajus* à fleurs magenta

Parmi toutes ces espèces, *Antirrhinum majus* est depuis longtemps utilisée en tant qu'espèce modèle en génétique du développement des plantes (Schwarz-Sommer *et al.* 2003) et notamment dans l'étude des voies de biosynthèse des anthocyanes, qui sont des pigments responsables de la couleur des fleurs (Schwinn, 1999). Des gènes responsables de la forme de la fleur (*DICHOTOMA*) ainsi que des gènes de régulation intervenant dans la voie de biosynthèse des anthocyanes (*PALLIDA* et *ROSEA*) ont été isolés (Schwinn, 1999). Dans le cadre de cette thèse, nous nous sommes plus particulièrement intéressés à deux sous espèces, *Antirrhinum majus pseudomajus* et *A. m. striatum* qui présentent des morphologies florales et végétales similaires ne différant que par la couleur de la corolle, magenta chez *A. m. pseudomajus* et jaune chez *A. m. striatum* (Figure 7). L'accès à l'intérieur de la corolle est fermé et nécessite que le pollinisateur exerce, avec son propre poids, une pression sur la lèvre inférieure de la corolle, la contraignant ainsi à s'ouvrir (Figure 8). Les insectes de petite taille ne sont pas assez lourds pour ouvrir la corolle et ne sont donc pas, *a priori*, des agents de pollinisation. La forme de la fleur est supposée être une adaptation à la morphologie des pollinisateurs qui sont donc majoritairement de grande taille, tels que les bourdons (*Bombus ssp.*) ou les abeilles charpentières (*Xylocopa ssp.*). De part leur anatomie, et parce qu'elles sont généralement auto-incompatibles, ces espèces sont dépendantes des pollinisateurs pour le transfert de pollen lors de la reproduction.

A/

B/

Figure 8: Visite d'un bourdon pollinisateur d'une fleur d'*Antirrhinum.*

Le bourdon (*Bombus hortorum*) atterrit sur le lobe inférieur de la corolle,

A/ en faisant pression avec son poids il ouvre la corolle et pénètre à l'intérieur de la fleur,

B/ à l'intérieur de laquelle il aura accès au nectar situé à la base de la corolle et entrera en contact avec les anthères et le stigmate avec la surface dorsale de son thorax.

Source (Whibley, 2004)

Ces deux sous-espèces sont géographiquement distribuées de manière parapatrique du sud-ouest de la France au nord-est de l'Espagne (Figure 9). Bien que la parapatrie constitue le mode le plus commun de distribution des espèces proches, la distribution de ces deux sous-espèces est originale dans le sens où contrairement à ce qui a été décrit dans la littérature (Vargas *et al.*, 2009; Wilson, Hudson, 2011), l'aire de distribution d'*A. m. striatum* n'est pas adjacente à celle d'*A. m. pseudomajus* mais au contraire incluse dans cette dernière (Figure 9). Les deux sous-espèces sont donc au contact l'une de l'autre aux marges de l'aire de distribution d'*A. m. striatum* (Figure 9), ce qui peut donner naissance à des zones hybrides.

Une telle zone hybride entre ces deux sous-espèces a récemment été décrite dans la vallée de Tosses, dans les Pyrénées espagnoles, en Catalogne (Whibley *et al.*, 2006). La couleur des fleurs ségrège chez les hybrides qui présentent une grande variété de phénotypes floraux dans la zone hybride (Figure 10).

Une étude de génétique portant sur un transect le long de cette zone hybride a révélé l'existence de deux clines abrupts concordants entre le gène *ROSEA* et la coloration florale. En revanche, aucune distribution clinale n'a été détectée pour des locus physiquement liés à *ROSEA* (*PALLIDA* et *DICHOTOMA*). Ce résultat a conduit les auteurs à avancer l'hypothèse de la contre sélection des phénotypes hybrides, et donc des génotypes non parentaux de ROS-1 pour expliquer la stabilité de cette zone hybride depuis le début de son suivi, commencé une dizaine d'années plus tôt.

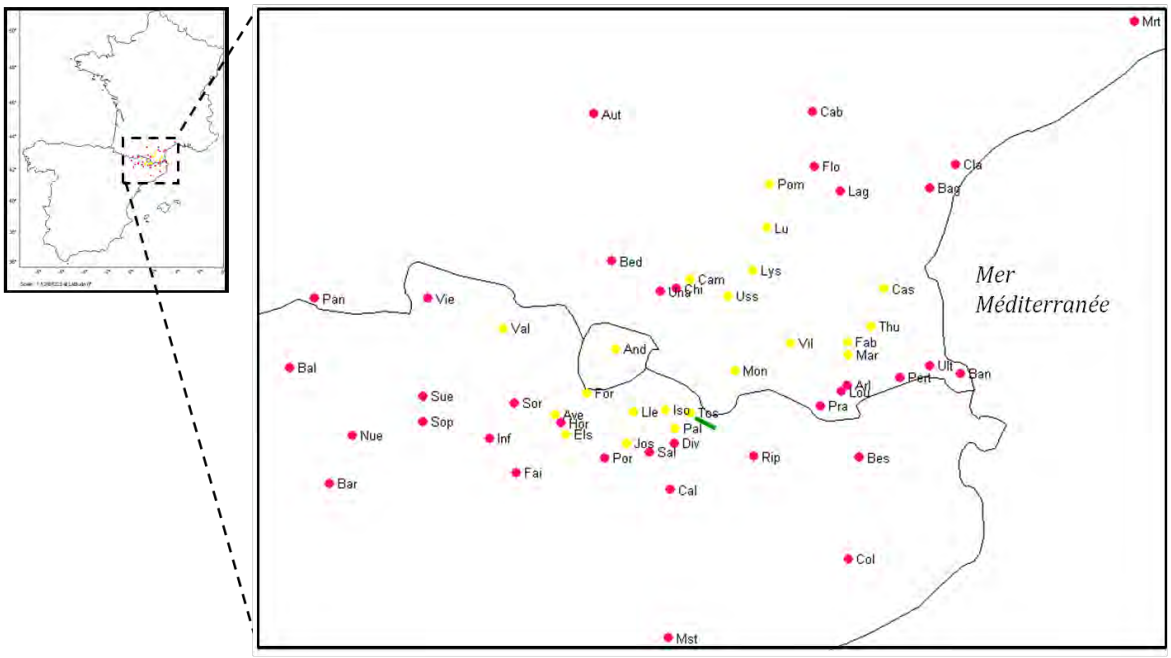


Figure 9 : Diversité des phénotypes floraux hybrides au sein de la zone hybride entre *A. m. pseudomajus* et *A. m. striatum.*

Bien que *A. m. pseudomajus* et *A. m. striatum* partagent leurs pollinisateurs, ceux-ci semblent jouer un rôle à deux niveaux différents dans l'isolement reproducteur de ces deux sous-espèces. D'une part, les pollinisateurs semblent adopter un comportement de constance vis-à-vis des phénotypes majoritaires lorsqu'ils sont confrontés à une grande variété de phénotypes floraux comme c'est le cas dans la zone hybride étudiée (Figure 10).

Figure 10 : Variabilité des phénotypes floraux des hybrides entre *A. m. pseudomajus* et *A. m. striatum*.

Ce comportement se traduit par de l'homogamie chez les plantes et constitue une barrière pré-zygotique à la reproduction des deux sous-espèces en diminuant la fréquence de formation des hybrides. D'autre part, certains phénotypes hybrides souffrent d'une limitation du nombre de visites affectant ainsi leur succès reproducteur. Le choix des phénotypes floraux visités par les pollinisateurs constitue une barrière post-zygotique à leur reproduction. Ces éléments suggèrent que les pollinisateurs jouent un rôle dans la contre-sélection des hybrides, ce qui permettrait le maintien de l'intégrité des phénotypes parentaux aux deux extrémités de la zone hybride, même si l'implication d'autres facteurs n'est pas exclue (Tastard *et al.*, 2008; Tastard *et al.*, 2011).

Ces études nous ont renseignés quant à la nature de l'isolement reproducteur entre les deux sous-espèces et quant au rôle des pollinisateurs via leur comportement de choix. La divergence des deux sous-espèces via la divergence de leurs phénotypes floraux n'ayant été abordée que dans le contexte spatial particulier d'une zone hybride isolée. L'absence d'une perspective évolutive et écologique à grande échelle dans ces études ne permet cependant pas de généraliser ces résultats à l'espèce.

L'objectif de cette thèse est d'appréhender la divergence en cours des deux-sous espèces à une échelle globale selon le cadre conceptuel développé par Dieckman et al (2004). Nous nous sommes donc intéressés dans un premier temps à la phase d'initiation de la divergence entre les deux sous-espèces en attachant un intérêt

particulier pour le contexte spatial et écologique des populations. Dans un deuxième temps, nous nous sommes intéressés aux mécanismes qui permettent le maintien de la divergence phénotypique des deux sous-espèces dans les zones de contact, et plus particulièrement l'impact de l'environnement des populations des deux sous-espèces. La région Pyrénéenne est le siège d'une importante hétérogénéité environnementale susceptible de modifier l'issue des interactions entre les deux sous-espèces et leurs pollinisateurs. Nous avons étudié les processus impliqués dans la phase de renforcement de l'isolement reproducteur et écologique des deux sous-espèces, dans différentes zones de contact. À cette fin, cette thèse s'articule autour de cinq chapitres dont voici un bref aperçu :

CHAPITRE 1

Nous nous sommes intéressés aux facteurs qui ont pu modeler la distribution actuelle d'*A. m. pseudomajus* et *A. m. striatum* et de leur diversité génétique. Dans un premier temps, nous avons testé différents scenario de colonisation post-glaciaire pouvant expliquer la distribution actuelle d'*A. majus*. Nous avons aussi évalué la diversité génétique neutre des deux sous-espèces et étudié sa structuration afin de tester les rôles de la dispersion limité, des barrières physiques ou de l'adaptation à différents habitats dans la structuration génétique des populations des deux sous-espèces.

Article 1: "Test of divergence scenario of two *Antirrhinum majus* subspecies*"*. *En préparation*

Article 2 : "Past colonization and geography define population genetic structure in *Antirrhinum majus". En préparation*

CHAPITRE 2

Dans ce chapitre, nous avons cherché à savoir si l'isolement reproducteur entre les deux sous-espèces est complet ou si elles échangent des gènes dans les zones de contact. L'étude de différents marqueurs génétiques à l'échelle de l'aire de distribution d'*A. majus* a révélé des évènements d'introgression, témoignant de flux de gènes récurrents entre

les deux sous-espèces. Nous avons aussi mis en évidence des gradients d'expansion asymétriques des deux sous-espèces dans des directions opposées à l'ouest et à l'est de la zone de contact.

Article 3: "Locally asymmetric introgressions between subspecies suggest circular range expansion at the *Antirrhinum majus* global scale*". Publié dans Journal of Evolutionary Ecology*

CHAPITRE 3

Nous nous sommes intéressés au rôle des facteurs environnementaux dans l'origine et le maintien de la distribution parapatrique d'*A. m. pseudomajus* et *A. m. striatum*. Nous avons reconstruit les niches environnementales des deux sous-espèces, testé leur divergence et prédit, sur la base des facteurs environnementaux, la coexistence des deux sous-espèces dans les zones de contact. Nous avons montré que si la divergence de niche peut être à l'origine de la distribution paparapatrique des deux sous-espèces, d'autres processus écologiques, tels que la compétition entre les deux sous-espèces, sont probablement impliqués dans son maintien.

Article 4: "Ecology predicts parapatric distributions of two closely related subspecies of *Antirrinum majus". Accepté dans Evolutionary Ecology*

CHAPITRE 4

Nous nous sommes intéressés à la dynamique temporelle d'une zone hybride particulière entre *A. m. pseudomajus* et *A. m. striatum*. Nous avons étudié la distribution des fréquences d'haplotypes chloroplastiques et des fréquences allèliques d'un gène nucléaire lié à la couleur des fleurs. Nous avons mis en évidence un patron discordant entre le cline chloroplastique et le cline nucléaire qui est stable dans le temps. Ces résultats suggèrent l'influence de processus non neutres, tels que la sélection épistatique, dans le maintien de cette zone hybride.

Article 5: Cline discordance and cytonuclear disequilibrium in an *Antirrhinum* hybrid zone. *En preparation*

Chapitre 1

Genetic inference of evolution in closely related *Antirrhinum majus* subspecies

&

Past colonization and geography define population genetic structure in *Antirrhinum majus*

Running title: Testing for models of Antirrhinum evolution

Genetic inference of evolution in closely related *Antirrhinum majus* subspecies

---

En préparation

Khimoun A [1,2], Palero F [3], Burrus M [1,2], Andalo C [1,2], Thebaud C [1,2]

[1] Université de Toulouse; UPS; EDB (Laboratoire Evolution et Diversité Biologique); UMR5174; 118 route de Narbonne, F-31062 Toulouse cedex 9, France.

[2] CNRS; EDB (Laboratoire Evolution et Diversité Biologique); UMR5174; F-31062 Toulouse, France

[3] Unitat Mixta Genòmica i Salut CSISP-UV/Institut Cavanilles Universitat de Valencia, C/ Catedrático Jose Beltran 2, 46980 Paterna, Spain

Author for correspondence: khimoun@cict.fr

Phone number: 00 33 (0)5 61 55 74 86

Fax number: 00 33 (0)5 61 55 73 27

Abstract

The genetic patterns found in contemporary populations can be used to make inferences on recent evolutionary processes. As a paradigmatic example, the geographic distribution and recent evolution of several animal species have been strongly influenced by climatic changes occurring during the Quaternary ice age. In the present study, classic distance-based and the recently developed coalescent-based methods are used to test among different hypotheses on the evolutionary history of *Antirrhinum majus*, a herbaceous short-lived perennial plant characterized by a patchy distribution in southern Europe. In order to infer past demographic processes, a set of 8 populations was genotyped using selectively neutral microsatellite markers. Our results indicate that *Antirrhinum majus* subspecies do not form a monophyletic clade, but rather a well-mixed genetic pool. The pattern observed when using selectively neutral microsatellite markers is indicative of extensive gene flow occurring among yellow-flowered *A. m. striatum* and red-flowered *A. m. pseudomajus* populations, even though migration rates as estimated through ABC methods seemed to have been quite low. Finally, shared polymorphisms could also be related to incomplete lineage sorting due to very recent divergence, even though more markers and more samples are needed to better resolve among the alternative hypotheses.

Keywords: *Antirrhinum majus*, Approximate Bayesian Computation, Pyrennees, allopatry, historical demography.

Geographic distribution and evolution of species have greatly been influenced by climatic changes during the Quaternary ice age (Hewitt, 2011). Climatic oscillations on that time lead to repeated cycles of dry and glaciated periods interrupted by shorter warmer and moister inter-glacial periods (Webb, Bartlein, 1992). During glacial periods, temperate species have gone extinct or contracted in southern regions to track suitable climatic conditions and habitats. Within Europe, the southern Iberian Peninsula, Italy and Balkans are supposed to represent the most important glacial refugia for temperate species (Hewitt, 2000; Hewitt, 1999). Range contraction and population survival in southern refugia might then have promoted speciation through population isolation and accumulation of genetic divergence ("genetic revolutions" (Mayr, 1954), see also (Barton, Charlesworth, 1984)). During subsequent post-glacial warming, many species expanded their ranges and recolonized northern regions from southern refugia (Hewitt, 1996). Although different species have responded differently to the quaternary glaciations (depending on their abilities to adapt or to disperse in new favorable habitats), concordant patterns of genetic variation among European species have allowed identifying the main routes of post-glacial recolonization (mostly through the Pyrenees: see (Hewitt, 2011; Schmitt *et al.*, 2006)).

Two different scenarios can be set up to summarize most reports published to date on post-glacial range expansions and the role of the Pyrenees in shaping genetic patterns (Hewitt, 1999; Taberlet *et al.*, 1998). For some organisms, post-glacial range expansions allowed secondary contact and promoted hybridization between different genetic pools that were formerly isolated (Hewitt, 2001). In such a cases (e.g. the grasshopper *Chorthippus parallelus*), Pyrenean mountains would have acted as a barrier to the northern post-glacial colonization of a gene pool from Iberia (i.e. the *C. p. erythropus* subspecies) and the southern post-glacial colonization of another gene pool from the Balkans (i.e. the *C. p. parallelus*

subspecies) (see Figure 2 A) (Butlin, 1998; Hewitt, 1993). These two subspecies now meet and hybridize along the Pyrenees, in the so called "suture zones", where hybrid zones of different lineages resulting from range expansions cluster together (Hewitt, 1999; Remington, 1968). Nevertheless, the limitation of post-glacial recolonization of Europe driven by the Pyrenean Mountains did not affect all organisms equally. For some species, the Iberian gene pool may have colonized the north of Europe by taking lower-altitude routes at the extremes of the Pyrenean principal chain. For example, the hedgehog *Erinaceus europaeus* re-colonized the north of Europe from Iberia by crossing the principal chain of the Pyrenees through the western and eastern extremes (Figure 2B) (Seddon *et al.*, 2001). Interestingly, despite several studies have focused on the impact of post-glacial recolonization over the phenotypic and genetic diversity of animal species, not much evidence has been accumulated so far on closely related plant species.

Antirrhinum majus (Scrophulariaceae) is a herbaceous short-lived perennial plant characterized by a patchy distribution in southern Europe (Figure 1). Two subspecies have been described within *A. majus* based on flower color (mainly encoded by the nuclear gene *ROSEA* (Whibley *et al.*, 2006)), namely *A. m. striatum* (yellow-flowered) and *A. m. pseudomajus* (magenta-flowered) (Rothmaler, 1956; Sutton, 1988). These two subspecies are interfertile and share pollinators (Andalo *et al.*, 2010). They occupy largely parapatric geographic regions, centred over the Pyrenees, between north-eastern Spain and south-western France (Figure 1). In both subspecies, the geographic range includes both the northern and southern faces of the Pyrenean axial chain, with northern and southern populations being separated by mountain crests (>1800 m which is the highest altitude recorded for an *A. majus* population) that were covered in ice during the last glaciations (see Figure 1). Given their geographic distribution and the fact that they are endemic to this region,

A. majus subspecies can be considered a relevant model to study the impact of Quaternary glaciations on the geographic distribution and evolution of plant species.

According to the most parsimonious scenario (topology 1), we would expect *A. majus* populations to be genetically more similar according to their flower color rather than according to their spatial location. Under such scenario, the history of post-glacial colonization would not shape the current distribution of *A. majus* genetic diversity and we could infer that the yellow-flowered *A. m. striatum* have evolved independently from the red-flowered *A. m. pseudomajus*. In the present study, we aim to test whether the distribution of genetic diversity within *A. majus* is in fact following this flower-color null model or if it has been influenced by *A. majus* history of post-glacial colonization (either by North-South secondary contact or by East-West recolonization). Under the scenario of northern and southern colonizations of *A. majus* range, we expect northern *A. majus* populations to be genetically more similar despite differences in their flower color (topology 2: Figure 2-A). Alternatively, the second scenario considers the northern colonization of *A. majus* through lower altitudes along an eastern/western route (topology 3: Figure 2-B). Under this scenario of eastern colonization of *A. majus* range, we would expect *A. majus* populations to be genetically more similar according to their spatial location on the eastern *versus* western part of the Pyrenees rather than according to their flower color (Figure 2-B).

To test the flower-color null model against the alternative recolonization models, we rely on a set of microsatellite markers because they are thought to be selectively neutral and would thus reflect the past demographic processes. Multilocus genotypes from several *A. majus* populations were analyzed using both classic distance-based methods and the recent Approximate Bayesian Computation (ABC) framework to test the likelihood of alternative population topologies expected under the different scenarios of colonization. ABC methods

have been shown to be particularly useful for the inference of demographic parameters from genetic data since these methods do not need to evaluate likelihood functions analytically and can therefore be used even while assuming complex models.

MATERIALS AND METHODS

Microsatellite amplification

Leaf samples of *Antirrhinum* individuals (n = 160) were obtained from several localities covering most of the present distribution of the species and located in the northern, southern, eastern and western parts of the Pyrenees (Figure 1). The sampled area included 8 different localities, with sample sizes ranging from 12 to 41 individuals (Table 1). Total genomic DNA extraction was performed using the DNeasy Plant Mini kit (Qiagen). Microsatellite loci were amplified in four different multiplex reactions (Appendix 1 in the online supplementary material). Multiplex amplification reactions were carried out in a 10 µl reaction volume containing 2 µl genomic DNA, 3.5 µl MasterMix (Qiagen), and 4.5 µl of the primer mix (Appendix 1 in the online supplementary material). The PCR thermal profile used was 94°C for 10 min for initial denaturation, followed by 24 cycles of 94°C for 30 s, the primer annealing temperature (which differs depending on the different multiplexes: see Appendix 1 in the online supplementary material) for 1 min 30 s s and 72°C for 1 min, and a final 30 min extension at 60°C in a Mastercycler Gradient (Eppendorf). After preliminary analysis of chromatograms, a subset of 10 polymorphic microsatellite loci was kept for this study (Table 2). Amplified products were scored using an ABI 3730 automatic sequencer from the Plateforme Génomique de Toulouse (France). Alleles were sized by PeakScanner™ and

Microsatelight (Palero *et al.*, 2011) software, with an internal size marker Rox™ Size Standard 500 (Applied Biosystems).

Classic distance-based methods

We employed CONVERT 1.2 to transform the excel-based microsatellites dataset into different formats to be run by other population genetic programs (Glaubitz, 2004). Genetic diversity and pairwise differentiation estimates (F_{ST}) for microsatellite data were obtained using the GENEPOP package version 4.0.7 (Rousset, 2008). Isolation by distance patterns were evaluated by correlating geographic distance and genetic distance. If *Antirrhinum* speciation had occurred following a gradual South-to-North pattern, we would expect species most closely located geographically to be those separated by a smaller genetic distance, whereas we would expect a vicariance event and posterior contact between genetically divergent populations to reduce the correlation between genetic and geographic distance matrices. Finally, phylogenetic trees (based on individuals or species) were built using the D_A distance (Nei, Chesser, 1983) defined as:

$$D_A = 1 - \frac{1}{r}\sum_{j}^{r}\sum_{i}^{m_j}\sqrt{x_{ij}y_{ij}}$$

where xij and yij are the frequencies of the ith allele at the jth locus in populations X and Y, respectively, and mj is the number of alleles at the jth locus. This distance-based method was selected since it has been shown to be among the most robust methods for phylogenetic reconstruction using microsatellite data (Felsenstein, 2005; Wiens, 2000; Wiens, Servedio, 1998). A recent study also showed that the probability of obtaining the correct branching pattern of a tree is generally highest for D_A distance (Takezaki, Nei, 2008). The distance-

based Neighbor Joining algorithm was used as implemented in Populations v1.2.30 http://bioinformatics.org/~tryphon/populations/. A total of 1000 bootstrap replicates over loci were obtained to asses support for each clade. Given that *Antirrhinum molle* is closely related to *Antirrhinum majus*, it has been used to root the trees (Vargas *et al.*, 2009).

ABC implementation: Models to be tested using ABC methods

When inferring phylogenies under the "Isolation with migration" model, likelihoods can only be computed for relatively simple scenarios containing few parameters (Hey, Nielsen, 2007; Wakeley, 1996). Indeed, the likelihood function for complex demographic scenarios can be very difficult, and practically impossible, to solve analytically (Marjoram *et al.*, 2003). This is the main reason why the application of ABC methods to solve phylogenetic inference-related problems has become of great interest (Csillery *et al.*, 2010; Hickerson *et al.*, 2006). ABC methods have the advantage of facilitating the comparison of alternative models marginal to parameter values without the need for calculating likelihoods (Beaumont *et al.*, 2002). The method relies on the simulation of large numbers of data sets using known parameters under a given coalescent model, for which it is more realistic than standard sequence-based phylogenetic approaches (Estoup *et al.*, 2004). When dealing with coalescent-based inference, we rely on simulating genetic data based on a coalescence model and computing summary statistics from simulated datasets. A typical ABC approach involves two steps (Beaumont *et al.*, 2002): a rejection step and a regression adjustment and weighting step. The rejection step consists of accepting only the simulations whose summary statistics are close to the summary statistics obtained from the observed dataset. To assess this closeness, a Euclidian distance is computed between the entire set of normalized summary statistics and the normalized summary statistics calculated from the data. A set of parameter values is accepted when its

Euclidian distance is within a certain percentage of the closest points to the studied data (Beaumont, 2008). The second step is a local linear regression adjustment that attempts to model the relationship between the parameter values and the summary statistics. This linear regression is performed only for the accepted set of parameter values. We assume that the relation between parameters and summary statistics is close to linear in the proximity of the target summary statistics. By using this adjustment, more points can be accepted, which allows a better characterization of the space problem (Estoup *et al.*, 2004). Also in this step, each accepted set of parameter values is given a weight between zero and one that declines quadratically until a defined distance from the studied data set is reached (Legras *et al.*, 2007). To reduce heteroscedasticity in the regression, all demographic parameter values were transformed on a log scale. The transformed parameter values were adjusted one at a time using a general linear regression on the accepted points. Adjusted values were then back-transformed taking the exponential for all parameters, in order to present posterior densities on a normal scale (Beaumont *et al.*, 2002; Estoup *et al.*, 2004). The transformation also minimizes the appearance of values outside the prior ranges after performing the linear-regression correction. Previous studies indicated that the logistic and related transformations can lead to biases in the posterior densities estimated in the proximity of the prior boundaries under particular circumstances (Lopes *et al.*, 2009). To avoid this problem we choose a log transformation, which still allows for points at the lower boundary to be retained within the support of the model. In this case, the points that fell outside the upper boundary after regression were discarded, since this procedure has been shown to give a more efficient estimation (Lopes *et al.*, 2009). A standard backward coalescent process was implemented to simulate genetic data (Hudson, 1990; Nordborg, 2003). Simulated data are obtained by adding mutations under a stepwise mutation model for short tandem repeats (STRs) (Kimura, Ohta,

1978). Hamilton *et al.* (2005) suggest running several hundreds of thousands to millions of simulations, depending on the complexity of the underlying model. In our simulations 10,000,000 values of the summary statistics sets were generated and a tolerance $\delta = 0.01$ was used to give 100,000 points from which parameters were estimated. When performing model-choice between the suggested different scenarios 20,000,000 points were simulated and a tolerance of $\delta = 0.005$ was used. We used the mode of the posterior distributions as a point estimate of the parameter. Credible intervals were calculated around the mode, following previous studies (Beaumont, 2008; Hamilton *et al.*, 2005). The model-choice studies were performed by first carrying out the simulation in parallel on a 256-node cluster, and then combining the simulated output, in order to shorten the simulation time. A program developed by Lopes and co-workers was used to simulate genetic data in an "Isolation with migration" model for any number of modern populations (Lopes *et al.*, 2009). This software allows the use of STR's and single nucleotide polymorphism (SNP) data simultaneously. The regression step was performed using a script developed by Beaumont (makepd.r, http:// www.rubic.rdg.ac.uk/~mab/) under the free software environment R v2.14.0. The posterior density estimation from the adjusted sample of parameter values was carried out using the locfit function (Loader, 1996).

Prior distributions of parameters

The same priors for the demographic parameters (current and ancestral effective sizes following a uniform distribution ranging from 10 to 20,000) were used for inferences based on every model topology (see above for details). The priors for the demographic parameters were chosen according to information available from the literature. The priors for the splitting time estimates also followed a uniform prior, as for the effective sizes above. Mutation rates

for each locus were treated as a nuisance parameter, given that it was not intended to infer their exact values. Therefore, a broad prior was used for the loci mutation rates to account for the uncertainty on the estimates. The variation in mutation rate between loci was accounted for by using a hierarchical Bayesian framework (Storz, Beaumont, 2002). The mutation rates for each locus were drawn from a lognormal distribution (priors) with mean sampled from a normal distribution and the standard deviation being the absolute value sampled also from a normal distribution (hyper-priors) (Beaumont, 2008). In order to cover the proposed limits of the mutation rate, we used a standard deviation hyperprior of 0.25. The use of hyper-parameters within ABC methods has been previously described (Beaumont, 2008; Excoffier *et al.*, 2005).

Choice of summary statistics

The summary statistics were chosen according to their success in previous ABC studies (Beaumont, 2008; Fagundes *et al.*, 2007). For STR data, six summary statistics were calculated for each sampled deme: allele number, k; heterozygosity, H; curtosis of alleles length; Shanon's index; Nm estimator based on H and variance in allele length, Var(length). All these 6 statistics were computed for each population taken individually and for each pair of populations pooled together. Hence, the Euclidian distances were computed from a total of 80 normalized summary statistics.

Comparison of scenarios using Approximate Bayesian Computation

In order to test between the flower-color groups (topology 1) and the previously proposed post-glacial recolonization hypotheses (Figure 2), we considered several scenarios which differed in the population tree topology. Under the null model, populations are grouped

according to flower color, so that yellow-flowered plants from the southern and northern faces of the Pyrenees form a monophyletic group. Under topology 2 however, populations are grouped according to North-South geographical location, so that plants from the southern face of the Pyrenees form a monophyletic group independently of flower color. For testing the North-South secondary contact (topology 2) against the flower colour hypothesis (topology 1), several populations from both northern and southern faces were included for both the yellow-flowered *A. m. striatum* (Cam: n = 16 and Ave: n = 12) and the red-flowered *A. m. pseudomajus* (Una: n = 19 and Rip: n = 12). For testing the East/West re-colonization (topology 3) against the flower colour hypothesis (topology 1), populations from both western and eastern regions were also included for the yellow-flowered *A. m. striatum* (Thu: n = 15 and Ave: n = 12) and the red-flowered *A. m. pseudomajus* (Hor: n = 41 and Per: n = 33). A single *Antirrhinum molle* population Fai (n = 12) was used as an outgroup in all simulations carried out.

Then coalescent-based simulations were performed under an ABC framework in order to discriminate among these three different scenarios (Topology 1 – Flower-color; Topology 2 – North-South secondary contact; Topology 3 – East/West re-colonization). A model-selection step was performed before estimating the final demographic historic parameters, which were done conditional to the most likely scenario. The prior probability for each scenario in all the comparisons were set to be equal (*i.e.* 1/2 for each two-scenario comparison). The posterior probability of each model was then estimated by performing the rejection-step followed by a logistic regression (Beaumont, 2008). Priors for divergence times were made broad enough to consider alternative speciation patterns. Beaumont (2008) also indicated that it is possible to sample the model indicator (i.e. {1, 2,..., m}) for "m" models (M1, M2,..., Mm) from a prior and treat this as a categorical random variable, X, in the ABC simulations. We can then apply

a categorical regression to estimate P(X = x1|S = s'), where x = 1, 2,..., m is the indicator for model Mx and s' is the vector of the summary statistics that summarize our observed data. A scheme of weighting can also be used, with weights given by the Epanechnikov kernel, as done in a standard regression procedure. The regression-step was performed using Beaumont's R script calmod http:// www.rubic.rdg.ac.uk/~mab, which needs the VGAM package (Yee, Wild, 1996). This procedure has been shown to substantially improve previous methods to select among different models using ABC (Beaumont, 2008; Palero *et al.*, 2009).

RESULTS

Classic distance-based methods

The linear-regression analysis revealed a significant correlation between the genetic distance matrix and the matrix of pairwise geographic distances when the set of *A. majus* populations used for the North-South model was analyzed (R = 0.661; P = 0.04). Interestingly, no significant correlation between the genetic distance matrix and the matrix of pairwise geographic distances was found when analyzing the set of *A. majus* populations used for the East/West model (R = 2.5*10^{-} 8; P = 0.99). The distance measure of Nei et al. (1983) placed *A. m. pseudomajus* samples next to *A. m. striatum* samples when phylogenetic trees were built using the individual-based matrices and the Neighbor Joining algorithm (Figure 3). When dealing with populations instead of individuals, a reciprocal monophyly of the northern and southern populations was obtained as well, even though phylogenetic relationships among populations were not completely resolved (Figure 3).

ABC methods

We made use of the existing simulations in order to evaluate whether model selection with ABC is able to distinguish between the proposed models. The summary statistics from one of the simulations are considered as pseudo-observed summary statistics and classified using all the remaining simulations. Then, if the summary statistics contain sufficient information to discriminate among models, one expects that a large posterior probability should be assigned to the model that generated the pseudo-observed summary statistics. Nevertheless, the comparisons between the null model based on flower colour and the two alternative post-glacial recolonization models (Figure 2) did not favour any particular speciation model, providing us with a misclassification proportion above 70%. Therefore, the ABC runs are indicating us that there is not enough information in the molecular markers used to strongly favour any particular model.

In any case, it is worth pointing out that estimates of the modern population effective sizes using microsatellite markers were very similar among different models, with ranges going from 1,000 to 20,000 for all *A. majus* populations studied (results for the North-South model given in Table 3). Estimates of the effective size of the ancestor populations were not very informative either given the large confidence intervals obtained (supplementary material). Of course, it should be pointed out that these values correspond to the effective population size, so it is not straightforward to infer the census population sizes from them. Surprisingly, our results would seem to point out towards a contraction of current populations compared with the referred ancestral populations. When conditioning for a North-to-South speciation pattern, all the five splitting times showed a posterior distribution relatively broad around the mode (Table 3), but in all cases the splitting time estimates can be related to past glacial events.

DISCUSSION

Understanding the forces that influence natural variation within and among populations has been a major objective of evolutionary biologists for decades. One of the most complete studies of the *Antirrhinum* (Scrophulariaceae) speciation carried out to date showed that only a limited number of monophyletic groups of *Antirrhinum* accessions can be actually related to sections and species (Vargas *et al.*, 2009). The authors attributed the observed patterns to high levels of homoplasy (correspondence of characters acquired as the result of parallel, reversal or convergent evolution), incomplete lineage sorting (persistence of ancestral polymorphism through speciation events) or reticulation (non-hierarchical gene transfer). Indeed, based on the distribution of key morphological characters, (Rothmaler, 1956) had already envisioned isolation–contact–isolation processes during the dry and wet episodes of the Ice Ages, resulting in hybridization and subsequent character sharing. The results obtained in the present study using a set of neutrally-evolving microsatellite markers seem to further support a scenario of extensive hybridization among *Antirrhinum* species.

Lack of data resolution suggest a recent evolutionary history of *A. majus*

Motivated by the growth in computational power and data complexity, modern approaches to phylogepographic and speciation questions make intensive use of simulation methods. Approximate Bayesian Computation (ABC) methods have been shown to be particularly useful for the inference of demographic parameters from genetic data. These methods do not need to evaluate likelihood functions analytically and therefore are supposed to be useful even while assuming complex models. Nevertheless, despite the long ABC simulations run in the present study, we were not able to confidently discriminate between different scenarios of colonization or to infer accurately migration rates or effective population sizes. This limitation

may be related to the amount of information contained within our microsatellite markers or it might be due to the retention of ancestral polymorphisms resulting from a recent divergence.

The chronology of the Maximum Ice Extent (MIE, that is the farthest advance of ice out of a mountain crest zone just before the start of the melt season) in the eastern part of the Pyrenees does not coincide with the existing chronologies for other parts of the Pyrenees. The MIE occurred much later (between 26,500 and 20,000 years ago) than farther west (between 74,000 and 60,000 years ago) (Andrieu *et al.*, 1988; Jalut *et al.*, 1988). Explanations for this may lie with climatic conditions specific to the eastern Pyrenees, namely the drier Mediterranean climate (Calvet, 2004). Because of its milder and later glaciations, taking the South-to-North route of colonization through the East would have been easier to colonize the current range of *A. majus* than following the western route.

In this study, we aimed at answering the main questions on the origin and significance of the *Antirrhinum* flower color and relate it to past events. The *Antirrhinum* flower color is generally used as a key character to distinguish among *Antirrhinum majus* subspecies but, to which extent is this character really telling us about the monophyletic origin of the yellow versus red flower plants? As with many examples in biology, some characters previously thought to reliably indicate the grouping of samples as Evolutionary Significant Units may in fact be shaped by ecological and environmental factors, rather than past history. In fact, optimal ecological conditions may also have been important in *Antirrhinum* evolution (Whibley *et al.*, 2006).

Two neutral genetic pools within each subspecies (i.e. flower color)

From the different scenarios we tested and given our genetic data, the most likely topologies using distance-based methods never grouped populations on the basis of their flower colors.

A. m. striatum and *A. m. pseudomajus* populations from the same part of the Pyrenees were genetically more similar than populations of the same subspecies located in different parts of the Pyrenees. In other words, *A. majus* populations were grouped according to their east/west and north/south locations. Our results showed that the neutral genetic diversity of *A. majus* populations is not mainly structured according their flower color but rather according to their geographic location. Within *A. majus*, our results highlighted the presence of two neutral genetic pools spatially structured in different parts of the Pyrenees. Two hypotheses could explain such a spatial structuration of *A. majus* neutral genetic diversity. The first hypothesis would be that the two different genetic pools differentiated through vicariance. Indeed, the advance of glaciers during the cold quaternary periods might have spatially isolated populations that evolved and accumulated genetic differences under reproductive isolation. Another hypothesis to explain two different gene pools spatially structured is that *A. majus* range would result from two colonization events with different origins. The two neutral genetic pools might have differentiated in two different refugia and then independently colonized *A. majus* range during the post-glacial periods. As shown by the ABC simulation results, our set of microsatellite markers could not discriminate among these hypotheses.

Two kind of nuclear markers revealed two different structuration patterns

Although the data were not enough informative to discriminate the most likely route of colonization, our results indicate that *A. majus* neutral genetic diversity is spatially structured in two clusters that do not correspond to the flower color. The spatial distribution of flower color, which is encoded by the nuclear gene *ROSEA*, is different from the spatial structure

observed from microsatellite markers. In this study, we revealed discordant patterns between the *ROSEA* gene and neutral genetic markers. Neutral markers could be reflecting the historic and demographic events whereas genes encoding for phenotypic traits (like *ROSEA*) are more likely to be reflecting the ecology or any other factor why they can be selected for.

Previous works have suggested the potential adaptive role of flower color (Whibley *et al.*, 2006) via pollinator behavior (Tastard *et al.*, 2008).Because flower color and thus ROSEA alleles highlight a different picture from the one provided by microsatellites (Khimoun *et al.*, 2011), our results would support the hypothesis of an adaptive role of flower color. In fact, the two *A. majus* subspecies have been shown to be correlated with different environments. *A. m. striatum* populations are generally found in locations characterised by higher precipitation, lower temperatures, higher thermal amplitudes and more compacted and wetter, poor in nutrients soils than *A. m. pseudomajus* (Khimoun *et al.*, Accepted). Yellow-flowered plants would thus be more adapted to "mountain-like" environmental condition than magenta-flowered plants.

How to reconcile the picture from microsatellites and from flower color?

Although apparently discordant, the genetic structure patterns provided from the microsatellites and flower color can be explained by a divergence scenario of the two subspecies. The neutral genetic structure does not seem to match with the flower color distribution but rather matches with the spatial location of the populations. Two alternative hypotheses could explain this observation. Under the first hypothesis, flower color divergence occurred one time and flower colors, which are probably adaptive in differential environments, were thus fixed within populations. Because microsatellites are selectively neutral, so alleles are free to pass between the two subspecies, the absence of neutral genetic

structure compared to flower color would be the result of extensive gene flow during the two subspecies divergence or after secondary contact. The structure of neutral diversity depending on spatial location of populations would therefore reflect the better connectivity of populations when they are located in the same side of the Pyrenees. Another hypothesis would be that flower color polymorphism could have been present in the ancestral populations. Flower-color alleles would have become fixed in different locations only after populations were isolated in different regions of the Pyrenees, potentially as a result of local environmental selection.

CONCLUSION

The combination of several nuclear markers (microsatellites) under a classic distance-based method and an ABC-coalescent framework has proven to be effective for testing among alternative evolutionary hypotheses in *Antirrhinum* and highlights the importance of using multiple markers when dealing with closely-related species. The *Antirrhinum* speciation pattern is a typical example of a series of rapid speciation events occurring within a group, with different populations diverging in a very short period of time. These recent speciation events provide a great opportunity to analyze the speciation process in several taxa, since footprints of species formation are most likely to be identified when comparing recently diverged species, the initial differentiation of which can be correlated with the different proposed speciation processes.

Bibliography

Andalo C, Cruzan MB, Cazettes C, *et al.* (2010) Post-pollination barriers do not explain the persistence of two distinct *Antirrhinum* subspecies with parapatric distribution. *Plant Systematics and Evolution* **286**, 223-234.

Andrieu V, Hubschman J, Jalut G, Herail G (1988) Chronologie de la déglaciation des Pyrénées françaises. Dynamique de sédimentation et contenu pollinique des paléolacs: application à l'interprétation du retrait glaciaire. *Bulletin de l'Association Française pour l'Etude du Quaternaire* **2**, 55-67.

Barton NH, Charlesworth B (1984) Genetic revolutions, founder effects, and speciation. *Annual Review of Ecology and Systematics* **15**, 133-164.

Beaumont MA (2008) Joint determination of topology, divergence time, and immigration in population trees. In: *Simulations, Genetics, and Human Prehistory* (eds. Matsumura S, Forster P, Renfrew C), pp. 135-154. McDonald Institute for Archaeological Research, Cambridge.

Beaumont MA, Zhang WY, Balding DJ (2002) Approximate Bayesian Computation in population genetics. *Genetics* **162**, 2025-2035.

Butlin RK (1998) What do hybrid zones in general, and the *Chorthippus parallelus* zone in particular, tell us about speciation? In: *Endless forms: species, speciation* (eds. Howard DJ, Berlocher SH), pp. 367–378. Oxford University Press, New York

Calvet M (2004) The Quaternary glaciation of the Pyrenees. In: *Quaternary glaciations - Extent and chronology* (eds. Ehlers J, Gibbard PL), pp. 119-128. Elsevier, Amsterdam.

Csillery K, Blum MGB, Gaggiotti OE, Francois O (2010) Approximate Bayesian Computation (ABC) in practice. *Trends in Ecology & Evolution* **25**, 410-418.

Estoup A, Beaumont M, Sennedot F, Moritz C, Cornuet JM (2004) Genetic analysis of complex demographic scenarios: Spatially expanding populations of the cane toad, *Bufo marinus*. *Evolution* **58**, 2021-2036.

Excoffier L, Estoup A, Cornuet JM (2005) Bayesian analysis of an admixture model with mutations and arbitrarily linked markers. *Genetics* **169**, 1727-1738.

Fagundes NJR, Ray N, Beaumont M, *et al.* (2007) Statistical evaluation of alternative models of human evolution. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 17614-17619.

Felsenstein J (2005) PHYLIP (Phylogeny Inference Package) version 3.6. Department of Genome Sciences, University of Washington, Seattle.

Glaubitz JC (2004) CONVERT: A user-friendly program to reformat diploid genotypic data for commonly used population genetic software packages. *Molecular Ecology Notes* **4**, 309-310.

Hamilton G, Currat M, Ray N, *et al.* (2005) Bayesian estimation of recent migration rates after a spatial expansion. *Genetics* **170**, 409-417.

Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature* **405**, 907-913.

Hewitt GM (1993) After the ice - *parallelus* meets *erythropus* in the Pyrenees. In: *Hybrid zones and the evolutionary process* (ed. Harrison RG), pp. 140–164. Oxford University Press, Oxford.

Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society* **58**, 247-276.

Hewitt GM (1999) Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society* **68**, 87-112.

Hewitt GM (2001) Speciation, hybrid zones and phylogeography - or seeing genes in space and time. *Molecular Ecology* **10**, 537-549.

Hewitt GM (2011) Quaternary phylogeography: the roots of hybrid zones. *Genetica* **139**, 617-638.

Hey J, Nielsen R (2007) Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 2785-2790.

Hickerson MJ, Stahl EA, Lessios HA (2006) Test for simultaneous divergence using Approximate Bayesian Computation. *Evolution* **60**, 2435-2453.

Hudson RR (1990) Gene genealogies and the coalescent process. *Oxford Survey of Evolutionary Biology* **7**.

Jalut G, Andrieu V, Delibrias G, Fontugne M, Pages P (1988) Paleoenvironment of the valley of Ossau (French Western Pyrénées) during the last 27 000 years. *Pollen et Spores* **3**, 357-393.

Khimoun A, Burrus M, Andalo C, *et al.* (2011) Locally asymmetric introgressions between subspecies suggest circular range expansion at the *Antirrhinum majus* global scale. *Journal of Evolutionary Biology* **24**, 1433-1441.

Khimoun A, Cornuault J, Burrus M, *et al.* (Accepted) Ecology predicts parapatric distributions in two closely related *Antirrhinum majus* subspecies. *Evolutionary Ecology*.

Kimura M, Ohta T (1978) Stepwise mutation model and distribution of allelic frequencies in a finite population. *Proceedings of the National Academy of Sciences of the United States of America* **75**, 2868-2872.

Legras JL, Merdinoglu D, Cornuet JM, Karst F (2007) Bread, beer and wine: *Saccharomyces cerevisiae* diversity reflects human history. *Molecular Ecology* **16**, 2091-2102.

Loader CR (1996) Local Likelihood Density Estimation. *Annals of Statistics* **24**, 1602-1618.

Lopes JS, Balding D, Beaumont MA (2009) PopABC: a program to infer historical demographic parameters. *Bioinformatics* **25**, 2747-2749.

Marjoram P, Molitor J, Plagnol V, Tavare S (2003) Markov chain Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 15324-15328.

Mayr E (1954) Change of genetic environment and evolution. In: *Evolution as a Process* (eds. Huxley J, Hardy AC, Ford EB), pp. 157–180 Allen & Unwin, London.

Nei M, Chesser RK (1983) Estimation of fixation indexes and gene diversities. *Annals of Human Genetics* **47**, 253-259.

Nordborg M (2003) Coalescent theory. In: *Handbook of Statistical Genetics* (eds. Balding DJ, Bishop M, Cannings C), pp. 602-635. Wiley, Chichester.

Palero F, Gonzalez-Candelas F, Pascual M (2011) MICROSATELIGHT-Pipeline to Expedite Microsatellite Analysis. *Journal of Heredity* **102**, 247-249.

Palero F, Lopes J, Abello P, *et al.* (2009) Rapid radiation in spiny lobsters (*Palinurus spp*) as revealed by classic and ABC methods using mtDNA and microsatellite data. *Bmc Evolutionary Biology* **9**.

Remington CL (1968) Suture-zones of hybrid interaction between recently joined biotas. In: *Evolutionary biology* (eds. Dobzhansky T, Hecht MK, Steere WC), pp. 321–428. Plenum, New Yorl.

Rothmaler W (1956) Taxonomische monographie der Gattung *Antirrhinum* (ed. Akademie-Verlag), Berlin.

Rousset F (2008) GENEPOP ' 007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources* **8**, 103-106.

Schmitt T, Hewitt GM, Muller P (2006) Disjunct distributions during glacial and interglacial periods in mountain butterflies: *Erebia epiphron* as an example. *Journal of Evolutionary Biology* **19**, 108-113.

Seddon JM, Santucci F, Reeve NJ, Hewitt GM (2001) DNA footprints of European hedgehogs, *Erinaceus europaeus* and *E. concolor*. Pleistocene refugia, postglacial expansion and colonization routes. *Molecular Ecology* **10**, 2187-2198.

Storz JF, Beaumont MA (2002) Testing for genetic evidence of population expansion and contraction: An empirical analysis of microsatellite DNA variation using a hierarchical Bayesian model. *Evolution* **56**, 154-166.

Sutton DA (1988) *A revision of the tribe Antirrhineae*, London and New York.

Taberlet P, Fumagalli L, Wust-Saucy A-G, Cossons J-F (1998) Comparative phylogeography and post-glacial colonization routes in Europe. *Molecular Ecology* **7**, 453-464.

Takezaki N, Nei M (2008) Empirical tests of the reliability of phylogenetic trees constructed with microsatellite DNA. *Genetics* **178**, 385-392.

Tastard E, Andalo C, Giurfa M, Burrus M, Thebaud C (2008) Flower colour variation across a hybrid zone in *Antirrhinum* as perceived by bumblebee pollinators. *Arthropod-Plant Interactions* **2**, 237-246.

Vargas P, Carrio E, Guzman B, Amat E, Guemes J (2009) A geographical pattern of *Antirrhinum* (Scrophulariaceae) speciation since the Pliocene based on plastid and nuclear DNA polymorphisms. *Journal of Biogeography* **36**, 1297-1312.

Wakeley J (1996) The variance of pairwise nucleotide differences in two populations with migration. *Theoretical Population Biology* **49**, 39-57.

Webb T, Bartlein PJ (1992) Global change during the last 3 million years - Climatic controls and biotic responses. *Annual Review of Ecology and Systematics* **23**, 141-173.

Whibley AC, Langlade NB, Andalo C, *et al.* (2006) Evolutionary paths underlying flower color variation in *Antirrhinum*. *Science* **313**, 963-966.

Wiens JJ (2000) Reconstructing phylogenies from allozyme data: comparing method performance with congruence. *Biological Journal of the Linnean Society* **70**, 613-632.

Wiens JJ, Servedio MR (1998) Phylogenetic analysis and intraspecific variation: Performance of parsimony, likelihood, and distance methods. *Systematic Biology* **47**, 228-253.

Yee TW, Wild CJ (1996) Vector generalized additive models. *Journal of the Royal Statistical Society Series B-Methodological* **58**, 481-493.

List of Figures

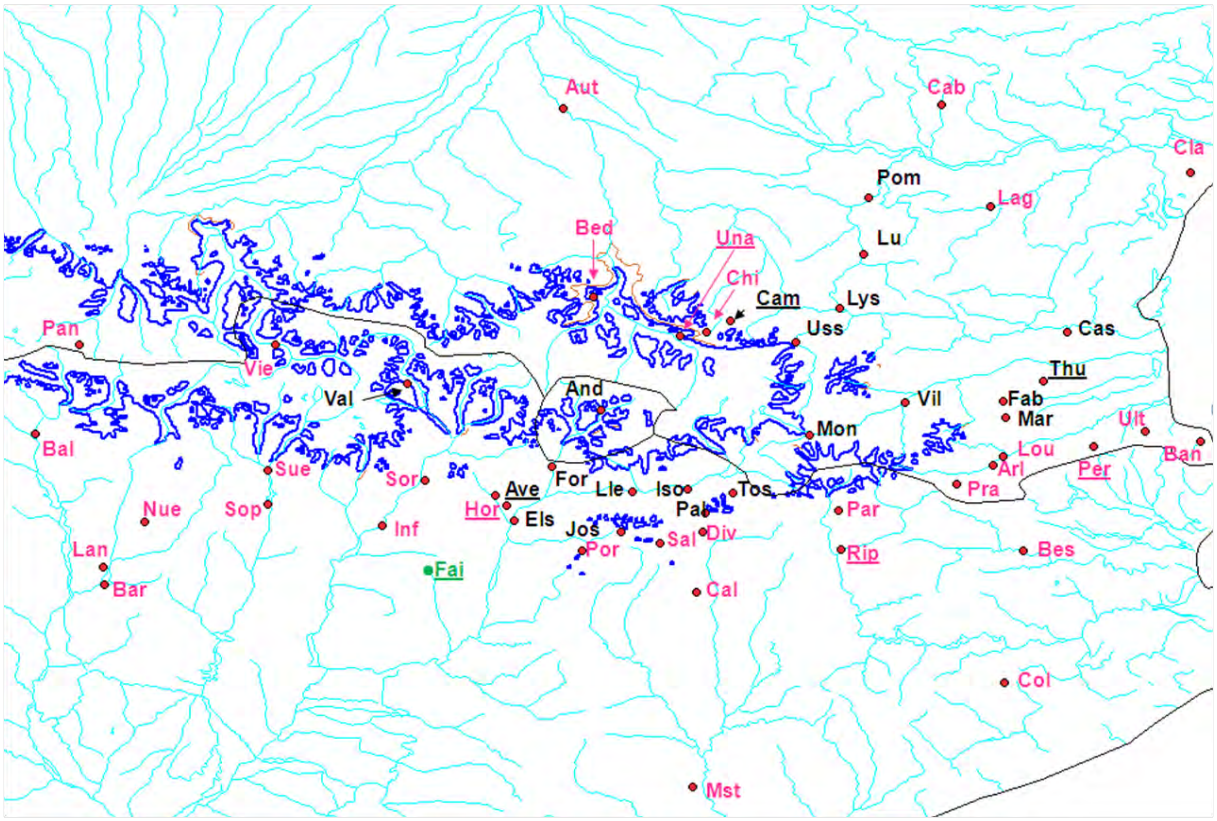


Figure 1: *Antirrhinum majus* range

The dark blue line represents the Maximum Ice Extend during the last glaciations (From Calvet 2004).

A. m. pseudomajus and *A. m. striatum* populations are represented with pink and black labels respectively. *A. molle* population used as an out group in the analysis is labeled in green. The 8 populations used in the analysis are underlined.

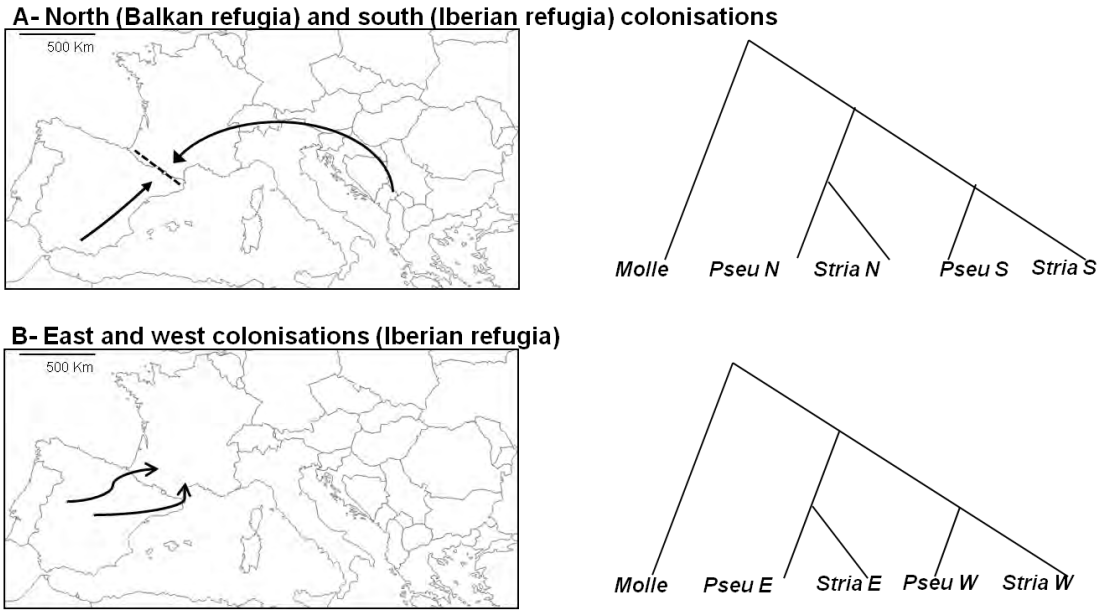


Figure 2: Tree topologies of *A. m. striatum* and *A. m. pseudomajus* populations expected for 2 alternative scenarios of post-glacial colonization.

"Pseu" is used for *A. m. pseudomajus* populations, "Stria" is used for *A. m. striatum* populations and "Molle" is used for A. molle populations. "N" refers to populations located in the northern side of the Pyrenees whereas "S" is used for populations located on the southern side of the Pyrenees. "W" is used for populations located in the western part of the Pyrenees whereas "E" is used for populations located in the eastern part of the Pyrenees. For the different scenarios, arrows indicate the colonization routes and the dashed line represent the location of the suture zone.

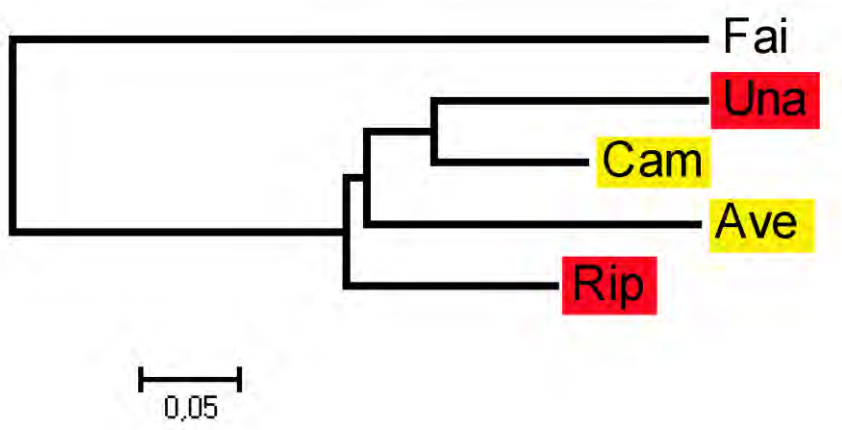


Figure 3: Tree topologies of *A. m. striatum* and *A. m. pseudomajus* populations obtained using distance-based phylogenetic methods. In all cases individuals/populations are mixed independently of flower-color.

List of tables

Table 1: Sampled localities

Population	n	Subspecies	North/South	East/West	Longitude	Latitude	Altitude
Fai	12	molle	South	West	1.16	42.167	1109
Hor	41	pseudomajus	South	East	1.348	42.325	1128
Per	33	pseudomajus	South	West	2.849	42.476	148
Rip	12	pseudomajus	South	West	2.202	42.215	709
Una	19	pseudomajus	North	East	1.79	42.76	915
Ave	12	striatum	South	East	1.32	42.353	1241
Cam	16	striatum	North	East	1.92	42.8	1241
Thu	15	striatum	South	West	2.721	42.644	130

Table 2: Microsatellite loci used in this analysis

Locus	Primer sequence	Repeats	Source
MAT5	TAAGAGGATAGAGAACCAAAC	$(AT)_5$	Feng (unpublished)
	AACACATACACAAACACACAAG		
An1	CCACACCAAAGTTTCCGACAG	$(CT)_{15}$	Zwettler et al. 2002
	CAACAAAAACCATAATCCTAG		
Ant11	GCATCAGCGTAATTTAATG	$(CA)_{10}$	Zwettler et al. 2002
	AAGAAGATGCCTTTGTGAG		
MSAT18	GGATTCTCTCCGATTGCTGT	$(CTT)_{10}$	Erasmus (unpublished)
	GCAGGTGATGTTGCCATTAG		
MSAT31	GACAAATCATCCGTAGAAGC	$(AT)_{14}$	Erasmus (unpublished)
	AAGAACTCCAAACATTCAAA		
FSAT26	TGGTGGCACCGCGACGGTGAAC	$(ATA)_{14}$	Palero (unpublished)
	CCGCACGGCTTTGCCGGAGA		
MSAT5	CTCAACCGCCACCAAAAC	$(TAA)_{13}$	Erasmus (unpublished)
	CGGGAAGGGTAAAACCGTC		
FSAT61	TCGCGACACAATCGGTTTGGT	$(TGA)_{11}$	Palero (unpublished)
	TCCGAGAGATGCAACAAGCCA		
FSAT34	AAAAATCCCTGCAACTGTCAC	$(TAA)_{13}$	Palero (unpublished)
	TTTGACGAATTTACCCCTGGA		
FSAT35	GGAGAGAAATCCCGACCTTTG	$(TAG)_{10}$	Palero (unpublished)
	TAATTTCTCGGCTGAAAGCGA		

Table 3: Estimates for the median and the 95% credible intervals for each demographic historic parameter when the simulation is conditional to the most North-to-South speciation scenario.

	Median	95% CI	
Mutation rate	-3.85886	-4.20062	-3.53634
Splitting time – T1	13109	812	29478
Splitting time – T2	27321	5636	51863
Splitting time – T3	41294	13481	72222
Splitting time – T4	56350	24992	90800
Ne1	5241	1084	9795
Ne2	7724	1656	18804
Ne3	6106	1136	18199
Ne4	9678	2043	19532
Ne5	8335	1560	19136
NeAnc1	13674	570	34513
NeAnc2	14552	435	34407
NeAnc3	16461	992	34398
NeAnc4	16444	1190	33951
Migration rate 1	0.000052	3.3909*10^-6	0.000098
Migration rate 2	0.000044	3.30932*10^-6	0.000097
Migration rate 3	0.000049	1.92138*10^-6	0.000096
Migration rate 4	0.000051	3.95748*10^-6	0.000097
Migration rate 5	0.000051	1.85222*10^-6	0.000098
Migration rate 6	0.000050	2.43182*10^-6	0.000098
Migration rate 7	0.000051	3.32577*10^-6	0.000098
Migration rate 8	0.000049	2.22335*10^-6	0.000097

Online supplementary material

Appendix 1: Summary of the microsatellite and primer sequences and of the PCR conditions

Locus	Foward and reverse primer sequences	Repeat struture	Primer volume	Multiplex	Primer annealing temperature	Reference
An3	CACAACAACAGATGTATTTAC	(CA)14	0.05 µl			Zwettler et al. 2002
	GGACAATGGCAGCAGCAACAG		0.05 µl			
MAT5	TAAGAGGATAGAGAACCAAAC	(AT)5	0.1 µl			Feng (unpublished data)
	AACACATACACAAACACACAAG		0.1 µl	A	58 °C	
MAG2	GCAAACATTCTGTAATCTTCTTC	(AG)2	0.1 µl			Feng (unpublished data)
	TAAAAATCAACCTATTAAACGG		0.1 µl			
MAAC2	AGGCGTGGCACTCATCGGTG	(AAC)2	0.05 µl			Feng (unpublished data)
	TTCCTAAACTAGACCTTACAAC		0.05 µl			
FSAT34	AAAAATCCCTGCAACTGTCAC	(TAA)13	0.025 µl			Palero (unpublished data)
	TTTGACGAATTTACCCCTGGA		0.025 µl			
FSAT35	GGAGAGAAATCCCGACCTTTG	(TAG)10	0.05 µl			Palero (unpublished data)
	TAATTTCTCGGCTGAAAGCGA		0.05 µl			
FSAT68	GTTGGTGTTGGGGGCTATTTG	(GAA)12	0.05 µl	B	55 °C	Palero (unpublished data)
	CAGATCGATGGCCAAGTGTCG		0.05 µl			
FSAT61	TCGCGACACAATCGGTTTGGT	(TGA)11	0.025 µl			Palero (unpublished data)
	TCCGAGAGATGCAACAAGCCA		0.025 µl			
FSA58	TAAGGACGATTTGCGGTGTCA	(TGTA)9	0.025 µl			Palero (unpublished data)

	ATCAAAGACTCCATCCGCCCT		0.025 µl			
Ant11A.M2	GCATCAGCGTAATTTAATG	(CA)10	0.05 µl			Zwettler et al. 2002
	AAGAAGATGCCTTTGTGAG		0.05 µl			
Ant10A	CAAATCGACGGCTCAAATC	(TG)6	0.05 µl			Zwettler et al. 2002
	TCGACTCTCTTACATCTCC		0.05 µl			
				C	55 °C	Erasmus (unpublished data)
MSAT26	TCAATGGGCAAAATCAAATG	(TTC)10	0.1 µl			
	CAGGCAACTCCTCAGACAAA		0.1 µl			
MAG4	TGTAAGCAAGAAATGGAAAT	(AG)4	0.05 µl			Feng (unpublished data)
	CTGACAAAGGTTGAAATAGGA		0.05 µl			
An1	CCACACCAAAGTTTCCGACAG	(CT)15	0.05 µl			Zwettler et al. 2002
	CAACAAAAACCATAATCCTAG		0.05 µl			
						Erasmus (unpublished data)
MSAT18	GGATTCTCTCCGATTGCTGT	(CTT)10	0.1 µl			
	GCAGGTGATGTTGCCATTAG		0.1 µl			
				D	55 °C	Erasmus (unpublished data)
MSAT2	AGCAGAGGAGCGTGGAAGT	(TAA)13	0.1 µl			
	TGCCAATGAGTTATGGTGGA		0.1 µl			
						Erasmus (unpublished data)
MSAT10	CGGTTTGGTTCACAACGAC	(TGA)11	0.1 µl			
	GAAAGCATCTCACTCACTCTCAA		0.1 µl			
						Erasmus (unpublished data)
MSAT5	CTCAACCGCCACCAAAAC	(TAA)13	0.05 µl			
	CGGGAAGGGTAAAACCGTC		0.05 µl			
						Erasmus (unpublished data)
MSAT31	GACAAATCATCCGTAGAAGC	(AT)14	0.05 µl	E	55 °C	
	AAGAACTCCAAACATTCAAA		0.05 µl			
FSAT26	TGGTGGCACCGCGACGGTGAAC	(ATA)14	0.1 µl			Palero (unpublished data)
	CCGCACGGCTTTGCCGGAGA		0.1 µl			

Running title: Population genetic structure in *Antirrhinum majus*

Past colonization and geography define population genetic structure in *Antirrhinum majus*

(En préparation)

Khimoun A [1,2*], Palero F [3*], Andalo C [1,2] , Burrus M [1,2] , Thebaud C [1,2]

[1] Université de Toulouse; UPS; EDB (Laboratoire Evolution et Diversité Biologique); UMR5174; 118 route de Narbonne, F-31062 Toulouse cedex 9, France.

[2] CNRS; EDB (Laboratoire Evolution et Diversité Biologique); UMR5174; F-31062 Toulouse, France

[3] Unitat Mixta Genòmica i Salut CSISP-UV/Institut Cavanilles Universitat de Valencia, C/ Catedrático Jose Beltran 2, 46980 Paterna, Spain

*AK and FP have contributed equally to this work and should be considered sharing first authorship

Author for correspondence: khimoun@cict.fr

Phone number: 00 33 (0)5 61 55 74 86

Fax number: 00 33 (0)5 61 55 73 27

Abstract

Understanding the relative role of geographic distance and landscape and habitat features on population connectivity is critical to understand microevolutionary processes. *Antirrhinum majus* is a herbaceous short-lived perennial plant for which two subspecies, *A. m. striatum* (yellow-flowered) and *A. m. pseudomajus* (magenta-flowered) have been described. Populations of the two subspecies may occur under different environmental conditions, suggesting that local adaptation could explain the parapatric distribution of the two subspecies. By using a set of neutral microsatellite markers, we find genetic diversity in *A. majus* populations to follow a latitudinal gradient, with northern populations showing lower diversity and number of alleles. Furthermore, Bayesian clustering algorithms and F_{ST} methods both support the presence of 2-3 main groups of populations following a SW to NE pattern rather than grouping populations based on flower color. Our results indicate that recent expansion towards northern latitudes and gene-flow restriction due to geographical boundaries (i.e. Pyrenees) are the main factors structuring genetic diversity in *A. majus*.

Keywords: *Antirrhinum majus*, Pyrennees, speciation, population expansion.

INTRODUCTION

Within species, geographic differentiation may result from limited dispersal ability that prevents gene flow over large distance (Cain *et al.*, 2000). In consequence, increasing geographic distance between populations may impede genetic homogeneity and result in a spatial pattern of isolation-by-distance (Rousset, 1997; Rousset, 2000; Slatkin, 1993; Wright, 1946). Such a pattern is common in plants because the limited geographic distance that seeds and pollen can cross reduces dispersal and gene flow. Moreover, geographical barriers (e.g. mountain range, rivers, lakes...) can also constitute an important obstacle to population connectivity leading to the restriction of gene flow between populations. Ecological characteristics of habitat may also be at the origin of population genetic differentiation. Local adaptation of populations to different abiotic or biotic environments may act as barriers to gene flow by preventing introgression of foreign non-adapted alleles. Understanding the relative role of geographic distance and landscape and habitat features on population connectivity is thus critical to understand microevolutionary processes that generate population genetic differentiation across space (Manel *et al.*, 2003).

Mountain regions offer peculiar environmental conditions mainly due to their topology and to ecological constrains resulting from different climatic variables (low temperature, daily and seasonal range variation in temperature, etc). Because of its west/east orientation, the Pyrennean chain, the largest mountain chain in southern Europe, is impacted by different climatic influences resulting in heterogeneous habitats. In addition to contemporary topological and ecological constrains, historical events of the Quaternary glaciations may have shaped the evolution of species and of their diversity (Hewitt, 2000). The Pyrenees have acted as a strong geographic barrier to the post-glacial colonisation of northern Europe by several species. Indeed, several “suture zones” have been described in the Pyrenees, where genetic lineages from Iberia and from the Balkan met after post-glacial expansion (Butlin,

1998; Hewitt, 1993). Despite these distinct climatic and historic characteristics and their potential impact on the evolution of species and on their diversity, little attention has been paid on the distribution of plant genetic diversity in this region (but see Lauga *et al.*, 2009).

Antirrhinum majus (Scrophulariaceae) is a herbaceous short-lived perennial plant characterized by a patchy distribution in southern Europe (Figure 1). Two subspecies, *A. m. striatum* (yellow-flowered) and *A. m. pseudomajus* (magenta-flowered) have been described based on the flower color, which is mainly encoded by a single nuclear gene ROS1 (Rothmaler, 1956; Sutton, 1988; Whibley *et al.*, 2006). These two subspecies occupy largely parapatric geographic regions, centred over the Pyrenees (Figure 1), so that the geographic ranges of the two subspecies encompass both the northern and southern sides of the Pyrenean axial chain. *A. m. striatum* and *A. m. pseudomajus* come into contact at the margins of their ranges, where introgressive hybridization occurs (local replacement of *A. m. pseudomajus* by *A. m. striatum* is observed in the west part of the contact zone and conversely in the east part; see Khimoun *et al.*, 2011). Populations (allopatric and contact zone populations) of the two subspecies have been found to occur under different environmental conditions, suggesting that local adaptation could explain the parapatric distribution of the two subspecies (Khimoun *et al.*, Accepted). Although environmental factors can have an influence on the spatial distribution of the two subspecies, the distribution of their genetic diversity has never been investigated.

In the present study, we first assessed the level of neutral genetic diversity of the two subspecies and described the global pattern of genetic structure. Then we tested the relative impacts of limited dispersal, and landscape and habitat features on the genetic differentiation of *A. majus* populations. To determine whether limited gene flow is the main factor structuring the genetic diversity of *A. majus* populations, we searched for isolation-by-distance pattern. To test the potential impact of landscape feature on population genetic

structure, we tried to detect cryptic boundaries corresponding to breaks in gene flow and tried to correlate them with physical features. Finally, to test the potential impact of adaptation to particular habitat features on population genetic structure, we tested whether genetic diversity is structured according to the two subspecies flower colors or according to particular geographical or climatic features.

METHODS

Plant material sampling strategy

A total of 818 *Antirrhinum majus* plants were sampled from 2002 to 2007 in 55 allopatric or parapatric populations distributed over the geographic range of the species (Figure 1). Geographic coordinates of populations were recorded by using a GPS device (Garmin, Olathe, Kansas, USA). A numerical scoring system was used to rank magenta and yellow flower colour phenotypes visually, following methods developed by (Whibley *et al.*, 2006). Obviously, plants that displayed yellow flowers were classified as *A. m. striatum* whereas plants that displayed magenta flowers were classified as *A. m. pseudomajus*. For each individual, young leaves and shoot tips were collected and stored at -20°C until DNA was extracted by using the DNeasy Plant Mini kit (Qiagen).

Microsatellite amplification and Genetic diversity analyses

Total genomic DNA extraction was performed using the DNeasy Plant Mini kit (Qiagen). Microsatellite loci were amplified in four different multiplex reactions (Appendix 1 in the online supplementary material). Multiplex amplification reactions were carried out in a 10 µl reaction volume containing 2 µl genomic DNA, 3.5 µl MasterMix (Qiagen), and 4.5 µl of the

primer mix (Appendix 1 in the online supplementary material). The PCR thermal profile used was 94°C for 10 min for initial denaturation, followed by 24 cycles of 94°C for 30 s, the primer annealing temperature which differs depending on the different multiplexes (Appendix 1 in the online supplementary material) for 1 min 30 s and 72°C for 1 min, and a final 30 min extension at 60°C in a Mastercycler Gradient (Eppendorf). After preliminary analysis of chromatograms, a subset of 10 polymorphic microsatellite loci was kept for this study (Table 1). Amplified products were scored using an ABI 3730 automatic sequencer from the Plateforme Génomique de Toulouse (France). Alleles were sized by PeakScanner™ and Microsatelight software (Palero *et al.*, 2011), with an internal size marker Rox™ Size Standard 500 (Applied Biosystems).

Mean number of alleles per locus, observed (H_O) heterozygosity and expected (H_E) heterozygosity were estimated for each *A. majus* population with Microsatellite Toolkit v3.1 (Park, 2001). F_{IS} estimates according to (Robertson, Hill, 1984) and exact tests for conformity to Hardy-Weinberg expectations were obtained using the GENEPOP package v4.0.7 (Rousset, 2008). Where multiple tests were involved, significance levels were adjusted according to the sequential Bonferroni procedure. FreeNA was used for estimating null allele frequency at different loci and population differentiation taking null alleles into account (ENA method for estimating F_{ST} values; (Chapuis, Estoup, 2007).

Genetic structure analysis

Pairwise F_{ST} values obtained with FreeNA were used to determine the degree of population subdivision among different populations, and significance was assessed by bootstrapping over loci. The patterns of spatial genetic structure described as isolation-by-distance (IBD) models were evaluated by correlating the matrix of pairwise population differentiation in terms of F_{ST}

and the matrix of geographic distances (shortest geographical distance between sampling locations) using GENEPOP v4.0.7 (Rousset, 2008). Then, we investigated how genetic variance is distributed at different levels. We tested whether genetic variance is mostly related to flower color differences or related to differences between populations within subspecies. To do so, we performed an analysis of molecular variance (AMOVA) using Arlequin v.3.5 (Excoffier, Lischer, 2010). In order to identify the presence of barriers without previously defining groups of populations, we used the Monmonier's maximum difference algorithm as implemented in the software Barrier v. 2.2 (Manni *et al.*, 2004). This program takes the geographical coordinates and the genetic distances that separate the samples, and traces the barriers separating those pairs of populations between which the genetic distances are proportionally greater (Manni *et al.*, 2004).

To further assess the global scale genetic structure of *A. majus* populations, we estimated the putative number of genetic clusters (i.e. gene pools), K, using two different Bayesian clustering algorithm implemented in STRUCTURE (Pritchard *et al.*, 2000) and TESS (Chen *et al.*, 2007; Francois *et al.*, 2006). Under certain conditions (infinite island model with moderate connectivity), TESS has been claimed to outperform other Bayesian clustering methods including STUCTURE (Chen *et al.*, 2007). TESS is a spatially explicit model that takes into account the geographic location of individuals (Chen *et al.*, 2007; Durand *et al.*, 2009) contrary to STRUTURE which does not assume any prior information about population spatial location. For the two models, we ran the MCMC algorithm assuming an admixture model. For STRUCTURE, 100 replicates of each K from and K=1 to K=10 were performed, with a burn-in of 30 000 steps followed by 50 000 MCMC iterations. For each K, we only kept the 10 runs with the highest likelihood values and determined the most likely number of genetic clusters using the ΔK statistic (Evanno *et al.*, 2005). For TESS, 100 replicates of each K from K=2 to 10 were performed, with a burn-in of 30 000 steps followed

by 50 000 MCMC iterations. The spatial interaction parameter, which represents the strength of the spatial component, was set as default. Genetic structure was inferred from the effective number of clusters identified by the *K*- model where the decreasing Deviance Information Criterium (DIC) averaged over the 10 runs reached a plateau (Durand *et al.*, 2009). The 10 highest likelihood runs for each *K* were averaged using CLUMMP (Jakobsson, Rosenberg, 2007) with the *Greedy* algorithm with random input order and 1 000 permutations. Finally we represented spatially the median individual assignation to the different clusters per population using Mathematica (Wolfram 2008).

Finally, a recently introduced multivariate method called Discriminant Analysis of Principal Components (DAPC) was used in order to identify clusters of genetically related genotypes (Jombart *et al.*, 2010). By using Principal Component Analysis of the transformed (centred and scaled) genetic data as a prior step, DAPC ensures that variables submitted to Discriminant Analysis are perfectly uncorrelated, and that their number is less than that of analyzed genotypes. K-means clustering was used to identify groups of genotypes by specifying the actual number of clusters (k = 10). In all analyses, 100 principal components of PCA were retained in the data transformation step, and 7 axes were retained in the Discriminant Analysis step. The DAPC was performed using the function dapc implemented in the adegenet package in R (http://www.r-project.org/).

RESULTS

Microsatellite amplification and genetic diversity analyses

We obtained genotype data for all 10 markers in 50 out of the 55 initially sampled populations, keeping a total of 752 individuals for genetic analysis. The mean number of alleles per locus varied from 2.9 to 7 among populations, with both *A. m. pseudomajus* and *A.*

m. striatum populations showing the same mean number of alleles (mean = 4.54). The estimated mean observed heterozygosity (H_O = 0.530 ± 0.046) and the expected heterozygosity (H_E = 0.600 ± 0.058) for the global dataset were very similar, even though *A. m. striatum* populations showed a slightly larger observed (H_O = 0.539 ± 0.069) and expected heterozygosity (H_E = 0.620 ± 0.053) than *A. m. pseudomajus* populations (H_O = 0.525 ± 0.066 and H_E = 0.600 ± 0.060). Gene diversity levels (H_E) were significantly correlated with latitudinal but not to the longitudinal position (Table 2). No significant genotypic linkage disequilibrium among loci was found after sequential Bonferroni correction. Except for one locus (MSAT31), which showed not enough information for HWE testing, observed genotypes conformed to Hardy-Weinberg expectations. Statistically significant F_{IS} values were only found for some loci in a few populations, which may be indicative of genotyping errors and missing alleles (Appendix 2 in the online supplementary material).

Genetic structure analysis

Despite some deviations from HWE, global F_{ST} estimates did not vary much when using the raw dataset (F_{ST} = 0.1158; 95% CI = 0.0997 - 0.1373) or when correcting for the presence of null alleles with FreeNA (F_{ST} = 0.1118; 95% CI = 0.0966 - 0.1338). Population pairwise F_{ST} values did not vary much either when correcting or not for the presence of null alleles (Appendix 3 in the online supplementary material). Significant population differentiation was found in pairwise comparisons between most populations, with the mean F_{ST} value obtained for pairwise comparisons among *A. m. striatum* populations (F_{ST_yellow} = 0.106 $\pm$ 0.059) being similar to that found among *A. m. pseudomajus* populations (F_{ST_red} = 0.115 $\pm$ 0.061) and when comparing *A. m. striatum* and *A. m. pseudomajus* populations ($F_{ST\_yellow\text{-}red}$ = 0.129 $\pm$ 0.062). When the correlation of pairwise genetic distances [Fst/(1-Fst)] to geographic

distances [a + b ln(distance)] was analyzed, significant correlation was detected (a = 0.0606435, b = 0.01676910; p-value = 0.015) (Figure 2).

According to the AMOVA analyses, significant genetic differentiation can be found at all levels, with most of the genetic variance being due to variability within individuals (76.17%) and among individuals within populations (11.41%). Surprisingly, only 1.67% of the genetic variance was explained by differences related to flower color, while differences among populations within flower-color groups represented 10.75% of the genetic variance. This result implies that the flower-color groups do not reflect properly the genetic variance in *A. majus* populations. Significant genetic differentiation among populations within color groups implies that some populations within subspecies are more different than populations among different subspecies. Interestingly, the largest genetic barriers (using pairwise F_{ST} values) detected by Monmonier's maximum difference algorithm did not match with previous expectations given flower color, but seemed to correspond with geographic barriers related to the Pyrenees (Figure 3).

Using the ΔK statistic (Evanno *et al.*, 2005), the STRUCTURE clustering analysis revealed a main genetic structure in two clusters and a sub-structure in 3 clusters (Figure 4 and Appendix 4 in the online supplementary material). Using the lowest DIC value before the plateau, TESS analysis corroborated the previous structuration into K=2 or K=3 genetic clusters (Figure 4). For each of the two optimal number of clusters, STRUCTURE and TESS provided equivalent individual cluster assignation (Figure 4). From the inferred q-values, we can see that most of the individuals are admixed, so that within most of the sampled populations individuals are mainly assigned to different clusters. Nevertheless, when we spatially represent the median individual admixture proportion per population, we can observe a spatial pattern of genetic structure (Figure 4). Finally, the two main components of the

DAPC analysis based on the actual number of localities (k = 50) also presented populations from each subspecies forming two partially overlapping groups (Figure 5).

DISCUSSION

Levels of neutral genetic diversity for the two main *Antirrhinum majus* subspecies are assessed in the present study for the first time. Parameter estimates for the 2-dimensional correlation analysis of changes in genetic diversity along the latitudinal and longitudinal ranges of *Antirrhinum majus* populations indicate that the latitudinal gradient is significant, with genetic diversity being lower as we move northwards. This result is in agreement with previous studies in several plant species using different molecular techniques such as allozyme analysis in the moss *Leucodon sciurides* (Cronberg, 2000) and in sedges *Schoenus spp.* (Hedren, 1997) which show genetic diversity to be higher in southern Europe (the Iberian peninsula, Italy and the Balkans) and lower in northern latitudes. Similar patterns were also found in the white oak *Quercus spp.* (Dumolin-Lapegue *et al.*, 1997), black alder *Alnus glutinosa* (King, Ferris, 1998), heather *Calluna vulgaris* (Rendell, Ennos, 2002) and in hornbeam *Carpinus betulus* (Grivet, Petit, 2003).

The greater genetic diversity in the southern Europe may be due to the tendency of southern refugia to accumulate higher genetic diversity owing to their persistence and relative stability. Evolutionary processes such as extensive past gene flow and hybridisation during the glacial cycles may also have contributed to the southern - northern gradient of genetic diversity (Hewitt, 1996). The favourable climatic, ecological and environmental conditions in southern Europe have provided suitable habitats not only for plants but also for animals that escaped the glaciers in northern Europe. However, reduction in genetic diversity along northern Europe can be explained also by the “leading edge model”. This model implies that the peripheral isolates from larger stable populations (i.e. southern refugia) became separated

followed by successive founder events after the Last Ice Age, leading to rapid colonization of ice-free territories (Hewitt, 1999; Nordal, 1987).

When we tested the relative impacts of limited dispersal, bootstrap resampling over loci and the corresponding 95% Confidence interval for the global and pairwise F_{ST} estimates indicated that there is indeed a significant genetic structure in *A. majus* populations. If we consider the dichotomic structure of *A. majus* genetic diversity observed when using bayesian methods, we can see that *A. majus* individuals from the northern populations are mainly assigned the one cluster whereas individuals from the south-western populations are mainly assigned to the other cluster. Individual assignation to one or the other cluster varies along a south-west/north-east gradient, and the same gradient is observed when considering the trichotomic structuration with a third cluster occupying the center of the *A. majus* range. The highest assignation of populations to this third cluster remain very low (Mon: 0.35), and this cluster is localised between the two previous ones, so these results may suggest that this third cluster would rather be the result of an admixture of the northern and the southern clusters.

However, the DAPC results may provide an alternative explanation and point to the fact that populations belonging to this central region are also those located at high-altitude. It should be mentioned that Northern and southern *A. majus* populations are separated by high altitudes mountain crests (>1800 m which is the highest altitude recorded for an *A. majus* population) and the third cluster could well represent those populations being genetically isolated due to altitudinal differences. Populations in the south may also have survived the glacial cycles by ascending the mountains during the warm interglacial periods and descending during the cold glacial periods. In any case, according to both the dichotomic and trichotomic structures, the inferred genetic clusters included populations from the two main subspecies. In consequence, the bayesian clustering algorithms indicate that genetic structure is not mainly defined by flower colors but rather by spatial locations of populations.

CONCLUSION

The set of neutral markers used in the present study indicates that genetic differentiation in *Antirrhinum majus* populations does not depend on flower color classification. Significant levels of population structure were found both within and among subspecies, indicating reduced levels of effective gene flow among different populations. Most importantly, the strongest genetic barriers did not separate subspecies but rather indicated the influence of the Pyrenees on defining population genetic structure. Bayesian clustering algorithms supported the presence of 2-3 clusters within *A. majus* populations, which correspond to two main South-East and North-West clusters, together with a third cluster related to high-altitude populations. Our results indicate the relevance of geographical and climatic features on shaping genetic structure of plant populations and provide further support for the hypothesis of an adaptive role of flower color in *A. majus*.

BIBLIOGRAPHY

Butlin RK (1998) What do hybrid zones in general, and the *Chorthippus parallelus* zone in particular, tell us about speciation? In: *Endless forms: species, speciation* (eds. Howard DJ, Berlocher SH), pp. 367–378. Oxford University Press, New York.

Cain ML, Milligan BG, Strand AE (2000) Long-distance seed dispersal in plant populations. *American Journal of Botany* **87**, 1217-1227.

Chapuis MP, Estoup A (2007) Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution* **24**, 621-631.

Chen C, Durand E, Forbes F, Francois O (2007) Bayesian clustering algorithms ascertaining spatial population structure: a new computer program and a comparison study. *Molecular Ecology Notes* **7**, 747-756.

Cronberg N (2000) Genetic diversity of the epiphytic bryophyte *Leucodon sciuroides* in formerly glaciated versus nonglaciated parts of Europe. *Heredity* **84**, 710-720.

Dumolin-Lapegue S, Demesure B, Fineschi S, LeCorre V, Petit RJ (1997) Phylogeographic structure of white oaks throughout the European continent. *Genetics* **146**, 1475-1487.

Durand E, Jay F, Gaggiotti OE, Francois O (2009) Spatial Inference of Admixture Proportions and Secondary Contact Zones. *Molecular Biology and Evolution* **26**, 1963-1973.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**, 2611-2620.

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10**, 564-567.

Francois O, Ancelet S, Guillot G (2006) Bayesian clustering using hidden Markov random fields in spatial population genetics. *Genetics* **174**, 805-816.

Grivet D, Petit RJ (2003) Chloroplast DNA phylogeography of the hornbeam in Europe: Evidence for a bottleneck at the outset of postglacial colonization. *Conservation Genetics* **4**, 47-56.

Hedren M (1997) Genetic variation and hybridization in Swedish *Schoenus* (Cyperaceae). *Plant Systematics and Evolution* **204**, 21-37.

Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature* **405**, 907-913.

Hewitt GM (1993) After the ice - *parallelus* meets *erythropus* in the Pyrenees. In: *Hybrid zones and the evolutionary process* (ed. Harrison RG), pp. 140–164. Oxford University Press, Oxford.

Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society* **58**, 247-276.

Hewitt GM (1999) Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society* **68**, 87-112.

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23**, 1801-1806.

Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *Bmc Genetics* **11**.

Khimoun A, Burrus M, Andalo C, *et al.* (2011) Locally asymmetric introgressions between subspecies suggest circular range expansion at the *Antirrhinum majus* global scale. *Journal of Evolutionary Biology* **24**, 1433-1441.

Khimoun A, Cornuault J, Burrus M, *et al.* (Accepted) Ecology predicts parapatric distributions in two closely related *Antirrhinum majus* subspecies. *Evolutionary Ecology*.

King RA, Ferris C (1998) Chloroplast DNA phylogeography of *Alnus glutinosa* (L.) Gaertn. *Molecular Ecology* **7**, 1151-1161.

Lauga B, Malaval S, Largier G, Regnault-Roger C (2009) Two lineages of *Trifolium alpinum* (Fabaceae) in the Pyrenees: evidence from random amplified polymorphic DNA (RAPD) markers. *Acta Botanica Gallica* **156**, 317-330.

Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution* **18**, 189-197.

Manni F, Guerard E, Heyer E (2004) Geographic patterns of (genetic, morphologic, linguistic) variation: How barriers can be detected by using Monmonier's algorithm. *Human Biology* **76**, 173-190.

Nordal I (1987) Tabularsa after all - Botanical evidence for ice-free refugia in scandinavia reviewed. *Journal of Biogeography* **14**, 377-388.

Palero F, Gonzalez-Candelas F, Pascual M (2011) MICROSATELIGHT-Pipeline to Expedite Microsatellite Analysis. *Journal of Heredity* **102**, 247-249.

Park SDE (2001) Trypanotolerance in West African Cattle and the Population Genetic Effects of Selection, University of Dublin.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-959.

Rendell S, Ennos RA (2002) Chloroplast DNA diversity in *Calluna vulgaris* (heather) populations in Europe. *Molecular Ecology* **11**, 69-78.

Robertson A, Hill WG (1984) Deviations from Hardy-Weinberg proportions - Sampling variances and use in estimation of inbreeding coefficients. *Genetics* **107**, 703-718.

Rothmaler W (1956) Taxonomische monographie der Gattung *Antirrhinum* (ed. Akademie-Verlag), Berlin.

Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* **145**, 1219-1228.

Rousset F (2000) Genetic differentiation between individuals. *Journal of Evolutionary Biology* **13**, 58-62.

Rousset F (2008) GENEPOP ' 007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources* **8**, 103-106.

Slatkin M (1993) Isolation by distance in equilibrium and nonequilibrium populations. *Evolution* **47**, 264-279.

Sutton DA (1988) A revision of the tribe *Antirrhineae*, London and New York.

Whibley AC, Langlade NB, Andalo C, *et al.* (2006) Evolutionary paths underlying flower color variation in *Antirrhinum*. *Science* **313**, 963-966.

Wolfram Research, Inc. (2008) Mathematica Edition: Version 7.0. Wolfram Research, Inc. Champaign, Illinois.

Wright S (1946) Isolation by distance under diverse systems of mating. *Genetics* **31**, 39-59.

Tables

Locus	Primer sequence	Repeat structure	Source
MAT5	TAAGAGGATAGAGAACCAAAC	(AT)5	Feng (unpublished data)
	AACACATACACAAACACACAAG		
An1	CCACACCAAAGTTTCCGACAG	(CT)15	Zwettler et al. 2002
	CAACAAAAACCATAATCCTAG		
Ant11A.M2	GCATCAGCGTAATTTAATG	(CA)10	Zwettler et al. 2002
	AAGAAGATGCCTTTGTGAG		
MSAT18	GGATTCTCTCCGATTGCTGT	(CTT)10	Erasmus (unpublished data)
	GCAGGTGATGTTGCCATTAG		
MSAT31	GACAAATCATCCGTAGAAGC	(AT)14	Erasmus (unpublished data)
	AAGAACTCCAAACATTCAAA		
FSAT26	TGGTGGCACCGCGACGGTGAAC	(ATA)14	Palero (unpublished data)
	CCGCACGGCTTTGCCGGAGA		
MSAT5	CTCAACCGCCACCAAAAC	(TAA)13	Erasmus (unpublished data)
	CGGGAAGGGTAAAACCGTC		
FSAT61	TCGCGACACAATCGGTTTGGT	(TGA)11	Palero (unpublished data)

	TCCGAGAGATGCAACAAGCCA		
FSAT34	AAAAATCCCTGCAACTGTCAC	(TAA)13	Palero (unpublished data)
	TTTGACGAATTTACCCCTGGA		
FSAT35	GGAGAGAAATCCCGACCTTTG	(TAG)10	Palero (unpublished data)
	TAATTTCTCGGCTGAAAGCGA		

Table 1 : Microsatellite loci used in this analysis

Subspecies	Population	Longitude	Latitude	Altitude	N	H_E	H_O	Alleles
A. m. pseudomajus	**Arl**	2.617	42.451	369	12	0.690 ± 0.052	0.620 ± 0.047	4.9 ± 1.9
A. m. pseudomajus	**Bal**	0.145	42.509	750	16	0.617 ± 0.073	0.530 ± 0.041	5.3 ± 2.3
A. m. pseudomajus	**Ban**	3.12	42.49	61	19	0.567 ± 0.067	0.533 ± 0.041	4.5 ± 1.8
A. m. pseudomajus	**Bar**	0.322	42.124	480	12	0.552 ± 0.089	0.501 ± 0.051	3.5 ± 1.7
A. m. pseudomajus	**Bes**	2.667	42.211	195	11	0.634 ± 0.058	0.549 ± 0.050	4.7 ± 1.8
A. m. pseudomajus	**Cal**	1.832	42.104	838	12	0.673 ± 0.060	0.557 ± 0.050	4.8 ± 2.2
A. m. pseudomajus	**Chi**	1.86	42.77	1336	16	0.536 ± 0.082	0.463 ± 0.042	3.9 ± 2.1
A. m. pseudomajus	**Cla**	3.096	43.177	120	12	0.485 ± 0.076	0.434 ± 0.052	2.9 ± 1.6
A. m. pseudomajus	**Col**	2.621	41.874	250	12	0.612 ± 0.071	0.522 ± 0.049	4.3 ± 1.6
A. m. pseudomajus	**Div**	1.85	42.26	842	10	0.626 ± 0.072	0.616 ± 0.051	4.6 ± 2.2
A. m. pseudomajus	**Fai**	1.149	42.161	1215	12	0.579 ± 0.075	0.486 ± 0.052	4.2 ± 1.9
A. m. pseudomajus	**Hor**	1.348	42.325	1128	41	0.586 ± 0.086	0.487 ± 0.028	5.2 ± 2.0
A. m. pseudomajus	**Inf**	1.031	42.274	584	12	0.616 ± 0.087	0.500 ± 0.050	4.7 ± 2.1
A. m. pseudomajus	**Lag**	2.584	43.09	149	16	0.595 ± 0.089	0.554 ± 0.044	4.6 ± 2.2
A. m. pseudomajus	**Lan**	0.316	42.169	1090	12	0.534 ± 0.109	0.492 ± 0.056	4.5 ± 2.5
A. m. pseudomajus	**Mrt**	3.89	43.645	40	12	0.487 ± 0.071	0.392 ± 0.047	3.2 ± 1.1
A. m. pseudomajus	**Mst**	1.822	41.609	500	13	0.644 ± 0.083	0.548 ± 0.048	4.8 ± 2.1
A. m. pseudomajus	**Nue**	0.425	42.285	850	18	0.542 ± 0.113	0.439 ± 0.046	4.4 ± 2.4
A. m. pseudomajus	**Pan**	0.256	42.738	1370	13	0.582 ± 0.089	0.490 ± 0.056	3.7 ± 1.4
A. m. pseudomajus	**Par**	2.195	42.314	1117	23	0.619 ± 0.075	0.561 ± 0.037	5.3 ± 1.5
A. m. pseudomajus	**Per**	2.849	42.476	148	33	0.697 ± 0.044	0.572 ± 0.031	5.2 ± 2.0
A. m. pseudomajus	**Por**	1.54	42.21	1851	11	0.480 ± 0.085	0.438 ± 0.055	3 ± 1.6
A. m. pseudomajus	**Pra**	2.496	42.382	1146	12	0.620 ± 0.079	0.488 ± 0.048	4.6 ± 1.7
A. m. pseudomajus	**Rip**	2.202	42.215	709	12	0.554 ± 0.069	0.477 ± 0.047	4.5 ± 1.7
A. m. pseudomajus	**Sal**	1.74	42.23	1126	12	0.668 ± 0.064	0.686 ± 0.046	4.9 ± 2.0
A. m. pseudomajus	**Sop**	0.738	42.331	854	16	0.684 ± 0.077	0.657 ± 0.042	7 ± 3.4
A. m. pseudomajus	**Sor**	1.14	42.392	960	12	0.591 ± 0.100	0.539 ± 0.052	4.2 ± 2.4
A. m. pseudomajus	**Sue**	0.737	42.415	865	13	0.599 ± 0.089	0.503 ± 0.045	4.8 ± 2.4
A. m. pseudomajus	**Ult**	2.98	42.516	500	12	0.694 ± 0.061	0.601 ± 0.046	5.5 ± 2.6
A. m. pseudomajus	**Una**	1.79	42.76	915	19	0.614 ± 0.059	0.551 ± 0.037	4.7 ± 2.3
A. m. pseudomajus	**Vie**	0.757	42.738	881	12	0.645 ± 0.071	0.488 ± 0.050	4.5 ± 1.8
A. m. striatum	**And**	1.59	42.57	1600	10	0.637 ± 0.077	0.540 ± 0.053	4.1 ± 1.5
A. m. striatum	**Ave**	1.32	42.353	1241	12	0.599 ± 0.098	0.450 ± 0.049	4.5 ± 2.2
A. m. striatum	**Cam**	1.92	42.8	1241	16	0.657 ± 0.060	0.692 ± 0.038	4.9 ± 2.1
A. m. striatum	**Els**	1.368	42.289	594	12	0.635 ± 0.042	0.524 ± 0.050	4.6 ± 1.8
A. m. striatum	**Fab**	2.615	42.593	381	13	0.618 ± 0.090	0.568 ± 0.054	3.9 ± 2.0
A. m. striatum	**For**	1.463	42.425	849	10	0.633 ± 0.084	0.583 ± 0.054	4.9 ± 2.7
A. m. striatum	**Iso**	1.81	42.37	1066	17	0.544 ± 0.086	0.524 ± 0.038	3.7 ± 1.6
A. m. striatum	**Jos**	1.64	42.26	1456	14	0.672 ± 0.071	0.471 ± 0.047	5.1 ± 2.3
A. m. striatum	**Lle**	1.668	42.363	960	12	0.660 ± 0.069	0.515 ± 0.050	4.7 ± 2.1
A. m. striatum	**Lu**	2.26	42.97	227	17	0.572 ± 0.075	0.591 ± 0.053	3.8 ± 1.3
A. m. striatum	**Mar**	2.622	42.552	628	26	0.661 ± 0.054	0.555 ± 0.032	5.9 ± 2.7
A. m. striatum	**Mon**	2.122	42.507	1564	11	0.451 ± 0.081	0.361 ± 0.051	3.1 ± 1.3
A. m. striatum	**Pal**	1.856	42.307	1920	11	0.626 ± 0.079	0.547 ± 0.051	4.4 ± 1.4
A. m. striatum	**Pom**	2.272	43.112	262	17	0.616 ± 0.041	0.467 ± 0.049	3.8 ± 1.7
A. m. striatum	**Thu**	2.721	42.644	130	15	0.646 ± 0.060	0.572 ± 0.043	5.2 ± 2.0
A. m. striatum	**Tos**	1.928	42.36	1499	35	0.672 ± 0.074	0.560 ± 0.030	6.3 ± 3.2

Table 2: Geographic location and diversity statistics for the *A. majus* populations sampled along the Pyrenees. Sample size (N), observed (H_O) and expected (H_E) heterozygosities and mean number of alleles per population.

	Structure	Variation (%)	F-statistic	Probability
Yellow / Red flowers (n = 752)	Among groups	1.67	F_{CT}: 0.0167	**P < 0.001**
	Among populations within groups	10.75	F_{SC}: 0.1094	**P < 0.001**
	Among individuals within populations	11.41	F_{IS}: 0.1310	**P < 0.001**
	Within individuals	76.17	F_{IT}: 0.1413	**P < 0.001**

Table 3: Global Analyses of Molecular Variance as a weighted average over loci carried out to compare the effect of flower color on the genetic structuring of *A. majus* populations. Significant *P* values are in bold.

List of Figures

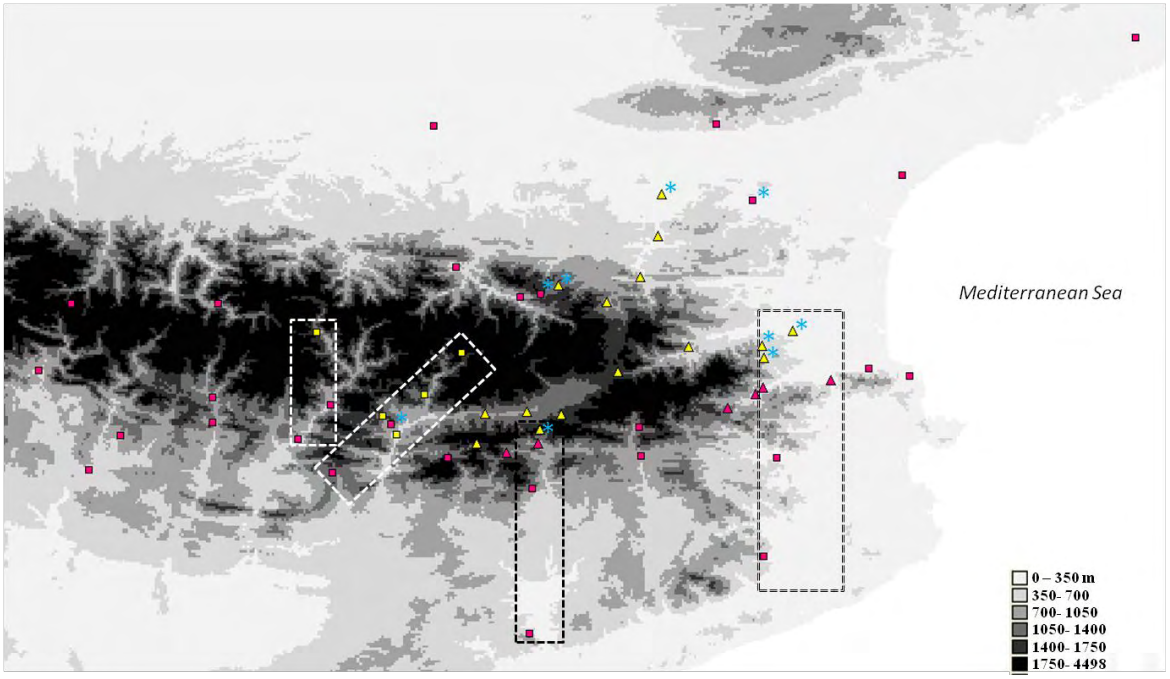


Figure 1: Spatial distribution of *Antirrhinum majus* populations

A. m. pseudomajus and *A. m. striatum* non-introgressed populations are represented by pink squares and yellow triangles respectively. *A. m. pseudomajus* and *A. m. striatum* introgressed populations are represented by pink triangles and yellow squares respectively. Populations that were polymorphic for the ROS1 locus are highlighted using a blue asterisk. Population grouping according to the 4 different valleys is symbolized by dashed lines, altitudinal ranges are represented using black and white shading for clarity.

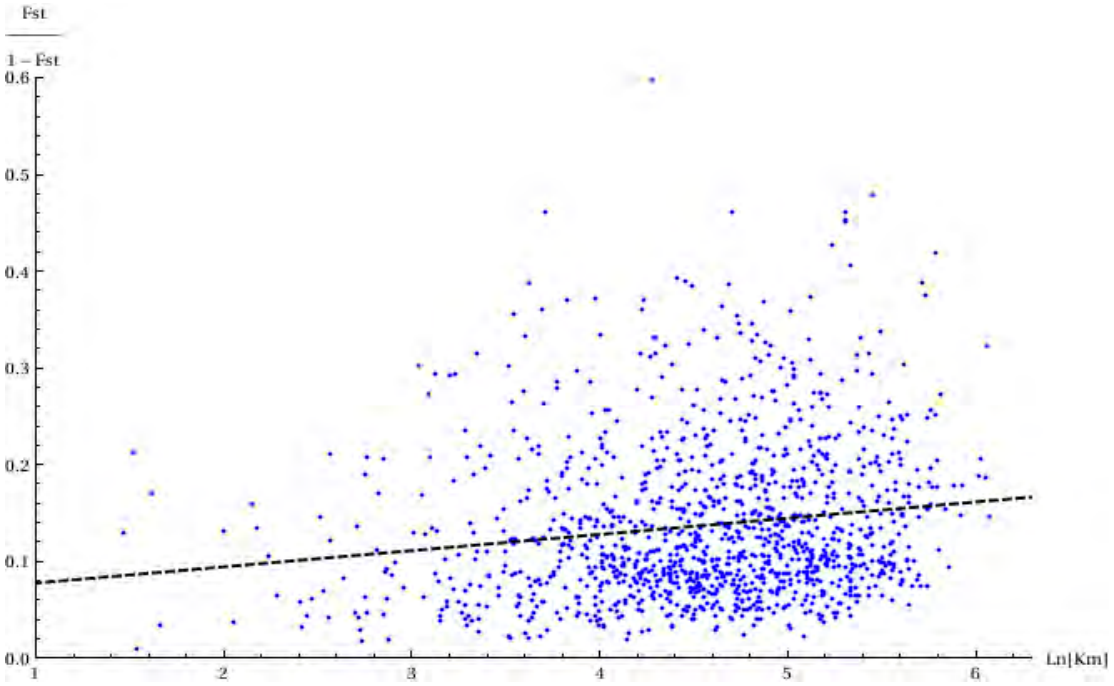


Figure 2: Scatterplot of genetic distance versus spatial distance between different *A. majus* populations.

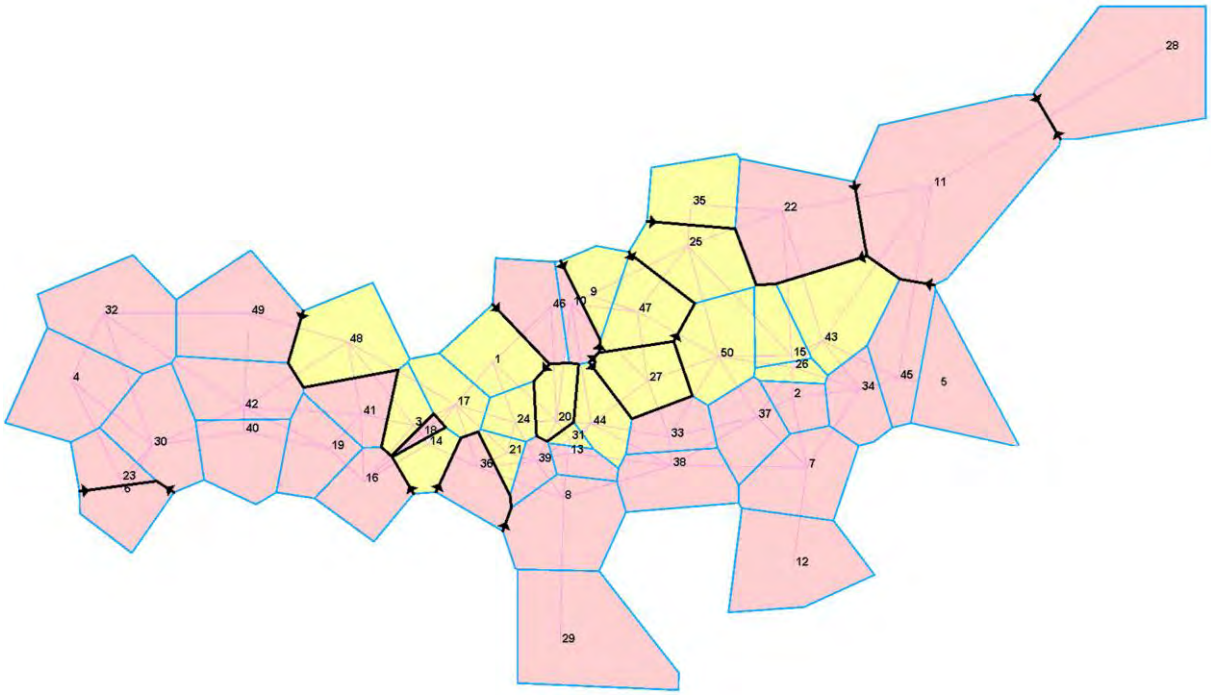


Figure 3: Main genetic barriers identified through Monmonier's maximum difference algorithm (black thick lines).

A. majus striatum populations (yellow) seem to show a different level of genetic differentiation among them and with *A. majus pseudomajus* populations (pink). Genetic barriers separate *A. majus* populations from Eastern-Western localities rather than by flower color.

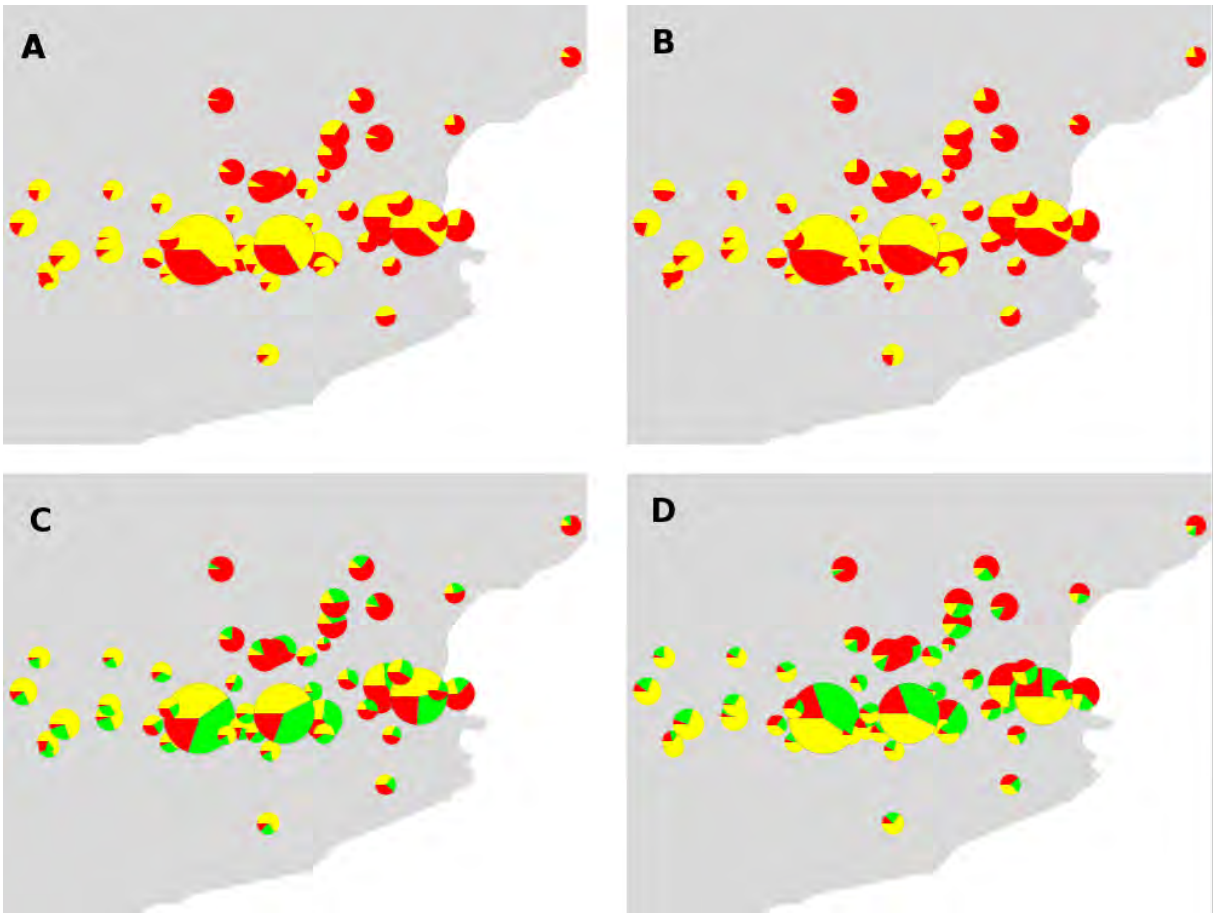


Figure 4 : Plot of the estimates of *Q* (median of estimated membership coefficients for each individual within a sampled population) for each cluster (*K*).

A and C were obtained with TESS and the most probable number of genetic clusters were K=2 and K=3 respectively. B and D were obtained with STRUCTURE and the most probable number of genetic clusters were K=2 and K=3 respectively. The vertical lines are broken into coloured segments showing the proportion of each individual assigned to each of the inferred clusters. Letters at the bottom of the graph correspond to the codes for the sampling locations.

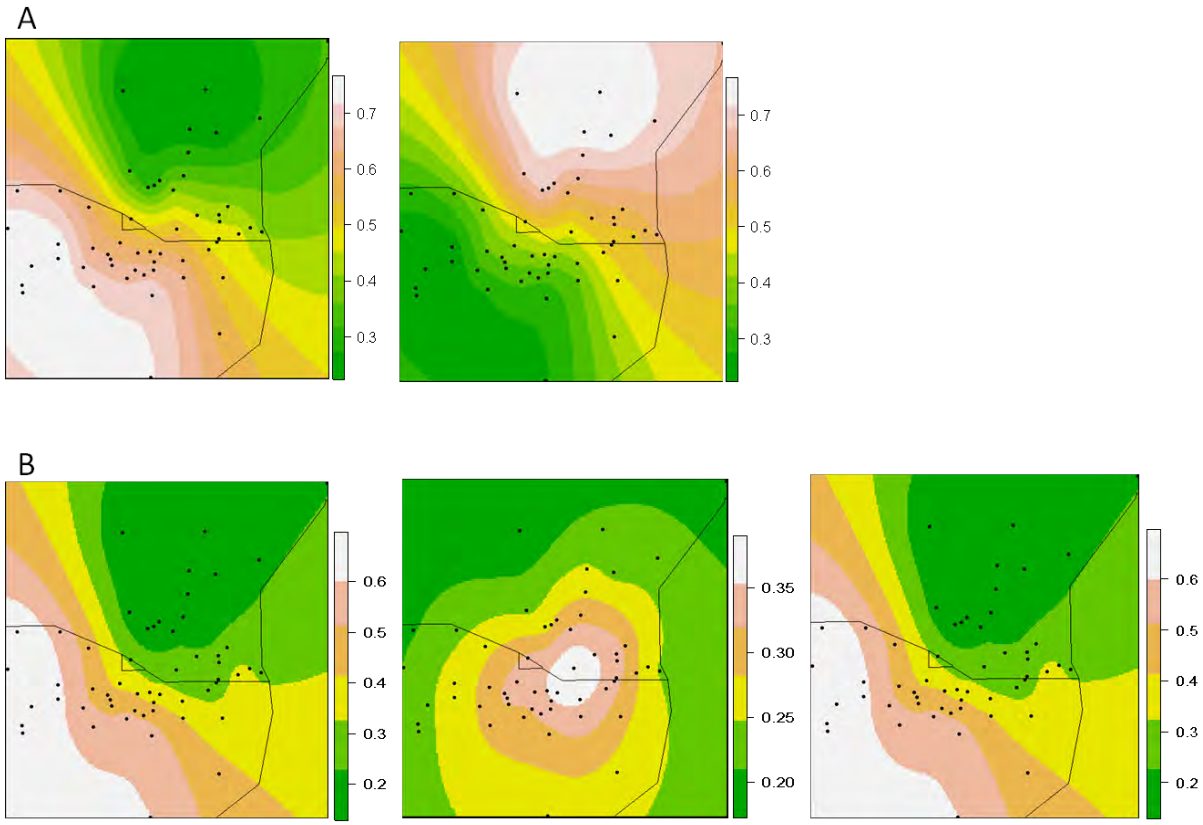


Figure 4 : Spatial display of the population membership obtained from TESS with K=2 in A/ and K=3 in B

Figure 5. Discriminant Analysis of Principal Components (DAPC) of *Antirrhinum majus* allelic data.

The scatterplot shows the first two principal components of the DAPC, using sampling localities as prior clusters. *A. m. pseudomajus* (red) and *A. m. striatum* (yellow) populations are shown by inertia ellipses, while dots represent individuals.

SUPPLEMENTARY MATERIALS

Locus	Foward and reverse primer sequences	Repeat struture	Primer volume	Multiplex	Primer annealing temperature	Reference
An3	CACAACAACAGATGTATTTAC	(CA)14	0.05 µl			Zwettler et al. 2002
	GGACAATGGCAGCAGCAACAG		0.05 µl			
MAT5	TAAGAGGATAGAGAACCAAAC	(AT)5	0.1 µl			Feng (unpublished data)
	AACACATACACAAACACACAAG		0.1 µl	A	58 °C	
MAG2	GCAAACATTCTGTAATCTTCTTC	(AG)2	0.1 µl			Feng (unpublished data)
	TAAAAATCAACCTATTAAACGG		0.1 µl			
MAAC2	AGGCGTGGCACTCATCGGTG	(AAC)2	0.05 µl			Feng (unpublished data)
	TTCCTAAACTAGACCTTACAAC		0.05 µl			
FSAT34	AAAAATCCCTGCAACTGTCAC	(TAA)13	0.025 µl			Palero (unpublished data)
	TTTGACGAATTTACCCCTGGA		0.025 µl			
FSAT35	GGAGAGAAATCCCGACCTTTG	(TAG)10	0.05 µl			Palero (unpublished data)
	TAATTTCTCGGCTGAAAGCGA		0.05 µl	B	55 °C	
FSAT68	GTTGGTGTTGGGGGCTATTTG	(GAA)12	0.05 µl			Palero (unpublished data)
	CAGATCGATGGCCAAGTGTCG		0.05 µl			

FSAT61	TCGCGACACAATCGGTTTGGT	(TGA)11	0.025 µl			Palero (unpublished data)
	TCCGAGAGATGCAACAAGCCA		0.025 µl			
FSA58	TAAGGACGATTTGCGGTGTCA	(TGTA)9	0.025 µl			Palero (unpublished data)
	ATCAAAGACTCCATCCGCCCT		0.025 µl			
Ant11A.M2	GCATCAGCGTAATTTAATG	(CA)10	0.05 µl			Zwettler et al. 2002
	AAGAAGATGCCTTTGTGAG		0.05 µl			
Ant10A	CAAATCGACGGCTCAAATC	(TG)6	0.05 µl			Zwettler et al. 2002
	TCGACTCTCTTACATCTCC		0.05 µl			
MSAT26	TCAATGGGCAAAATCAAATG	(TTC)10	0.1 µl	C	55 °C	Erasmus (unpublished data)
	CAGGCAACTCCTCAGACAAA		0.1 µl			
MAG4	TGTAAGCAAGAAATGGAAAT	(AG)4	0.05 µl			Feng (unpublished data)
	CTGACAAAGGTTGAAATAGGA		0.05 µl			
An1	CCACACCAAAGTTTCCGACAG	(CT)15	0.05 µl			Zwettler et al. 2002
	CAACAAAAACCATAATCCTAG		0.05 µl			
MSAT18	GGATTCTCTCCGATTGCTGT	(CTT)10	0.1 µl	D	55 °C	Erasmus (unpublished data)
	GCAGGTGATGTTGCCATTAG		0.1 µl			
MSAT2	AGCAGAGGAGCGTGGAAGT	(TAA)13	0.1 µl			Erasmus (unpublished data)

	TGCCAATGAGTTATGGTGGA		0.1 µl			
MSAT10	CGGTTTGGTTCACAACGAC	(TGA)11	0.1 µl			Erasmus (unpublished data)
	GAAAGCATCTCACTCACTCTCAA		0.1 µl			
MSAT5	CTCAACCGCCACCAAAAC	(TAA)13	0.05 µl			Erasmus (unpublished data)
	CGGGAAGGGTAAAACCGTC		0.05 µl			
MSAT31	GACAAATCATCCGTAGAAGC	(AT)14	0.05 µl	E	55 °C	Erasmus (unpublished data)
	AAGAACTCCAAACATTCAAA		0.05 µl			
FSAT26	TGGTGGCACCGCGACGGTGAAC	(ATA)14	0.1 µl			Palero (unpublished data)
	CCGCACGGCTTTGCCGGAGA		0.1 µl			

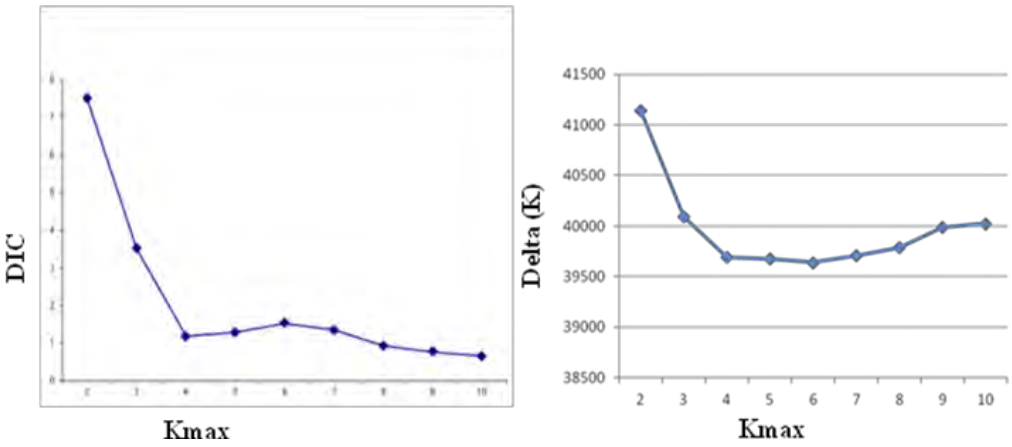
Appendix 1: Summary of the microsatellite and primer sequences and of the PCR conditions

Appendix 2. F_{IS} estimates according to Robertson and Hill (1984) and exact tests for conformity to Hardy-Weinberg expectations as obtained using GENEPOP. Significant values after Bonferroni correction **(P < 0.001)** are highlighted. Blank cells correspond to cases where F_{IS} could not be estimated due to missing data.

	MAT5	AN1	Ant11	MSAT18	MSAT31	FSAT26	MSAT5	FSAT61	FSAT34	FSAT35
And	0.056	-0.133	0.114	0.011		0.531	0.179	0.358	0.108	0.039
Arl	-0.061	-0.164	0.098	-0.037	0.239	0.118	0.363	0.476	-0.044	0.000
Ave	0.025	-0.015	0.193	-0.146		0.164	**0.377**	0.750	0.100	
Bal	0.014	0.033	0.141	-0.062	-0.116	0.053	0.151	0.499	0.073	-0.028
Ban	0.250	-0.135	0.034	-0.240	-0.079	0.141	0.384	-0.044	-0.091	-0.016
Bar	-0.141	0.073	0.543	-0.385		0.063	0.368	0.609	0.017	-0.002
Bes	-0.025	-0.125	0.894	-0.236	0.138	0.122	0.148	0.287	0.009	**0.407**
Cal	-0.129	-0.004	0.299	0.693	-0.519	0.203	-0.007	0.442	0.037	0.548
Cam	-0.060	0.156	-0.247	0.070	-0.086	-0.067	-0.029	-0.031	-0.096	-0.076
Chi	-0.027	0.016	0.028	-0.043	-0.074	0.010	0.688	0.467	0.689	0.064
Cla	-0.034	0.113	0.156	-0.009	1.200		0.138	0.242	-0.117	
Col	-0.059	0.115	**0.533**		-0.125	0.408	0.095	0.235	-0.046	-0.104
Div	0.012	0.196	-0.045	0.056	-0.049	-0.056	0.264	0.310	-0.203	
Els	-0.048	0.185	0.475	0.104	0.313	0.053	0.108	0.531	0.085	-0.241
Fab	0.022	0.217	-0.065	-0.066			0.760	-0.179	0.035	-0.010
Fai	0.031	0.072	0.344	-0.108		0.414	0.118	0.296	-0.019	-0.028
For	0.046	-0.031	0.430	-0.276		0.017	0.563	0.039	0.032	-0.120
Hor	0.122	-0.055	**0.315**	0.030		0.144	**0.317**	0.119	0.185	0.022
Inf	-0.079	**0.359**	0.069	-0.002		-0.063	0.988	0.400	0.015	-0.067
Iso	0.013	-0.052	0.054	-0.250		0.033	**0.308**	1.063	0.275	-0.087
Jos	0.003	0.053	0.566	0.140	-0.062	0.056	0.558	0.474	**0.658**	**0.407**
Lag	-0.053	0.404	0.010	0.206		0.375	-0.022	-0.082	-0.069	-0.089
Lan	0.146	0.075	-0.033	-0.032		-0.037	0.225	-0.133	0.275	
Lle	0.568	-0.204	**0.200**	0.178		0.011	0.377	0.786	-0.127	-0.007
Lu	0.228	-0.031	-0.065	-0.125		0.031	0.007	-0.134	-0.067	-0.222
mars	0.013	-0.006	0.495	0.013	0.109	**0.165**	0.311	**0.131**	0.208	0.056
Mon	0.050	0.130	0.470	0.021		0.039	-0.150	0.285		0.068
Mrt	-0.236	-0.085	0.135	0.016		0.144	**1.125**	0.344	0.394	-0.002
Mst	-0.036	0.042	1.083	-0.036	0.069	-0.107	0.536	0.330	-0.044	-0.050
Nue	0.060	0.042	-0.054			-0.057	**0.602**	0.369	**0.361**	
Pal	0.043	-0.081	0.055	0.015		0.135	**0.458**	0.282	-0.120	
Pan	-0.117	0.191	0.111	-0.026		-0.250	1.143	0.393	0.319	-0.063
Par	-0.038	-0.040	0.121	-0.017	-0.060	0.027	0.161	0.209	0.069	-0.004
Per	-0.053	0.099	**0.216**	0.255	0.535	0.160	**0.540**	-0.019	0.081	-0.095
Pom	-0.034	-0.112	0.354	0.231	2.000	0.071	0.823	0.151	0.154	-0.286
Por	-0.032	-0.147	0.415	-0.382		-0.004	**1.143**		0.045	-0.267
Pra	0.018	0.192	0.116	0.162	1.091	0.179	0.722	0.169	-0.060	-0.005
Rip	-0.050	-0.111	0.245	-0.024		0.008	**0.528**	0.352	0.011	-0.008
Sal	-0.004	0.120	0.232	-0.046	-0.164	-0.095	0.269	-0.188	0.083	-0.175
Sop	0.028	-0.082	0.128	-0.036		0.153	0.073	0.169	-0.038	-0.046
Sor	-0.038	-0.033	0.361	0.101			0.540	0.245	-0.121	
Sue	-0.008	0.007	0.349	0.220		0.088	-0.038	**0.553**	-0.122	-0.050
Thu	-0.094	0.020	0.062	-0.126	0.691	-0.024	0.374	0.259	0.344	-0.024
Tos	-0.001	-0.059	**0.171**	0.127	-0.085	0.006	**0.378**	**0.329**	0.064	-0.070
Ult	-0.067	0.165	0.136	0.013	-0.183	0.092	0.355	0.082	0.044	-0.053
Una	0.071	-0.029	0.482	0.168	0.217	-0.026	0.387	0.050	-0.044	-0.062
Uss	-0.020	0.011	0.284	0.214		0.069	0.372	0.254	0.031	0.331
Val	-0.007	-0.085	-0.113	0.202		0.098	-0.076	0.273	-0.100	-0.143
Vie	0.129	0.028	0.050	0.128		0.322	0.523	**0.943**	0.067	-0.172
Vil	-0.029	-0.121	0.123	0.217		0.072	-0.115	0.382	-0.160	-0.056

Appendix 3. - F_{ST} values observed in population pairwise comparisons before (below diagonal) and after (above diagonal) correcting for the presence of null alleles. Significant values are shown in bold (P < 0.05). See attached excel file.

		Arl	Ave	Bal	Ban	Bar	Bes	Cal	Cam	Chi	Cla	Col	Div	Els	Fab	Fai	For	Hor	Inf	Iso	Jos	Lag	Lan	Lle	Lu	Mar	Mon	Mrt	Mst	Nue	Pal	Pan	Par	Per	Pom	Por	Pra	Rip	Sal	Sop	Sor	Sue	Thu	Tos	Ult	Una	Uss	Val	Vie	Vil
And		0,150	0,109	0,127	0,205	0,219	0,151	0,131	0,098	0,185	0,211	0,215	0,160	0,179	0,116	0,138	0,058	0,177	0,149	0,116	0,108	0,166	0,192	0,076	0,152	0,128	0,171	0,201	0,196	0,246	0,149	0,173	0,139	0,154	0,184	0,264	0,149	0,178	0,154	0,143	0,170	0,147	0,159	0,086	0,106	0,172	0,131	0,165	0,138	0,104
Arl	0,156		0,123	0,074	0,119	0,138	0,032	0,058	0,079	0,129	0,172	0,024	0,050	0,113	0,044	0,080	0,127	0,062	0,042	0,162	0,080	0,081	0,089	0,122	0,120	0,031	0,185	0,117	0,054	0,077	0,048	0,088	0,044	0,054	0,072	0,225	0,027	0,081	0,049	0,067	0,055	0,074	0,054	0,084	0,054	0,078	0,065	0,112	0,112	0,059
Ave	0,122	0,127		0,083	0,133	0,211	0,104	0,103	0,133	0,164	0,189	0,127	0,104	0,118	0,103	0,099	0,084	0,115	0,099	0,152	0,025	0,164	0,152	0,061	0,118	0,079	0,207	0,137	0,156	0,166	0,077	0,155	0,079	0,093	0,152	0,179	0,096	0,141	0,058	0,123	0,114	0,068	0,108	0,047	0,049	0,147	0,089	0,111	0,091	0,070
Bal	0,139	0,077	0,090		0,149	0,171	0,091	0,039	0,115	0,165	0,213	0,051	0,098	0,158	0,129	0,064	0,123	0,088	0,028	0,186	0,088	0,143	0,116	0,100	0,171	0,073	0,248	0,127	0,083	0,105	0,085	0,092	0,083	0,079	0,089	0,204	0,056	0,060	0,064	0,080	0,070	0,073	0,084	0,074	0,068	0,143	0,053	0,077	0,082	0,083
Ban	0,217	0,120	0,151	0,156		0,198	0,085	0,120	0,133	0,111	0,225	0,137	0,126	0,053	0,094	0,139	0,087	0,117	0,117	0,172	0,085	0,113	0,133	0,142	0,088	0,085	0,315	0,102	0,118	0,150	0,106	0,163	0,070	0,099	0,109	0,215	0,084	0,152	0,145	0,153	0,087	0,147	0,126	0,089	0,058	0,119	0,125	0,111	0,123	0,087
Bar	0,235	0,140	0,221	0,186	0,204		0,129	0,128	0,184	0,205	0,294	0,167	0,168	0,224	0,198	0,114	0,211	0,139	0,139	0,247	0,141	0,200	0,144	0,198	0,226	0,174	0,311	0,244	0,162	0,143	0,173	0,152	0,161	0,183	0,199	0,238	0,166	0,195	0,142	0,094	0,191	0,189	0,171	0,159	0,162	0,215	0,162	0,137	0,159	0,167
Bes	0,178	0,032	0,114	0,102	0,080	0,137		0,071	0,108	0,093	0,189	0,068	0,042	0,129	0,078	0,090	0,106	0,066	0,061	0,147	0,057	0,062	0,083	0,106	0,136	0,056	0,264	0,060	0,069	0,102	0,058	0,128	0,040	0,059	0,109	0,213	0,037	0,100	0,055	0,067	0,052	0,087	0,041	0,066	0,059	0,093	0,100	0,131	0,117	0,051
Cal	0,144	0,055	0,109	0,037	0,122	0,140	0,079		0,113	0,169	0,214	0,047	0,081	0,090	0,092	0,074	0,125	0,081	0,037	0,164	0,070	0,133	0,074	0,107	0,141	0,052	0,250	0,142	0,038	0,078	0,079	0,072	0,093	0,047	0,087	0,208	0,062	0,081	0,042	0,072	0,077	0,095	0,082	0,078	0,048	0,133	0,048	0,083	0,043	0,084
Cam	0,108	0,077	0,146	0,120	0,136	0,195	0,120	0,124		0,116	0,216	0,129	0,126	0,147	0,032	0,110	0,075	0,103	0,094	0,118	0,079	0,076	0,145	0,096	0,077	0,051	0,174	0,168	0,147	0,166	0,102	0,137	0,093	0,117	0,108	0,244	0,080	0,119	0,114	0,106	0,103	0,125	0,093	0,052	0,057	0,119	0,068	0,092	0,122	0,045
Chi	0,205	0,134	0,187	0,187	0,126	0,222	0,108	0,193	0,128		0,231	0,154	0,175	0,186	0,132	0,201	0,120	0,137	0,113	0,157	0,086	0,045	0,173	0,175	0,121	0,122	0,315	0,162	0,178	0,235	0,125	0,226	0,099	0,156	0,143	0,237	0,096	0,212	0,157	0,177	0,104	0,186	0,134	0,104	0,111	0,036	0,155	0,184	0,160	0,092
Cla	0,217	0,175	0,198	0,220	0,221	0,306	0,199	0,222	0,219	0,243		0,229	0,215	0,238	0,169	0,252	0,222	0,229	0,186	0,271	0,141	0,204	0,200	0,185	0,252	0,196	0,269	0,248	0,192	0,272	0,179	0,203	0,161	0,173	0,187	0,315	0,173	0,236	0,170	0,232	0,190	0,197	0,184	0,150	0,137	0,196	0,201	0,239	0,191	0,211
Col	0,227	0,025	0,134	0,058	0,137	0,165	0,066	0,039	0,137	0,168	0,237		0,091	0,149	0,116	0,091	0,187	0,052	0,034	0,223	0,094	0,117	0,106	0,182	0,178	0,060	0,277	0,143	0,065	0,093	0,076	0,119	0,079	0,067	0,096	0,237	0,047	0,076	0,047	0,086	0,074	0,101	0,081	0,107	0,081	0,127	0,078	0,096	0,111	0,103
Div	0,167	0,048	0,110	0,099	0,132	0,174	0,037	0,080	0,126	0,193	0,217	0,089		0,120	0,076	0,057	0,148	0,084	0,073	0,173	0,073	0,161	0,073	0,116	0,148	0,067	0,207	0,128	0,060	0,073	0,032	0,092	0,043	0,063	0,133	0,190	0,062	0,050	0,064	0,096	0,072	0,032	0,080	0,076	0,058	0,137	0,077	0,118	0,109	0,098
Els	0,182	0,112	0,131	0,161	0,049	0,232	0,131	0,094	0,144	0,202	0,248	0,149	0,122		0,067	0,159	0,089	0,175	0,141	0,156	0,076	0,155	0,150	0,109	0,083	0,080	0,280	0,154	0,106	0,167	0,123	0,146	0,108	0,079	0,111	0,232	0,132	0,162	0,146	0,159	0,121	0,159	0,128	0,087	0,062	0,153	0,129	0,124	0,133	0,114
Fab	0,110	0,041	0,120	0,140	0,092	0,209	0,078	0,100	0,029	0,141	0,169	0,120	0,074	0,068		0,096	0,064	0,094	0,060	0,068	0,049	0,091	0,075	0,080	0,072	0,008	0,178	0,104	0,099	0,118	0,065	0,098	0,031	0,069	0,070	0,250	0,051	0,073	0,077	0,121	0,076	0,093	0,040	0,030	0,027	0,112	0,072	0,099	0,104	0,036
Fai	0,150	0,080	0,113	0,078	0,149	0,127	0,096	0,082	0,118	0,228	0,258	0,100	0,060	0,162	0,104		0,117	0,052	0,078	0,169	0,096	0,161	0,075	0,110	0,154	0,087	0,260	0,133	0,096	0,071	0,042	0,077	0,055	0,109	0,105	0,218	0,074	0,054	0,058	0,058	0,096	0,076	0,098	0,075	0,049	0,187	0,048	0,058	0,118	0,065
For	0,060	0,119	0,089	0,121	0,086	0,214	0,109	0,127	0,075	0,126	0,215	0,184	0,144	0,080	0,063	0,119		0,145	0,126	0,072	0,065	0,109	0,158	0,047	0,073	0,076	0,239	0,126	0,180	0,198	0,129	0,172	0,090	0,124	0,130	0,226	0,110	0,150	0,148	0,134	0,138	0,150	0,120	0,036	0,046	0,126	0,108	0,102	0,130	0,040
Hor	0,177	0,057	0,127	0,097	0,117	0,143	0,064	0,078	0,100	0,144	0,229	0,053	0,083	0,178	0,094	0,059	0,139		0,053	0,152	0,086	0,123	0,097	0,170	0,160	0,072	0,244	0,134	0,089	0,087	0,031	0,110	0,050	0,114	0,090	0,225	0,038	0,078	0,038	0,097	0,092	0,076	0,100	0,092	0,088	0,122	0,068	0,103	0,096	0,075
Inf	0,147	0,037	0,105	0,032	0,116	0,143	0,062	0,033	0,090	0,113	0,188	0,032	0,074	0,141	0,059	0,085	0,115	0,053		0,161	0,053	0,095	0,079	0,111	0,154	0,041	0,247	0,084	0,048	0,071	0,063	0,069	0,038	0,062	0,060	0,204	0,021	0,046	0,042	0,081	0,017	0,058	0,044	0,061	0,061	0,090	0,049	0,071	0,051	0,069
Iso	0,119	0,166	0,173	0,201	0,185	0,259	0,163	0,179	0,122	0,175	0,285	0,236	0,184	0,159	0,074	0,179	0,071	0,156	0,169		0,118	0,136	0,182	0,158	0,113	0,111	0,279	0,174	0,216	0,208	0,137	0,184	0,125	0,157	0,179	0,261	0,142	0,185	0,170	0,184	0,164	0,196	0,169	0,107	0,116	0,141	0,163	0,197	0,167	0,087
Jos	0,124	0,074	0,022	0,086	0,083	0,153	0,061	0,079	0,082	0,092	0,142	0,091	0,069	0,070	0,047	0,099	0,057	0,091	0,043	0,128		0,080	0,119	0,057	0,081	0,051	0,197	0,090	0,100	0,134	0,049	0,123	0,041	0,065	0,099	0,127	0,047	0,089	0,041	0,092	0,073	0,062	0,055	0,021	0,029	0,087	0,058	0,078	0,038	0,050
Lag	0,169	0,075	0,171	0,147	0,107	0,205	0,067	0,140	0,076	0,048	0,204	0,121	0,159	0,150	0,084	0,163	0,103	0,115	0,085	0,136	0,076		0,166	0,157	0,110	0,083	0,281	0,098	0,149	0,209	0,119	0,184	0,083	0,100	0,105	0,263	0,061	0,158	0,134	0,159	0,091	0,170	0,085	0,098	0,080	0,066	0,138	0,153	0,145	0,054
Lan	0,200	0,085	0,161	0,116	0,130	0,150	0,075	0,070	0,144	0,189	0,193	0,097	0,073	0,146	0,074	0,081	0,146	0,102	0,079	0,191	0,113	0,160		0,149	0,151	0,096	0,310	0,156	0,073	0,057	0,070	0,019	0,094	0,090	0,139	0,244	0,090	0,103	0,081	0,096	0,048	0,110	0,069	0,095	0,074	0,176	0,100	0,136	0,115	0,112
Lle	0,087	0,124	0,075	0,107	0,150	0,219	0,122	0,124	0,103	0,194	0,185	0,195	0,126	0,110	0,088	0,127	0,043	0,185	0,116	0,172	0,059	0,163	0,152		0,119	0,088	0,202	0,151	0,131	0,162	0,113	0,151	0,095	0,094	0,119	0,214	0,102	0,122	0,099	0,122	0,119	0,092	0,063	0,026	0,049	0,168	0,078	0,092	0,120	0,087
Lu	0,156	0,115	0,127	0,175	0,089	0,235	0,139	0,149	0,075	0,131	0,255	0,179	0,149	0,076	0,071	0,157	0,070	0,156	0,149	0,116	0,069	0,106	0,149	0,124		0,079	0,271	0,187	0,176	0,206	0,102	0,177	0,099	0,116	0,171	0,251	0,095	0,169	0,139	0,172	0,100	0,173	0,103	0,079	0,077	0,133	0,125	0,140	0,145	0,044
Mar	0,134	0,029	0,091	0,079	0,086	0,183	0,064	0,058	0,046	0,132	0,195	0,071	0,065	0,077	0,001	0,093	0,071	0,074	0,042	0,112	0,050	0,079	0,091	0,094	0,070		0,180	0,091	0,082	0,080	0,063	0,086	0,036	0,044	0,087	0,198	0,037	0,051	0,048	0,088	0,071	0,064	0,044	0,054	0,021	0,091	0,053	0,081	0,072	0,038
Mon	0,181	0,189	0,210	0,252	0,329	0,331	0,285	0,257	0,186	0,339	0,273	0,287	0,216	0,287	0,184	0,281	0,242	0,249	0,247	0,292	0,206	0,292	0,322	0,207	0,281	0,183		0,324	0,266	0,299	0,249	0,289	0,227	0,233	0,270	0,374	0,222	0,232	0,221	0,225	0,278	0,162	0,239	0,172	0,203	0,270	0,190	0,256	0,234	0,239
Mrt	0,219	0,116	0,154	0,140	0,109	0,250	0,057	0,144	0,173	0,170	0,258	0,142	0,142	0,160	0,109	0,148	0,125	0,135	0,084	0,192	0,084	0,086	0,163	0,161	0,191	0,101	0,342		0,149	0,158	0,133	0,170	0,068	0,085	0,153	0,279	0,081	0,129	0,137	0,151	0,099	0,127	0,081	0,106	0,114	0,147	0,165	0,169	0,151	0,094
Mst	0,212	0,055	0,178	0,095	0,120	0,166	0,066	0,041	0,155	0,203	0,204	0,068	0,060	0,111	0,107	0,105	0,180	0,097	0,058	0,234	0,108	0,158	0,068	0,146	0,181	0,094	0,283	0,163		0,058	0,065	0,060	0,080	0,080	0,082	0,248	0,066	0,085	0,074	0,115	0,062	0,080	0,092	0,090	0,080	0,142	0,078	0,121	0,084	0,130
Nue	0,248	0,070	0,170	0,102	0,147	0,147	0,093	0,063	0,161	0,247	0,269	0,084	0,073	0,160	0,118	0,067	0,184	0,090	0,068	0,212	0,125	0,200	0,056	0,163	0,198	0,075	0,309	0,162	0,052		0,092	0,085	0,087	0,113	0,135	0,256	0,067	0,052	0,073	0,080	0,090	0,075	0,088	0,124	0,103	0,197	0,090	0,125	0,126	0,141
Pal	0,154	0,043	0,086	0,090	0,108	0,180	0,056	0,077	0,101	0,143	0,182	0,077	0,031	0,120	0,064	0,037	0,121	0,032	0,067	0,143	0,043	0,112	0,068	0,117	0,096	0,060	0,259	0,137	0,072	0,087		0,089	0,019	0,075	0,085	0,215	0,027	0,074	0,017	0,129	0,061	0,051	0,066	0,061	0,046	0,104	0,064	0,111	0,083	0,060
Pan	0,183	0,103	0,183	0,108	0,173	0,180	0,144	0,083	0,149	0,264	0,206	0,139	0,101	0,153	0,106	0,093	0,171	0,136	0,091	0,204	0,138	0,193	0,015	0,171	0,186	0,098	0,311	0,198	0,081	0,103	0,104		0,100	0,068	0,127	0,224	0,086	0,087	0,085	0,091	0,068	0,101	0,080	0,109	0,068	0,193	0,084	0,128	0,072	0,115
Par	0,149	0,040	0,091	0,093	0,073	0,167	0,042	0,099	0,093	0,109	0,161	0,086	0,043	0,111	0,030	0,058	0,087	0,050	0,039	0,133	0,038	0,081	0,091	0,103	0,096	0,038	0,239	0,077	0,087	0,083	0,018	0,118		0,072	0,067	0,176	0,019	0,055	0,042	0,103	0,040	0,051	0,055	0,049	0,038	0,090	0,074	0,076	0,101	0,041
Per	0,164	0,051	0,104	0,080	0,095	0,191	0,066	0,047	0,116	0,172	0,178	0,069	0,062	0,077	0,066	0,111	0,122	0,117	0,062	0,169	0,069	0,096	0,083	0,105	0,110	0,044	0,239	0,090	0,092	0,104	0,070	0,076	0,075		0,098	0,177	0,054	0,078	0,066	0,104	0,057	0,090	0,041	0,075	0,039	0,119	0,081	0,108	0,078	0,068
Pom	0,194	0,075	0,166	0,098	0,111	0,216	0,123	0,095	0,110	0,152	0,196	0,107	0,142	0,117	0,065	0,112	0,125	0,094	0,053	0,188	0,096	0,102	0,136	0,123	0,171	0,088	0,282	0,158	0,101	0,124	0,086	0,135	0,070	0,100		0,235	0,083	0,116	0,091	0,114	0,086	0,127	0,098	0,079	0,064	0,120	0,067	0,064	0,124	0,093
Por	0,276	0,228	0,180	0,210	0,221	0,255	0,228	0,218	0,251	0,260	0,314	0,245	0,191	0,231	0,252	0,229	0,229	0,239	0,200	0,278	0,125	0,267	0,236	0,221	0,255	0,198	0,386	0,295	0,260	0,249	0,216	0,226	0,180	0,178	0,235		0,221	0,248	0,171	0,171	0,229	0,217	0,209	0,163	0,153	0,216	0,183	0,180	0,155	0,190
Pra	0,161	0,030	0,109	0,065	0,088	0,170	0,039	0,069	0,081	0,105	0,181	0,056	0,068	0,140	0,047	0,078	0,105	0,032	0,016	0,148	0,045	0,057	0,088	0,113	0,093	0,041	0,233	0,079	0,080	0,060	0,026	0,103	0,018	0,061	0,088	0,227		0,037	0,041	0,095	0,032	0,043	0,039	0,056	0,054	0,073	0,039	0,105	0,069	0,042
Rip	0,190	0,085	0,160	0,068	0,164	0,214	0,110	0,084	0,123	0,242	0,250	0,085	0,048	0,162	0,081	0,059	0,151	0,092	0,055	0,202	0,095	0,167	0,108	0,134	0,171	0,054	0,235	0,155	0,096	0,051	0,083	0,110	0,069	0,083	0,122	0,253	0,048		0,069	0,099	0,093	0,043	0,068	0,079	0,069	0,188	0,045	0,075	0,094	0,106
Sal	0,159	0,043	0,055	0,068	0,146	0,154	0,062	0,044	0,112	0,169	0,168	0,053	0,062	0,146	0,077	0,068	0,142	0,040	0,044	0,175	0,038	0,131	0,080	0,102	0,134	0,046	0,223	0,145	0,083	0,068	0,011	0,099	0,039	0,060	0,095	0,177	0,039	0,079		0,070	0,081	0,045	0,046	0,066	0,051	0,127	0,045	0,092	0,063	0,078
Sop	0,152	0,067	0,127	0,080	0,156	0,099	0,080	0,074	0,109	0,193	0,235	0,086	0,094	0,162	0,126	0,066	0,133	0,101	0,079	0,197	0,102	0,158	0,095	0,133	0,173	0,091	0,236	0,162	0,115	0,076	0,129	0,107	0,103	0,103	0,120	0,176	0,097	0,102	0,074		0,118	0,094	0,095	0,104	0,089	0,169	0,062	0,080	0,117	0,094
Sor	0,171	0,051	0,125	0,073	0,086	0,197	0,048	0,075	0,102	0,113	0,188	0,071	0,071	0,125	0,073	0,105	0,133	0,091	0,010	0,178	0,064	0,086	0,045	0,134	0,103	0,073	0,286	0,099	0,068	0,087	0,063	0,079	0,040	0,055	0,084	0,231	0,025	0,104	0,079	0,117		0,090	0,038	0,072	0,077	0,101	0,093	0,114	0,093	0,063
Sue	0,166	0,071	0,083	0,073	0,155	0,194	0,087	0,092	0,131	0,208	0,197	0,098	0,027	0,168	0,098	0,085	0,154	0,079	0,057	0,220	0,065	0,175	0,105	0,107	0,180	0,072	0,173	0,144	0,078	0,066	0,059	0,115	0,057	0,095	0,138	0,223	0,045	0,045	0,042	0,087	0,084		0,076	0,072	0,079	0,149	0,057	0,114	0,081	0,109
Thu	0,170	0,053	0,126	0,092	0,127	0,184	0,052	0,095	0,088	0,146	0,183	0,091	0,081	0,131	0,034	0,106	0,114	0,103	0,046	0,181	0,056	0,080	0,065	0,064	0,100	0,047	0,243	0,085	0,103	0,085	0,064	0,092	0,055	0,044	0,101	0,210	0,038	0,076	0,044	0,099	0,036	0,074		0,065	0,067	0,138	0,073	0,100	0,096	0,070
Tos	0,090	0,080	0,055	0,071	0,092	0,164	0,070	0,083	0,047	0,114	0,152	0,106	0,075	0,083	0,029	0,078	0,026	0,093	0,059	0,107	0,018	0,096	0,090	0,032	0,071	0,050	0,175	0,113	0,099	0,118	0,059	0,118	0,051	0,080	0,081	0,165	0,058	0,081	0,063	0,102	0,072	0,080	0,067		0,023	0,108	0,048	0,053	0,081	0,032
Ult	0,119	0,054	0,065	0,076	0,059	0,175	0,065	0,061	0,059	0,125	0,134	0,089	0,059	0,061	0,031	0,056	0,047	0,098	0,063	0,127	0,032	0,079	0,066	0,059	0,074	0,021	0,217	0,129	0,092	0,098	0,045	0,071	0,038	0,039	0,066	0,150	0,061	0,075	0,052	0,093	0,080	0,090	0,068	0,024		0,095	0,030	0,041	0,059	0,026
Una	0,178	0,080	0,159	0,153	0,124	0,223	0,102	0,144	0,123	0,030	0,209	0,137	0,143	0,160	0,114	0,201	0,126	0,122	0,084	0,148	0,087	0,064	0,179	0,180	0,136	0,094	0,282	0,147	0,157	0,198	0,110	0,217	0,092	0,124	0,126	0,230	0,080	0,203	0,132	0,175	0,102	0,159	0,143	0,111	0,101		0,133	0,189	0,139	0,084
Uss	0,147	0,062	0,101	0,058	0,130	0,182	0,113	0,055	0,076	0,178	0,204	0,081	0,080	0,132	0,079	0,054	0,111	0,074	0,042	0,182	0,062	0,144	0,097	0,087	0,132	0,056	0,205	0,177	0,086	0,077	0,064	0,101	0,079	0,082	0,062	0,187	0,044	0,044	0,048	0,070	0,097	0,062	0,078	0,050	0,032	0,141		0,040	0,049	0,050
Val	0,177	0,107	0,117	0,075	0,117	0,149	0,138	0,086	0,091	0,202	0,242	0,095	0,116	0,125	0,103	0,062	0,102	0,103	0,062	0,208	0,074	0,153	0,130	0,100	0,144	0,078	0,265	0,177	0,126	0,110	0,107	0,133	0,077	0,103	0,064	0,178	0,104	0,074	0,090	0,078	0,114	0,113	0,099	0,049	0,038	0,195	0,038		0,108	0,077
Vie	0,143	0,110	0,099	0,083	0,123	0,168	0,121	0,041	0,118	0,179	0,193	0,111	0,102	0,129	0,097	0,125	0,123	0,096	0,048	0,174	0,041	0,138	0,110	0,100	0,140	0,071	0,229	0,162	0,096	0,120	0,082	0,076	0,106	0,078	0,125	0,158	0,071	0,095	0,060	0,117	0,086	0,084	0,101	0,082	0,063	0,145	0,056	0,103		0,095
Vil	0,112	0,053	0,077	0,087	0,091	0,175	0,058	0,089	0,047	0,104	0,216	0,107	0,097	0,112	0,036	0,069	0,037	0,071	0,064	0,087	0,045	0,053	0,109	0,096	0,043	0,032	0,246	0,096	0,139	0,130	0,054	0,123	0,040	0,065	0,093	0,193	0,038	0,108	0,076	0,097	0,067	0,116	0,070	0,029	0,026	0,085	0,056	0,081	0,092	



Appendix 4: Estimation of the most probable number of genetic clusters

Chapitre 2

Locally asymmetric introgressions between subspecies suggest circular range expansion at the *Antirrhinum majus* global scale

JOURNAL OF Evolutionary Biology

doi: 10.1111/j.1420-9101.2011.02276.x

Locally asymmetric introgressions between subspecies suggest circular range expansion at the *Antirrhinum majus* global scale

A. KHIMOUN*†, M. BURRUS*†, C. ANDALO*†, Z.-L. LIU‡, C. VICÉDO-CAZETTES*†, C. THÉBAUD*† & B. PUJOL*†

*Université de Toulouse; UPS; EDB (Laboratoire Évolution et Diversité Biologique); UMR5174; Toulouse Cedex 9, France
†CNRS; EDB (Laboratoire Évolution et Diversité Biologique); UMR5174; Toulouse, France
‡Key Laboratory of Resource Biology and Biotechnology in Western China (Northwest University), Ministry of Education, Xi'an, China

Keywords:

Antirrhinum majus;
chloroplast DNA;
contact zone;
genetic introgression;
haplotype sharing.

Abstract

Assessing processes of geographic expansion in contact zones is a crucial step towards an accurate prediction of the evolution of species genetic diversity. The geographic distribution of cytonuclear discordance often reflects genetic introgression patterns across a species geographic range. *Antirrhinum majus pseudomajus* and *A. m. striatum* are two interfertile subspecies that occupy nonoverlapping areas but enter in contact in many locations at the margin of their geographic distribution. We found that genetic introgression between both subspecies was asymmetric at the local scale and geographically oriented in opposite directions at both ends of their contact zone perimeter in the Pyrenees. Our results suggest that the geographic expansion of *A. majus* subspecies was circular around the perimeter of their contact zone and pinpoint the need to integrate different spatial scales to unravel complex patterns of species geographic expansion.

Introduction

Species range expansion is a key mechanism that shapes the genetic diversity of species (Hewitt, 2000; Excoffier *et al.*, 2009) and modifies their evolutionary potential (Lavergne & Molofsky, 2007; Pujol & Pannell, 2008; Pujol *et al.*, 2009). Range expansion frequently leads to the formation of contact zones between populations of differentiated species (Anderson, 1949; Grant, 1971; Arnold, 1997; Barton, 2001). In such cases, genes from a foreign species might replace genes of, and therefore introgress, the gene pool of the native species (Potts & Reid, 1988, 1990; Schemske & Morgan, 1990). Genetic introgression (i.e. the transfer of genetic material in the genome of another species) is often rendered possible by fertile hybrids that occupy the contact zones and act as 'bridges to gene flow', therefore allowing gene exchange between species to occur (Broyles, 2002). Recurrent pollen exchanges between species and recurrent backcrossing between hybrids and parental species are then likely to generate large geographic areas of genetic introgression inside and outside contact zones (Campbell *et al.*, 1998; Leebens-Mack & Milligan, 1998). The characterization of genetic introgression is a key step towards a better understanding of the expansion dynamics of species in contact and of the evolution of their diversity (Currat *et al.*, 2008; Excoffier *et al.*, 2009).

Recent work on *Antirrhinum majus* showed that floral trait segregation, in combination with pollinator behaviour, can explain, at least partly, the maintenance of flower colour polymorphism in one particularly narrow hybrid zone between *A. majus pseudomajus* and *A. m. striatum* subspecies (Whibley *et al.*, 2006; Tastard *et al.*, 2008). It is, however, currently unknown whether contact zones and gene exchanges between *A. m. pseudomajus* and *A. m. striatum* are widespread across the geographic range of both these subspecies. The general aim of our study is to understand the biogeography of these two interfertile subspecies. We expect that contact might be frequent across the species range if one subspecies progressively expands its range into the range

Correspondence: Aurélie Khimoun, Université de Toulouse; UPS; EDB (Laboratoire Évolution et Diversité Biologique); UMR5174; 118 route de Narbonne, F-31062 Toulouse Cedex 9, France.
Tel.: 00 33 05 61 55 74 86; fax: 00 33 05 61 55 73 27;
e-mail: khimoun@cict.fr

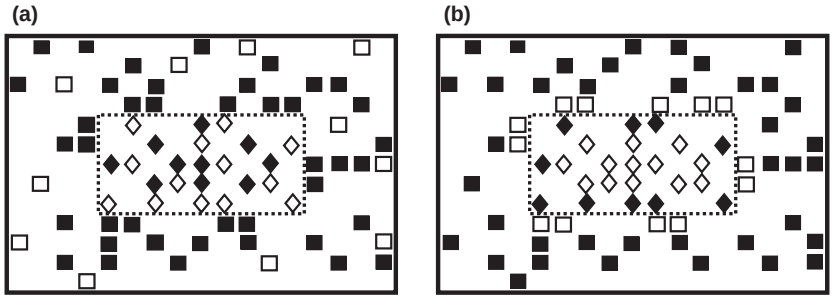


Fig. 1 Expected patterns of chloroplast sharing between species. Black and white colours represent Species 1 and Species 2, respectively. The contact zone perimeter between both species is symbolized by a dotted line. Squares represent the chloroplast haplotype 1 most often associated with Species 1, and diamonds represent the chloroplast haplotype 2 most often associated with Species 2. Under the hypothesis that black squares and white diamonds reflect the ancestral state, then the two less frequent associations (i.e. black diamonds and white squares) are called 'discordant'. (a) Random geographic distribution of discordant associations within each species range expected that results from the retention of ancestral polymorphism or convergence. (b) Geographic structure of discordant associations found only around the contact zone perimeter that results from local introgression between both species where genes can be exchanged.

occupied formerly by the other subspecies because the boundary between *A. majus* subspecies is not linear. *A. m. pseudomajus* is distributed around the range of *A. m. striatum*. Ultimately, moving boundaries sometimes result in the local replacement of the invaded species by the species pushing off the contact zone on its front of colonization (Buggs & Pannell, 2007; Pannell & Pujol, 2009). Alternatively, stable boundaries between both taxa can be maintained at equilibrium between migration and selection (Barton & Hewitt, 1989; Bull, 1991). When the contact zone between two species ranges is not linear, one could expect geographically complex expansion patterns and/or reciprocal gene exchanges to occur and result in multiple sites of genetic introgression. The detection of such widespread pattern is important because it might result in the long term in the genetic admixture of their formerly differentiated genomes. Evidence to determine whether taxa replacement, maintenance of species boundaries or admixture are the most likely evolutionary outcomes can be provided by the analysis of the geographic distribution of relict uniparentally inherited DNA (i.e. mitochondrial or chloroplastic DNA) where the nuclear genome is being replaced (Potts & Reid, 1988, 1990; Schemske & Morgan, 1990).

To establish whether parapatric boundaries between *A. majus* subspecies are moving following a complex geographic expansion pattern, we studied the geographic distribution of the association between chloroplast haplotypes (maternally inherited) and a nuclear gene (biparentally inherited) regulating the main taxonomic criterion, which is the magenta flower colour for *A. m. pseudomajus* and the yellow flower colour for *A. m. striatum* (see the study system section for details on the taxonomy of the species; Rothmaler, 1956 and Sutton, 1988) over the range of the species. We then searched for evidence of genetic introgression between *A. m. pseudomajus* and *A. m. striatum.* Although chloroplast haplotype sharing across taxa boundaries is often the outcome of genetic introgression (Rieseberg & Soltis, 1991; Wendel & Doyle, 1998; Linder & Rieseberg, 2004), caution must be taken when interpreting patterns of haplotype sharing because convergence or incomplete sorting of ancestral polymorphism might generate similar patterns (Muir & Schlotterer, 2005; Lexer *et al.*, 2006). In cases of retention of ancestral polymorphism or convergence, we would expect cytonuclear associations to be randomly distributed in a mosaic pattern over the species geographic range (Fig. 1a). In contrast, if chloroplast sharing between both subspecies is the result of introgression, we would expect discordant cytonuclear associations to be located close to the perimeter zone formed by the contact between subspecies (Fig. 1b). Geographic sectors characterized by the high frequency of one cytonuclear association are also expected if heterogeneous selection spatially structured *A. majus* ancestral polymorphism. In this article, we confront those hypotheses to establish the most likely scenario of evolutionary history that can explain the observed geographic distribution of cytonuclear associations in *A. majus*. Our investigation of those scenarios was rendered possible by the broad scale at which our study was conducted, i.e. the species geographic range, which allowed us to uncover the geographic direction of genetic introgression between both subspecies around their geographic boundaries.

Materials and methods

Study system

Antirrhinum majus (Scrophulariaceae) is a herbaceous short-lived perennial plant characterized by a patchy distribution in southern Europe. Its geographic distribution is centred over the Pyrenees, between north-eastern Spain and south-western France. The two subspecies *A. m. striatum* and *A. m. pseudomajus* occupy largely parapatric geographic regions. The geographic area

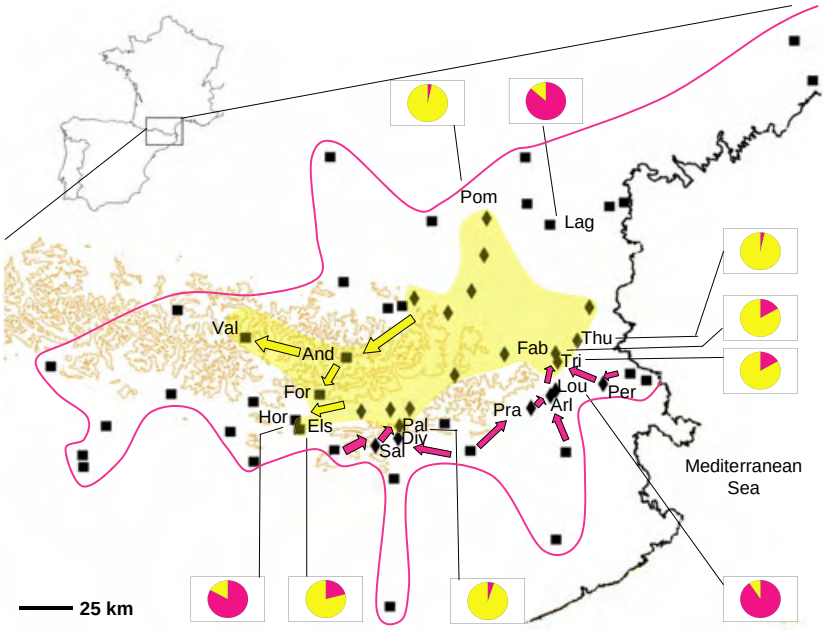


Fig. 2 Geographic distribution of *A. majus* cytonuclear associations. Magenta and yellow layers represent, respectively, *Antirrhinum majus pseudomajus* and *A. m. striatum* geographic ranges. Within each subspecies, black symbols represent populations sampled for this study. Squares and diamonds represent populations characterized by Haplotype I and Haplotype II, respectively. Pie charts are only presented for populations that are polymorphic at the *ROS1* locus. The magenta and the yellow proportion of the pie charts represent the respective frequencies of *ROS1-M* and *ROS1-Y* alleles in the population. Magenta and yellow arrows indicate the hypothetical scenario of range expansion followed by *A. m. pseudomajus* and *A. m. striatum* that is supported by our data. Brown lines represent elevation isoclines above 1800 m.

occupied by *A. m. striatum* is surrounded by the geographic area occupied by *A. m. pseudomajus* (Fig. 2). Taxonomic determination of *A. majus* subspecies is mostly based on the colour of flower corolla. *A. m. pseudomajus* is characterized by magenta flowers. It is referred to interchangeably in the literature as *A. m.* ssp. *majus* and *A. m.* ssp. *linkianum*. Some authors include the ssp. *cirrhigerum* as a variety of *A. m.* ssp. *linkianum*. *A. m. striatum* is characterized by yellow flowers. It is referred to interchangeably in the literature as *A. latifolium* ssp. *striatum*, *A. huetii* and *A. braun-blanquetii* (Rothmaler, 1956; Sutton, 1988). It is important to note that *A. m. pseudomajus* and *A. m. striatum* are interfertile and share pollinators (Whibley *et al.*, 2006; Andalo *et al.*, 2010).

Plant material sampling strategy

A total of 685 plants were sampled from 2002 to 2007 in 55 allopatric or parapatric populations distributed over the geographic range of the species. Geographic coordinates of populations were recorded by using a GPS device (Garmin, Olathe, KS, USA). A numerical scoring system was used to rank magenta and yellow flower colour phenotypes visually, following methods developed by Whibley *et al.* (2006). Obviously, plants that displayed yellow flowers were classified as *A. m. striatum* whereas plants that displayed magenta flowers were classified as *A. m. pseudomajus*. Population characteristics are summarized in Table S1a and S1b. For each individual, young leaves and shoot tips were collected and stored at −20 °C until DNA was extracted by using the DNeasy Plant Mini kit (Qiagen, Hilden, Germany).

Molecular analyses

ROSEA genotyping

The *ROSEA* locus is made of 2 MYB – myeloblastosis – regulatory genes controlling floral pigmentation intensity, out of which *ROS1* has the main role in flower colour variation (Schwinn *et al.*, 2006). *ROS1* sequences can be grouped in three main haplotypes *ROS1*-Ma, *ROS1*-Mb and *ROS1*-Y (Whibley, 2004). *ROS1*-Ma and *ROS1*-Mb haplotypes are diagnostic of *A. m. pseudomajus* and are grouped under the name of *ROS1*-M whereas the *ROS1*-Y haplotype is diagnostic of *A. m. striatum* (Whibley, 2004). *ROS1* genotypic data were available for the 14 populations ($n = 166$ plants) that were previously examined by Whibley *et al.* (2006). We obtained *ROS1* genotypic data for the remaining 41 populations ($n = 519$ plants) using the RG4/RR21, RG6/RR21 and RG1/RR21 primers in a single PCR, following the protocol established by Whibley (2004).

PCR–RFLP analysis of chloroplast DNA

Maternal lineages were determined in the 55 populations ($n = 685$ plants) by genotyping the 1.6-kb *psbC* [*psII*

44-kDa protein] – *trnS* [*tRNA-Ser*(*UGA*)] intergenic region, using the CS universal primers (Demesure *et al.*, 1995). Sequencing of this chloroplast region revealed two haplotypes that differed at two SNP loci, one of which was included in a *MseI* restriction site. We therefore obtained two different haplotypes after digestion of the *psbC-trnS* fragment by the *Mse I* enzyme. Haplotype I was characterized by eight *Mse I* restriction sites that generated a nine-band profile on agarose gel. Haplotype II was characterized by a 10-band profile. The PCR amplification protocol is presented in the supplementary online material.

Data analyses

To examine cytonuclear associations, we calculated *ROS1* allelic frequencies and chloroplast haplotype frequencies within each population and mapped these frequencies using ArcGis (ESRI, Redlands, CA, USA) software. To determine whether subspecific patterns of chloroplast haplotype sharing were the result of evolutionary convergence, incomplete sorting of ancestral polymorphism or introgression, we assessed the role of the geographic distance between populations of different subspecies on the geographic distribution of chloroplast haplotypes. To do so, we calculated the Euclidian geographic distance to the closest population of the other subspecies for every population within each subspecies. We then tested whether this geographic distance differed between populations that share chloroplast haplotypes with the other subspecies and populations that do not share chloroplast haplotypes with the other subspecies. We performed a two-sample *t*-test built on the basis of 2000 permutations (Good, 2000) in R (R Development Core Team, 2007, Vienna, Austria).

Results

Relationship between flower colour and *ROS1* genotype

At the population level, 67% of the 55 populations were assigned to *A. m. pseudomajus* and 33% to *A. m. striatum* on the basis of their flower colour phenotype (n = 37 *A. m. pseudomajus* populations and n = 18 *A. m. striatum* populations). Forty-six populations out of 55 were monomorphic at the *ROS1* locus. All plants in those populations presented the same homozygote genotype at the *ROS1* locus, being either *ROS1-M/ROS1-M* in 34 *A. m. pseudomajus* populations or *ROS1-Y/ROS1-Y* in 12 *A. m. striatum* populations (Table 1). Six *A. m. striatum* populations (*Els, Fab, Pal, Pom, Thu* and *Tri*) and three *A. m. pseudomajus* populations (*Hor, Lag* and *Lou*) were polymorphic at the *ROS1* locus (allelic frequencies are presented on Fig. 2).

Over all individuals, plants that displayed yellow flowers were characterized by the genotypes *ROS1-Y/ROS1-Y, ROS1-Y/ROS1-M* or *ROS1-M/ROS1-M* at the respective frequencies of 94%, 4.5% and 1.5%. Plants that displayed magenta flowers were characterized by the genotypes *ROS1-Y/ROS1-Y, ROS1-Y/ROS1-M* or *ROS1-M/ROS1-M* at respective frequencies of 0.8%, 2.2% and 97%. Such correlation between *ROS1-Y* and the yellow colour and between *ROS1-M* and the magenta colour is in agreement with the previous study conducted by Whibley *et al.* (2006).

Distribution of chloroplast DNA genotypes

Each particular population was characterized by a unique *psbC-trnS* chloroplast haplotype. The geographic distribution of population haplotypes was nonoverlapping across the geographic range of the species (Fig. 2). Haplotype I was found in 79% of *A. m. pseudomajus* populations and in 20% of *A. m. striatum* populations. Haplotype II was found in the remaining 21% of *A. m. pseudomajus* populations and 80% of *A. m. striatum* populations (Table 1).

Among *A. m. pseudomajus* populations, the chloroplast haplotype depended significantly on whether populations were located closely to *A. m. striatum* populations. Most of the *A. m pseudomajus* populations that were characterized by Haplotype I were distant from *A. m. striatum* populations (Fig. 2). In contrast, *A. m. pseudomajus* characterized by Haplotype II could only be found in populations located closely to the contact zone perimeter. The mean distance between *A. m. pseudomajus* popula-

Table 1 Cytonuclear associations.

Chloroplast haplotype	Population subspecies	ROS-1 genotypes
Haplotype I (35)	*Antirrhinum majus pseudomajus* (31)	**ROS1-M/ROS1-M** (29)
		ROS1-M/ROS1-M; ROS1-M/ROS1-Y; ROS1-Y/ROS1-Y (2)
	A. m. striatum (4)	**ROS1-Y/ROS1-Y** (3)
		ROS1-M/ROS1-Y; **ROS1-Y/ROS1-Y** (1)
Haplotype II (20)	*A. m. pseudomajus* (6)	**ROS1-M/ROS1-M** (5)
		ROS1-M/ROS1M; ROS1-M/ROS1-Y (1)
	A. m. striatum (14)	**ROS1-Y/ROS1-Y** (9)
		ROS1-M/ROS1-M; ROS1-M/ROS1-Y; **ROS1-Y/ROS1-Y** (5)

*Most frequent genotype in bold. The number of populations is indicated between parentheses.

tions characterized by Haplotype I and the closest *A. m. striatum* population (mean distance ± SD = 40.3 ± 30.1 km) was significantly larger (using a permutation *t*-test $P < 0.05$) than the mean distance between *A. m. pseudomajus* populations characterized by Haplotype II and the closest *A. m. striatum* population (mean distance ± SD = 12.9 ± 5.7 km). Within the *A. m. striatum* geographic range, no correlation between the occurrence of a chloroplast haplotype and the distance to the nearest *A. m. pseudomajus* population was detected. The mean distance between *A. m. striatum* populations characterized by Haplotype II and the closest *A. m. pseudomajus* population (mean distance ± SD = 16.6 ± 7.0 km) was not significantly different (using a permutation *t*-test $P > 0.05$) than the mean distance between *A. m. striatum* populations characterized by Haplotype I and the closest *A. m. pseudomajus* population (mean distance ± SD = 18.3 ± 11.1 km). It is important to note that the few *A. m. striatum* populations characterized by chloroplast Haplotype I ($n = 4$) were all grouped on the west border of *A. m. striatum* geographic distribution (see Fig. 2). It is also important to note that such analysis in *A. m. striatum* was limited by the small number of *A. m. striatum* populations that are located far from the contact zone perimeter, which is a direct consequence of the narrower geographic area occupied by *A. m. striatum*. Furthermore, *A. m. pseudomajus* populations characterized by Haplotype II were located in the east of the contact zone perimeter whereas *A. m. striatum* populations characterized by Haplotype I were located in the west of the contact zone perimeter (Fig. 2). The direction of the geographic gradient formed by *ROS1* allele frequencies in the east was different from the one in the west of the contact zone.

Cytonuclear association

Because of the correlation between *ROS1* and flower colour, the overall pattern of cytonuclear association was very similar to the pattern presented above. Most of the *A. majus* populations were characterized either by the association of chloroplast Haplotype I and the *ROS1*-M allele or by the association of chloroplast Haplotype II and the *ROS1*-Y allele. Among populations characterized by Haplotype I, most of them (83%) were also characterized by the fixation of the *ROS1*-M allele whereas the remaining populations were characterized either by polymorphism at the *ROS1* locus (8.5%) or by the fixation of the *ROS1*-Y (8.5%). Among populations characterized by Haplotype II, only 45% were characterized by the fixation of the *ROS1*-Y allele whereas the other populations were characterized either by polymorphism at the *ROS1* locus (30%) or by the fixation of the *ROS1*-M allele (25%) (Table 1).

In most of the *A. m. pseudomajus* populations ($n = 29$ out of 37), all individuals were characterized by the cytonuclear association of chloroplast Haplotype I and *ROS1*-M. This includes all the *A. m. pseudomajus* populations that were distant from the contact zone perimeter (Fig. 2). The most frequent cytonuclear association that characterized *A. m. striatum* populations was found in 50% of *A. m. striatum* populations ($n = 9$). In those populations, all individuals were characterized by the same cytonuclear association (chloroplast Haplotype II and *ROS1*-Y). Around the contact zone perimeter (see Fig. 2), we found five *A. m. pseudomajus* populations (*Arl*, *Div*, *Per*, *Pra* and *Sal*) where all individuals were characterized by the cytonuclear association of chloroplast Haplotype II and *ROS1*-M. Those populations were located at the eastern side of the contact zone perimeter (Fig. 2). Around the contact zone perimeter, we also found three *A. m. striatum* populations (*And*, *For* and *Val*) where all individuals were characterized by the association of the chloroplast Haplotype I and *ROS1*-Y. Those populations were located at the western side of the contact zone perimeter (Fig. 2).

In two populations of the three *A. m. pseudomajus* populations that were polymorphic at the *ROS1* locus, Haplotype I was associated with a high frequency of *ROS1*-M. Similarly, in five populations of the six *A. m. striatum* populations that were polymorphic at the *ROS1* locus, Haplotype II was associated with a high frequency of *ROS1*-Y alleles (Table 1). Interestingly, such populations at an intermediary stage of genetic introgression were always very close to the contact zone perimeter (Fig. 2).

Discussion

Heterogeneous selection of ancestral polymorphism vs. genetic introgression

One hypothesis explaining that chloroplast haplotypes are shared between *A. m. pseudomajus* and *A. m. striatum* is that such pattern results from the retention of ancestral polymorphism without selection being involved. Under such scenario, we would expect cytonuclear associations to be widespread across the entire range of *A. majus* (Fig. 1). This was however not the case. We found them to be grouped in four discrete geographic areas. We therefore discarded this hypothesis (Fig. 2). Another hypothesis that can be invoked is that local heterogeneous selection is responsible for the geographic distribution of the four cytonuclear associations in four discrete geographic sectors in the absence of interspecific introgression. Under such scenario, natural selection would have differently advantaged four ancestral cytonuclear associations between the chloroplast Haplotypes I and II and *ROS1* alleles in four regions. In populations located between those four regions where cytonuclear associations were fixed, we found populations that were polymorphic for *ROS1* alleles but not for chloroplast haplotypes (Fig. 2). These polymorphic populations formed geographically orientated gradients in *ROS1* allele

frequencies that were all located onto the contact zone perimeter between *A. m. pseudomajus* and *A. m. striatum*. Gradients were found on the east side and on the west side of the contact zone perimeter. Because the contact zone perimeter is where we expect gene exchanges between both subspecies to occur, the geographic distribution of chloroplast haplotypes between subspecies and the gradients of *ROS1* allele frequencies that we found in the contact zone perimeter are more likely reflecting genetic introgression between subspecies than geographically heterogeneous selection on cytonuclear ancestral polymorphism.

Local patterns of genetic introgression reflect a circular range expansion scenario

The geographic distribution of cytonuclear associations suggests that chloroplast Haplotype I was historically associated with *A. m. pseudomajus*. This is because chloroplast Haplotype I was more frequent in *A. m. pseudomajus* populations, especially those that were geographically isolated from *A. m. striatum* populations (i.e. allopatric populations) whereas Haplotype II was only found in *A. m. pseudomajus* populations located in the contact zone perimeter (i.e. parapatric populations). The distribution of chloroplast Haplotypes was less strikingly structured among *A. m. striatum* populations. It would seem nevertheless logical, in regard of Haplotypes I and II distribution in *A. m. pseudomajus*, that Haplotype II was historically associated with *A. m. striatum*. Under the assumption that Haplotype I and Haplotype II were originally associated specifically with *A. m. pseudomajus* and *A. m. striatum*, respectively, the geographic distribution of subspecies, chloroplast haplotypes and nuclear *ROS1* alleles revealed areas of cytonuclear discordance. In those areas, chloroplast haplotypes were not associated with the expected subspecies. This was the case on the east side of the contact zone perimeter for six *A. m. pseudomajus* populations characterized by chloroplast Haplotype II and a high frequency of *ROS1*-M alleles that had often reached fixation. This is probably because *A. m. striatum* plants were previously occupying the sites where those *A. m. pseudomajus* populations are nowadays found. Historically, those six populations were probably displaying yellow flowers and were characterized by matching chloroplast nuclear genotypes, i.e. Haplotype II and *ROS1*-Y. It is plausible that cytonuclear discordance emerged because nuclear genes of foreign populations were dispersed and introgressed the gene pool of local populations. The exact inverse scenario can be observed on the west side of the contact zone perimeter in four populations of *A. m. striatum*, which habitat was probably occupied previously by *A. m. pseudomajus* populations. Such geographic distribution of cytonuclear associations could be interpreted as reflecting asymmetric introgression between subspecies at a local scale, i.e. unidirectional introgression of *ROS1* alleles of one subspecies into the gene pool of the second subspecies. At the broad scale of the species geographic distribution, such directional genetic introgression, however, appeared to be inverted between the east side and the west side of the contact zone perimeter. Because the genotype at the *ROS1* locus determines whether a plant belongs to *A. m. pseudomajus* or to *A. m. striatum*, the spread of *ROS1* alleles reflects the spread of the corresponding subspecies. Our results therefore reflect a progressive shift in the geographic range of both *A. m. pseudomajus* and *A. m. striatum*. Under such scenario, both subspecies expanded and/or still expand their ranges in opposite directions on the east and the west side of the contact zone perimeter, which ultimately results in their global range expansion being articulated around each other into a circular pattern (Fig. 2).

The relative role of selection and dispersal in the spread of *ROS1* alleles

Either selection or dispersal can generate and maintain genetic introgression patterns, such as those detected in our study (Currat *et al.*, 2008). Local selection might explain the local asymmetry in the introgression pattern, even in the presence of bidirectional gene flow. In such case, we would expect cytonuclear discordant associations 'Haplotype II/*ROS1*-M' and 'Haplotype I/*ROS1*-Y' to provide a selective advantage, respectively, on the east side and on the west side of the contact zone perimeter. When patterns of introgression are asymmetric, they might result from intrinsic attributes of species, such as prezygotic asymmetric barriers [e.g. asymmetric pollen-style incompatibilities (Cruzan & Arnold, 1994)], sex-biased dispersal (Petit *et al.*, 2003) or post-zygotic asymmetric barriers [e.g. partial hybrid sterility (Shuker *et al.*, 2005), which are commonly attributed to cytonuclear interactions (Levin, 1971; Tiffin *et al.*, 2001)]. The hypothesis of one subspecies having an intrinsic advantage over the other subspecies when introgressing a foreign gene pool can be discarded because reciprocal patterns of introgressive hybridization between subspecies were detected on the west and the east side of the contact zone perimeter. Our results bring evidence that genes of each subspecies have the potential to introgress the other subspecies. They therefore corroborate the absence of intrinsic post-pollination barriers to reproduction between both subspecies previously found in an experimental study by Andalo *et al.* (2010). Local selection might also be driven by extrinsic factors. Environmental conditions might exert selective pressures on the *ROS1* locus that vary between regions where genetic introgression was found. Such selective pressures might also target nuclear genes that are linked with ROSEA. We acknowledge the limits of our genetic assay based only on the single-locus ROSEA, which is responsible for the taxonomic criterion determining to which subspecies a plant belongs. Investigating more markers would bring

a more complete picture about the extent of genetic introgression between both subspecies and would be informative on the role played by local selection on the spread of *ROS1* alleles. Local asymmetric introgression patterns might also be explained by recurrent unidirectional gene exchanges. Because asymmetric introgression was restricted to specific geographic areas, local environmental barriers to gene flow (valleys, mountains, etc.) might be responsible for local unidirectional gene flow. Our study therefore calls for testing whether specific environmental or physical conditions on each side of the contact zone might exert directional constraints to gene flow. Finally, biotic interactions might also be involved in the spread of *ROS1* alleles at the local scale. Experimental pollination studies brought evidence of a constancy phenomenon in the pollinating behaviour of bumblebees that was driven by *A. majus* flower colour, i.e. pollinators visited preferentially the same morph during a foraging sequence (Jones & Reithel, 2001; Tastard, 2009). Such pollinator behaviour was already shown to affect the evolution of a floral trait coded by a single locus (Jones & Reithel, 2001). In our case, such behaviour might result in positive frequency-dependent selection on flower colour that would ultimately reinforce or accelerate the spread of *ROS1* alleles. Such process would counteract the spread of rare variants in a population and is therefore not expected to be at the origin of the asymmetric introgression of *ROS1* alleles. It might, however, participate to the fixation of a new variant in a population that is submitted to massive unidirectional gene flow from the other subspecies.

Cytonuclear discordance as a result of pollen flow

Genetic introgression patterns such as those detected between *A. m. pseudomajus* and *A. m. striatum* are a common outcome when invading populations can spread their nuclear genes at a long distance by means of pollen flow and seed dispersal is limited (Petit *et al.*, 2003). Such hypothesis is not exclusive because the geographic distribution of cytonuclear associations that we observed could also be explained by demographic expansion through seed dispersal. In such case, the demographic imbalance between invaders and residents would result in the asymmetric introgression of genes from the resident species genome into the invader genome (Currat *et al.*, 2008). Dispersal characteristics of *A. majus*, however, bring support to the first hypothesis, i.e. genetic introgression by pollen flow. Indeed, *A. majus* seeds are very small and light [<15 mg (Andalo *et al.*, 2010)] and can mostly be dispersed at a short distance of the maternal plant by gravity. In contrast, *A. majus* pollen is transported by bumblebees (several *Bombus* species) and carpenter bees (*Xylocopa* sp.) and is therefore likely to migrate across long distances (Whibley, 2004). Indeed, distance covered by carpenter bees of the species *Xylocopa violacea* can reach 1.2 km (Molitor, 1937) whereas bumblebees of the species *Bombus terrestris* can cover up to 2.8 km (Chapman *et al.*, 2003; Darvill *et al.*, 2004). In the light of such dispersal characteristics, the geographic scale at which we observed the signature of genetic introgression reinforces our view that the spread of nuclear genes across subspecies boundaries in *A. majus* was/is progressive. Such progressive spread certainly involved populations, either disappeared or still present, that were separated by close distances suitable for pollinator browsing. Such populations would then play the role of a relay for pollinators and act as directional bridges to gene flow.

Conclusion

Documented examples of species geographic expansion in a contact zone generally imply a unique geographic direction at the scale of the species (Martinsen *et al.*, 2001; Rohwer *et al.*, 2001; Melo-Ferreira *et al.*, 2005). Here, we found that *A. m. pseudomajus* invaded what was previously the habitat of *A. m. striatum* by expanding its range northward on the east side of the contact zone perimeter whereas *A. m. striatum* expanded its range southward within the initial habitat of *A. m. pseudomajus* on the west side of the contact zone perimeter. Both subspecies appear thus to replace each other in a rotation movement at the scale of the species geographic range. Ultimately, this circular mode of geographic expansion might result in the global admixture of both subspecies nuclear genomes. Evolutionary consequences of genetic admixture in *A. majus* might therefore be expected to influence the evolutionary dynamics of the species at a global scale. This system, because it integrates reciprocal gradients of range expansion and genetic admixture in the two subspecies, constitutes a unique opportunity to evaluate their relative impact on the evolutionary potential of a species. It was possible to detect this surprising geographic pattern because we evaluated the geographic distribution of few but spatially structured chloroplastic and nuclear loci in multiple populations from geographically distinct sectors of the whole contact zone perimeter between *A. m. pseudomajus* and *A. m. striatum*. Our study therefore reinforces the current view that direction and speed of hybrid zone displacement can vary across replicates (Hairston *et al.*, 1992; Britch *et al.*, 2001; Buggs & Pannell, 2007). It also pinpoints the need to take into account multiple sites when studying contact zones between species because a broad geographic scope might reveal different patterns than those observed at a local scale. Indeed, focusing on a restricted area of the contact zone might shed light on species geographic range expansion patterns that are not representative of the whole species expansion dynamics.

Acknowledgments

All the work presented in this article was supported by core funding from the French CNRS and the University

of Toulouse Paul Sabatier, France. AK is supported by a PhD grant from the French ministry of research. We thank Nick Barton, Pauline Garnier-Géré, Ferran Palero and the anonymous reviewers for their useful comments on the manuscript and the Coen group from the JIC Norwich for useful discussions in the early stage of the study.

References

Andalo, C., Cruzan, M.B., Cazettes, C., Pujol, B., Burrus, M. & Thebaud, C. 2010. Post-pollination barriers do not explain the persistence of two distinct *Antirrhinum* subspecies with parapatric distribution. *Plant Syst. Evol.* **286**: 223–234.

Anderson, E. 1949. *Introgressive Hybridization*. Whiley & Sons, New York.

Arnold, M.L. 1997. *Natural Hybridization and Evolution*. Oxford University Press, Oxford.

Barton, N.H. 2001. The role of hybridization in evolution. *Mol. Ecol.* **10**: 551–568.

Barton, N.H. & Hewitt, G.M. 1989. Adaptation, speciation and hybrid zones. *Nature* **341**: 497–503.

Britch, S.C., Cain, M.L. & Howard, D.J. 2001. Spatio-temporal dynamics of the *Allonemobius fasciatus-A. socius* mosaic hybrid zone: a 14-year perspective. *Mol. Ecol.* **10**: 627–638.

Broyles, S.B. 2002. Hybrid bridges to gene flow: a case study in milkweeds (*Asclepias*). *Evolution* **56**: 1943–1953.

Buggs, R.J.A. & Pannell, J.R. 2007. Ecological differentiation and diploid superiority across a moving ploidy contact zone. *Evolution* **61**: 125–140.

Bull, C.M. 1991. Ecology of parapatric distributions. *Annu. Rev. Ecol. Syst.* **22**: 19–36.

Campbell, D.R., Waser, N.M. & Wolf, P.G. 1998. Pollen transfer by natural hybrids and parental species in an *Ipomopsis* hybrid zone. *Evolution* **52**: 1602–1611.

Chapman, R.E., Wang, J. & Bourke, A.F.G. 2003. Genetic analysis of spatial foraging patterns and resource sharing in bumble bee pollinators. *Mol. Ecol.* **12**: 2801–2808.

Cruzan, M.B. & Arnold, M.L. 1994. Assortative mating and natural selection in an iris hybrid zone. *Evolution* **48**: 1946–1958.

Currat, M., Ruedi, M., Petit, R.J. & Excoffier, L. 2008. The hidden side of invasions: massive introgression by local genes. *Evolution* **62**: 1908–1920.

Darvill, B., Knight, M.E. & Goulson, D. 2004. Use of genetic markers to quantify bumblebee foraging range and nest density. *Oikos* **107**: 471–478.

Demesure, B., Sodzi, N. & Petit, R.J. 1995. A set of universal primers for amplification of polymorphic noncoding regions of mitochondrial and chloroplast DNA in plants. *Mol. Ecol.* **4**: 129–131.

Excoffier, L., Foll, M. & Petit, R.J. 2009. Genetic consequences of range expansions. *Annu. Rev. Ecol. Evol. Syst.* **40**: 481–501.

Good, P. 2000. *Permutation Tests*. Springer, New York.

Grant, V. 1971. *Plant Speciation*. Columbia University Press, New York.

Hairston, N.G., Wiley, R.H., Smith, C.K. & Kneidel, K.A. 1992. The dynamics of 2 hybrid zones in Appalachian salamanders of the genus *Plethodon*. *Evolution* **46**: 930–938.

Hewitt, G. 2000. The genetic legacy of the Quaternary ice ages. *Nature* **405**: 907–913.

Jones, K.N. & Reithel, J.S. 2001. Pollinator-mediated selection on a flower color polymorphism in experimental populations of *Antirrhinum* (Scrophulariaceae). *Am. J. Bot.* **88**: 447–454.

Lavergne, S. & Molofsky, J. 2007. Increased genetic variation and evolutionary potential drive the success of an invasive grass. *Proc. Natl. Acad. Sci. USA* **104**: 3883–3888.

Leebens-Mack, J. & Milligan, B.G. 1998. Pollination biology in hybridizing *Baptisia* (*Fabaceae*) populations. *Am. J. Bot.* **85**: 500.

Levin, D.A. 1971. The origin of reproductive isolating mechanisms in flowering plants. *Taxon* **20**: 91–113.

Lexer, C., Kremer, A. & Petit, R.J. 2006. Shared alleles in sympatric oaks: recurrent gene flow is a more parsimonious explanation than ancestral polymorphism. *Mol. Ecol.* **15**: 2007–2012.

Linder, C.R. & Rieseberg, L.H. 2004. Reconstructing patterns of reticulate evolution in plants. *Am. J. Bot.* **91**: 1700–1708.

Martinsen, G.D., Whitham, T.G., Turek, R.J. & Keim, P. 2001. Hybrid populations selectively filter gene introgression between species. *Evolution* **55**: 1325–1335.

Melo-Ferreira, J., Boursot, P., Suchentrunk, F., Ferrand, N. & Alves, P.C. 2005. Invasion from the cold past: extensive introgression of mountain hare (*Lepus timidus*) mitochondrial DNA into three other hare species in northern Iberia. *Mol. Ecol.* **14**: 2459–2464.

Molitor, A. 1937. Zur vergleichenden Psychobiologie der akuleaten Hymenopteren auf experimenteller Grundlage. *Biologia Generalis.* **13**: 294–333.

Muir, G. & Schlotterer, C. 2005. Evidence for shared ancestral polymorphism rather than recurrent gene flow at microsatellite loci differentiating two hybridizing oaks (*Quercus* spp.). *Mol. Ecol.* **14**: 549–561.

Pannell, J.R. & Pujol, B. 2009. The paradoxical spread of a new Y chromosome – a novel explanation. *Trends Ecol. Evol.* **24**: 59–63.

Petit, R.J., Bodenes, C., Ducousso, A., Roussel, G. & Kremer, A. 2003. Hybridization as a mechanism of invasion in oaks. *New Phytol.* **161**: 151–164.

Potts, B.M. & Reid, J.B. 1988. Hybridization as a dispersal mechanism. *Evolution* **42**: 1245–1255.

Potts, B.M. & Reid, J.B. 1990. The evolutionary significance of hybridization in *Eucalyptus*. *Evolution* **44**: 2151–2152.

Pujol, B. & Pannell, J.R. 2008. Reduced responses to selection after species range expansion. *Science* **321**: 96.

Pujol, B., Zhou, S.R., Vilas, J.S. & Pannell, J.R. 2009. Reduced inbreeding depression after species range expansion. *Proc. Natl. Acad. Sci. USA* **106**: 15379–15383.

Rieseberg, L.H. & Soltis, D.E. 1991. Phylogenetic consequences of cytoplasmic gene flow in plants. *Evol. Trends Plants* **5**: 65–84.

Rohwer, S., Bermingham, E. & Wood, C. 2001. Plumage and mitochondrial DNA haplotype variation across a moving hybrid zone. *Evolution* **55**: 405–422.

Rothmaler, W. 1956. *Taxonomische monographie der Gattung Antirrhinum*. Akademie-Verlag, Berlin.

Schemske, D.W. & Morgan, M.T. 1990. The evolutionary significance of hybridization in eucalyptus. *Evolution* **44**: 2150–2151.

Schwinn, K., Venail, J., Shang, Y.J., Mackay, S., Alm, V., Butelli, E. *et al.* 2006. A small family of MYB-regulatory genes controls floral pigmentation intensity and patterning in the genus *Antirrhinum*. *Plant Cell* **18**: 831–851.

Shuker, D.M., Underwood, K., King, T.M. & Butlin, R.K. 2005. Patterns of male sterility in a grasshopper hybrid zone imply

accumulation of hybrid incompatibilities without selection. *Proc. Biol. Sci.* **272**: 2491–2497.
Sutton, D.A. 1988. *A revision of the Tribe Antirrhineae*. British Museum (Natural History), London and New York.
Tastard, E. 2009. *Maintien d'une zone hybride de gueules de loup (Antirrhinum majus): role de quelques intéractions biologiques*. PhD dissertation of the University of Toulouse Paul Sabatier, Toulouse.
Tastard, E., Andalo, C., Giurfa, M., Burrus, M. & Thebaud, C. 2008. Flower colour variation across a hybrid zone in *Antirrhinum* as perceived by bumblebee pollinators. *APIS* **2**: 237–246.
Tiffin, P., Olson, M.S. & Moyle, L.C. 2001. Asymmetrical crossing barriers in angiosperms. *Proc. R. Soc. Lond. B Biol. Sci.* **268**: 861–867.
Wendel, J. & Doyle, J. 1998. *Phylogenetic Incongruence: Window into Genome History and Molecular Evolution in Molecular Systematics of Plants*. Kluwer Academic Publishers, Boston.
Whibley, A.C. 2004. *Molecular and Genetic Variation Underlying the Evolution of Flower Colour in Antirrhinum*. PhD dissertation of the University of East Anglia, Norwich.
Whibley, A.C., Langlade, N.B., Andalo, C., Hanna, A.I., Bangham, A., Thebaud, C. *et al.* 2006. Evolutionary paths underlying flower color variation in *Antirrhinum*. *Science* **313**: 963–966.

Supporting information

Additional Supporting Information may be found in the online version of this article:
Table S1a *Antirrhinum majus pseudomajus* population characteristics.
Table S1b *Antirrhinum majus striatum* population characteristics.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

Received 7 December 2010; revised 15 March 2011; accepted 16 March 2011

Chapitre 3

Ecology predicts parapatric distributions in two closely related *Antirrhinum majus* subspecies

Running title: Ecological niches and parapatric distributions

Ecology predicts parapatric distributions in two closely related *Antirrhinum majus* subspecies

---

Accepté dans Evolutionary Ecology

Khimoun A [1,2], Cornuault J [1,2], Burrus M [1,2], Pujol B [1,2], Thebaud C [1,2], Andalo C [1,2]

AK and JC have contributed equally to this work and should be considered sharing first authorship

[1] Université de Toulouse; UPS; EDB (Laboratoire Evolution et Diversité Biologique); UMR5174; 118 route de Narbonne, F-31062 Toulouse cedex 9, France.

[2] CNRS; EDB (Laboratoire Evolution et Diversité Biologique); UMR5174; F-31062 Toulouse, France

Author for correspondence: khimoun@cict.fr

Phone number: 00 33 (0)5 61 55 74 86

Fax number: 00 33 (0)5 61 55 73 27

Abstract

Using a species distribution model, we reconstructed the environmental niches of *Antirrhinum majus pseudomajus* and *A. m. striatum*, two closely related species with parapatric distributions. We tested whether retention of ancestral environmental niche (i.e. niche conservatism) or adaptation to different ecological conditions (i.e. niche divergence) could explain the maintenance of their non-overlapping geographic ranges. We found that the environmental niche of *A. m. pseudomajus* is almost twice as large as that of *A. m. striatum*, with substantial overlap indicating that *A. m. pseudomajus* and *A. m. striatum* should co-occur frequently within the geographic range of *A. m. striatum*. By analysing contact zones where both subspecies are geographically close, we found that the presence of one subspecies instead of the other was significantly influenced by particular combinations of climatic factors. Since independent genetic evidence indicates that the two subspecies have experienced phases of range overlap at or near contact zones over the course of their evolutionary history, we propose that ecological niche displacement might be an important factor in explaining the absence of current range overlap between *A. majus* subspecies.

Keywords: *Antirrhinum majus*, parapatry, niche modeling, niche divergence, ecological character displacement.

INTRODUCTION

Parapatry (*i.e.* geographically separated ranges abutting along common boundaries) is widespread among closely related species of plants and animals (Anderson and Evensen 1978; King 1993). Yet, it often remains a challenge to identify the mechanisms that prevent range overlap between parapatric species. It has long been hypothesized that parapatric distributions may be associated with spatial changes in environmental factors, species interactions in areas of contact, or dispersal limitation even in the absence of physical barriers (Bull 1991). When environmental factors produce spatial segregation, parapatric distributions are often found to match sharp or gradual environmental transitions (*e.g.* temperature, precipitation). In such cases of ecotonal changes, closely related species may be specifically adapted to the environmental conditions defining their respective habitats across their parapatric boundary. If so, geographic isolation of the two species may be maintained through local adaptation, which could drive divergence, reproductive isolation, and ultimately speciation between parapatric incipient species (Dobzhansky 1951; Funk 1998; Schluter 2001; Rundle and Nosil 2005; Schluter 2009). Under such an ecological divergence scenario, parapatric species should occupy different environmental niches both in regions where they are found to be completely isolated from each other (*i.e.* allopatry) and in regions where they are found in contact on each side of their common boundary (*i.e.*contact zone).

Competitive exclusion between closely related species is thought to prevent range overlap and therefore shape instances of parapatric distributions (Hutchinson 1953; Connor and Bowers 1987). This is expected when species have diverged when isolated geographically (*i.e.* allopatry) but retained the same environmental niche (Peterson et al. 1999; Wiens 2004; Wiens and Graham 2005). Under such an allopatric divergence scenario, both species should conserve their ancestral environmental niche in allopatry albeit they diverged (*i.e.* niche conservatism). However, when they meet secondarily after range expansion, ecological

character displacement may occur in the area of sympatry, which could result in a partitioning of their environmental niche on each side of their common boundary (Ricklefs 2010). Thus, if competition plays a role in the parapatric distributions of closely related species, one might expect greater differences in environmental niches in sympatry than in allopatry (Brown and Wilson 1956; Dayan and Simberloff 2005).

In addition to niche divergence and competition, dispersal limitation has also the potential to shape parapatric distributions (Garcia-Ramos et al. 2000). This is because limited dispersal can prevent range overlap between geographically isolated populations of closely related species, thereby maintaining them distributed in parapatry. As a consequence, niche differences between species might be observed that are caused by environmental differences associated to their separated distribution ranges, due to spatial autocorrelation in environmental variables between the regions over which the species are distributed, rather than actual niche divergence between species (McCormack et al. 2010).

Species Distribution Models (SDMs) provide a powerful tool to investigate the role of environmental conditions in shaping spatial patterns of biodiversity (Cicero 2004; Guisan and Thuiller 2005; Elith et al. 2006). Because they can predict habitat suitability in unsampled areas and help to track species range shift in response to climate change (Wiens et al. 2009), SDMs are extensively used in the context of biodiversity inventories and conservation planning (Kremen et al. 2008). More recently, SDMs have been used in another context at the interface of ecology and evolutionary biology, to assess environmental niche differentiation among species (Kozak et al. 2008; Warren et al. 2008) and explore divergence mechanisms at the origin of species formation (titmice, Cicero 2004; dendrobatid frogs, Graham et al. 2004; wild tomatoes, Nakazato et al. 2008; Mexican jays, McCormack et al. 2010).

In this study, we investigated the role of environmental conditions in the maintenance of geographic isolation between two subspecies of snapdragon plants, *Antirrhinum majus*

pseudomajus and *A. m. striatum*, by conducting an analysis of the geographic distribution of their environmental niche. Our aim was to examine whether niche divergence may explain parapatric distributions in this system, and infer indirectly from our results whether ecological processes might also be involved. *A. majus* provides an ideal study system to assess niche differentiation in a species divergence context since the two subspecies used in this study are endemic to the Pyrenean mountains and surrounding Mediterranean plains. However, while they both cover a large range of environmental conditions, their geographic distribution remains parapatric throughout their range.

MATERIALS AND METHODS

Study system

Antirrhinum majus (Scrophulariaceae) is an herbaceous short-lived perennial plant characterised by a patchy distribution centred over the Pyrenees, between north-eastern Spain and south-western France. The geographic range of *A. m. striatum* is surrounded by the range of *A. m. pseudomajus* (Figure 1), which do not overlap. *A. m. striatum* and *A. m. pseudomajus* come into contact at the margins of their ranges. In the contact zones between those two subspecies, introgressive hybridization occurs and local replacement of *A.m. pseudomajus* by *A.m. striatum* is observed in the west part of the contact zone and conversely in the east part (Khimoun et al. 2011).

Environmental data

A total of 31 environmental variables were used to construct the SDMs: Fifteen climatic variables (including annual trends, seasonality, extreme climatic parameters; Hijmans et al. 2005), eight soil variables and four vegetation variables (Table 1). Previous studies have shown that vegetation indices could improve niche models when used in combination with

climatic variables (Buermann et al. 2008). In particular, we used the mean Normalized Difference Vegetation Index (NDVI, average of the monthly NDVI values) that measures the density of vegetation and is therefore a good proxy of biotic competitive environment (Nakazato et al. 2010).

Environmental niche modeling

A. majus occurrence records: We first characterized the environmental niches of *A. m. pseudomajus* and *A. m. striatum* from allopatric populations, where interspecific interactions do not operate. Genetic analyses of chloroplast and nuclear genes revealed the absence of genetic introgression between these populations (Khimoun et al. 2011), indicating that they have remained in allopatry for a long time. Consequently, we considered that the two subspecies niches cannot differ as a result of inter-specific competition when only these allopatric populations are taken into account. In total, we used occurrence data of 26 *A. m. pseudomajus* populations and 9 *A. m. striatum* populations. The small number of *A. m. striatum* allopatric populations is inherent to the system since *A. m. striatum* has a spatially restricted range compared to *A. m. pseudomajus*.

Preliminary models: The MaxEnt approach (Phillips et al. 2006) was used to predict each subspecies occurrence outside its sampled range. This method is appropriate for presence-only species records and has been shown to perform well in comparison with alternative approaches (Elith et al. 2006). We did not perform an *a priori* procedure of variable selection and we included the 31 environmental variables to construct the models. We used default values for the convergence threshold (10^{-5}) and the minimum number of iterations (500). Following Phillips et al. (2006), we constructed five types of model that included different features: (i) only linear features (raw environmental variables); (ii) linear

and quadratic features (including the square of environmental variables); (iii) linear, quadratic and product features (adding the products of pairs of environmental variables); (iv) threshold features (using binary thresholds on environmental variables) and (v) hinge features (like a linear feature but constant beyond a threshold). The suitability scores obtained from the five models were then averaged to give a single model called the average model. It has been previously shown that the averaging of different model predictions (ensemble modeling) should outperform single model predictions, even when some of the models perform badly (Grenouillet et al. 2010).

Average prediction assessment and decision threshold: Model predictive performance is generally assessed by randomly dividing occurrence data into training (75%) and testing (25%) datasets (Fielding and Bell 1997; Araujo et al. 2005). However, this approach is not appropriate with limited occurrence records because the training dataset may be too small to calibrate the model correctly. Instead, we used a jackknife procedure thought to perform well with relatively small datasets (Pearson et al. 2007). This approach consists in alternately removing each locality from the training dataset, and calibrating the model with the N-1 remaining localities. For each of the N constructed models, MaxEnt suitability scores were converted into presence/absence (Pearson et al. 2004) by using the "lowest training presence score" as a decision threshold. Model predictive performance was then evaluated as the model's capacity to successfully predict presence at the left-out localities, taking into account the estimated prevalence (*i.e.* the proportion of the study area occupied by the subspecies). This procedure was carried out as described in Pearson et al. (2007), except that in our study, both successes and failures were weighted by prevalence. Thus, failures to predict an observed presence when the species is present in most of the study area (high prevalence) receive a high penalty in our analyses. This procedure has been implemented in a script

(available upon author's request) used with R software (R Development Core Team 2007). Further details on this test are provided in the supplementary online material (see Appendix 1).

Tests for environmental niche divergence between the two subspecies

Niche identity

This procedure allowed us to test whether there was a difference between the two subspecies' niches, regardless of the environmental conditions available in their respective backgrounds. Niche overlap was quantified using the Schoener's D metric (Schoener 1968). Significance was assessed using randomization tests which consisted in creating a series of SDMs from randomized datasets of occurrences (pseudoreplicates) and computing the Schoener's D metric for each pseudoreplicate. This procedure permitted to build a null hypothesis that we compared with the observed D values (Warren et al. 2008). We used MaxEnt in batch mode to construct the SDMs from 1,000 pseudoreplicates following the procedure described above, i.e. using ensemble modeling. We then used the R software (R Development Core Team 2007) to average the predictions of these models and compute the significance of the Schoener's D metric.

Background test

Because differences in environmental niches can be due to spatial autocorrelation, the *background test* (see Warren et al. 2008) was performed to assess whether the potential environmental niches of *A. m. pseudomajus* and *A. m. striatum* were more similar or divergent than would be expected given the environmental conditions available in the regions they occupy (*i.e.* their backgrounds). For this procedure, pseudoreplicates were generated through randomization of the occurrence locations of one subspecies by randomly sampling the same

number of points within its background. The test was carried out in both directions (randomization of *A. m. pseudomajus* and *A. m. striatum* occurrences). The background test is two-tailed because the observed values of the Schoener's D metric can be greater (niche conservatism) or lower (niche divergence) than the null hypothesis. We also adapted this test to average modeling. Both niche identity and niche background tests are described in more detail in the supplementary online material (see Appendix 2).

Test for niche divergence between the two subspecies in contact zones

First, we produced a graphical representation of the set of environmental variables that were suitable for each subspecies, both when they were in contact and in allopatry. To this aim, we conducted a Principal Component Analysis on environmental variables for introgressed populations of the contact zones and non-introgressed allopatric populations. Second, we tested if *A. m. pseudomajus* and *A. m. striatum* occupy different niches in sympatry and focussed on genetically introgressed populations at or near contact zones (n=5 for *A. m. striatum* and n=6 for *A. m. pseudomajus*; (see Khimoun et al. 2011). Since introgression reveals evolutionary interactions between subspecies during phases of geographic contact (Khimoun et al. 2011), we considered that the absence of current records for one or the other subspecies in the area of introgression might therefore reflect the influence of biotic and/or abiotic factors rather than dispersal limitation. We performed a logistic regression with quasi-binomial error to analyze the effect of climatic variables on presence/absence data of the two subspecies throughout the introgression area. Environmental variables were standardized to mean 0 and unit variance and summarized into principal coordinates to avoid multicollinearity between climatic, soil and vegetation variables. Because differences in environmental conditions generally increased with geographic distance, residuals were considered to be spatially autocorrelated. Autocorrelation was assumed to decrease exponentially with

geographic distance. All statistical analyses were performed using R software (R Development Core Team, 2007).

RESULTS

Environmental niches

The average models predicted occurrences at the test localities better than chance (P <0.001 for *A. m. pseudomajus* and *A. m. striatum* models). The SDM built with *A. m. striatum* allopatric populations yielded a projected distribution restricted to the Pyrenees Mountains and surrounding valleys whereas the SDM built with *A. m. pseudomajus* allopatric populations predicted that its range should extend beyond the Pyrenees to the Mediterranean coast and surrounding plains (Figure 1). The environmental niche of *A. m. pseudomajus* is almost twice as large as that of *A. m. striatum* (prevalence of 0.47 and 0.26, respectively). The predicted overlap of the two subspecies distributions is 0.19 (Schoener's D). This overlap reveals that environmental conditions should be suitable for the establishment of both subspecies over a large area within the actual range of *A. m. striatum* (Figure 1). In particular, both subspecies should be present at every sampled locality of the contact zones. Thus, based on the distribution of environmental conditions that are suitable for both subspecies, *A. m. pseudomajus* and *A. m. striatum* are expected to be frequently found in sympatry (Figure 1).

Niche divergence in allopatry

The niche identity test indicated that the current niche overlap between the two *Antirrhinum* subspecies is significantly lower than expected by chance, when considering allopatric populations only (P = 0.05 see Figure 2). The background test, which takes into account

242 background differences in environmental conditions, indicated divergence between the
243 potential niches of *A. m. striatum* and *A. m. pseudomajus* when the occurrence randomisation
244 procedure was applied for *A. m. striatum* ($P = 0.02$ for Schoener's D metric). Such divergence
245 was not found when the randomisation procedure was applied to *A. m. pseudomajus* ($P = 0.56$
246 for Schoener's D metric; see Figure 3). Because the observed difference between subspecies
247 niches was greater than the expected difference under the hypothesis that *A. m. striatum* was
248 randomly distributed within its background, our results imply that *A. m. striatum* occurs in a
249 part of its environmental background where conditions are particularly dissimilar to the
250 environmental niche of *A. m. pseudomajus*.
251
252 **Niche divergence in contact zones**
253 The first three principal components from the PCA explained 75% of the total variance (47%,
254 18% and 10% for PC1, PC2 and PC3, respectively). PC1 was mostly correlated with annual
255 mean temperatures, extreme values of temperature and precipitation, PC2 was correlated with
256 variables describing temperature variation, and PC3 was correlated with soil structure, soil
257 nutrient and water availability. Vegetation variables were poorly correlated to the first three
258 PCA axes (see supplementary online material, Appendix 3). According to the logistic
259 regression analysis, the presence of one subspecies instead of the other was significantly
260 affected by all pairwise interactions between PC1, PC2 and PC3 (Table 2). The association of
261 higher precipitation with lower temperatures, higher thermal amplitudes and wetter, more
262 compact and nutrient-deprived soils significantly increased the probability of observing *A. m.*
263 *striatum* instead of *A. m. pseudomajus* (Figure 4). Although the two subspecies were expected
264 to co-occur in contact zones on the basis of environmental factors, they show significant
265 ecological niche divergence throughout the area of introgression. Moreover, the contact zone

populations of the two subspecies were found in a subset of the environmental niche which only partly overlapped the environmental niche of allopatric populations.

DISCUSSION

Geographic segregation is not predicted by environmental niche modelling

Niche models based on environmental factors indicate that the predicted geographic range of both subspecies is larger than their actual range, with both subspecies occupying only partly the geographic range where environmental conditions are suitable for their establishment (Figure 1). They also show that the geographic range of *A. m. pseudomajus* should be larger than the predicted range of *A. m. striatum* and include part of it on the basis of their predicted environmental niches (Figure 1). Thus, on the basis of environmental conditions alone, *A. m. pseudomajus* and *A. m. striatum* should co-occur frequently within the *A. m. striatum* geographic range. Such co-occurrence is however not observed in nature. Even in localities where populations bear the signature of gene exchange between subspecies, populations of the two subspecies remain geographically separated. The predicted co-occurence of the two subspecies in the area of introgression and over most of *A. m. striatum* range could be due to the poor resolution of the environmental grids used for calibrating the models and/or the small number of occurrence records. Although the number of *A. m. striatum* populations that we used is relatively small, these populations cover the entire geographic range of the subspecies. Furthermore, the model predictive performances were good, suggesting that the set of environmental variables considered (31 variables) is sufficient for correctly representing both subspecies niches. This suggests that factors besides environmental factors, such as dispersal limitation and biotic interactions, could have important effects in explaining the absence of

range overlap between taxa where they may share similar environmental requirements (Sillero 2011).

Dispersal limitation can prevent organisms from colonizing an area of suitable habitat in its entirety (Holt 2003). This seems likely in *Antirrhinum* since seeds are mostly dispersed over short distances from maternal plants, even though their small size and weight (<15 mg) may allow occasional long-distance dispersal (Andalo et al. 2010). While dispersal limitation might therefore explain the global pattern of subspecies geographic segregation, it can hardly explain the complete absence of sympatry where the two subspecies were once geographically close enough to exchange genes (Khimoun et al. 2011). Thus, it is possible that biotic interactions (e.g. competition, predation or parasitism), possibly in interaction with environmental factors, prevent the two subspecies from occupying the whole common area that is suitable to their establishment (Miller 1967).

Niche divergence in contact zones

In contact zones, the presence of one or the other subspecies was correlated with environmental conditions. Our results suggest that *A. m. striatum* populations are ecologically distinct from *A. m. pseudomajus* populations wherever they could occur in sympatry (see Figure 4). Evidence for recent gene flow among subspecies populations in the contact zones suggests that differences in the environmental niche cannot be explained by dispersal limitation (Khimoun et al. 2011). Expansion processes might generate a geographically structured distribution of genetically introgressed populations (Currat et al. 2008). Under such scenario, the fact that *A. m. striatum* invaded the previous range of *A. m. pseudomajus* in the west part of the contact zone whereas *A. m. pseudomajus* reciprocally invaded *A. m. striatum* range in the east part of the contact zone could therefore be related to neutral demographic processes (Khimoun et al. 2011). However, our results show that such replacement is highly

correlated to environmental conditions (Figure 4). Thus, it seems possible that adaptation of each subspecies to different local conditions could explain the local asymmetry of subspecies replacement in the area of introgression. The two subspecies parapatric distribution could then reflect differential abilities to survive and reproduce in varying local environmental conditions. It is also possible that, depending upon environmental conditions, one subspecies has a superior ability over the other one to take up and/or use water and nutrient resources when they become available in a competitive environment. Although we do not have direct evidence to support this hypothesis, the observed pattern of niche displacement between the two subspecies in contact zones compared to allopatry suggests that competition between subspecies may be a major factor explaining why *A. m. pseudomajus* and *A. m. striatum* are prevented from occupying their entire potential niche, thereby maintaining their parapatric distributions.

To conclude, our results revealed that environmental factors alone could not be responsible for *Antirrhinum* subspecies parapatric distributions. We found that differences in environmental niches between subspecies in areas of contact were greater than expected by chance and related to possible differences in resource use, in particular water and soil nutrients that are often limiting in Mediterranean mountains. Thus, we argue that range overlap might be prevented in our study system by ecological niche displacement driven by competition, recognizing that more comprehensive geographic sampling and a functional characterization of differences in resource use between subspecies are required before any firm conclusion can be reached (Losos 2000).

Acknowledgements

All the work presented in this article was supported by core funding from CNRS and the University of Toulouse Paul Sabatier, France. AK and JC are supported by a PhD grant from the French ministry of research. We thank, Yann Bourgeois and Loïc Tudesque for technical help with GIS files. We thank Jérôme Chave, Jean-Baptiste Ferdy, and Ferran Palero for their useful comments on the manuscript. We also thank the associate editor Mario Vallejo-Marin and two anonymous reviewers for their useful suggestions. This work is part of the "Laboratoire d'Excellence (LABEX)" entitled TULIP (ANR -10-LABX-41).

346 **REFERENCES**

347

348 Andalo C, Cruzan MB, Cazettes C et al (2010) Post-pollination barriers do not explain the
349 persistence of two distinct *Antirrhinum* subspecies with parapatric distribution. Plant Syst
350 Evol 286:223-234

351 Anderson S, Evensen MK (1978) Randomness in allopatric speciation. Syst Zool 27:421-430

352 Araujo MB, Pearson RG, Thuiller W, Erhard M (2005) Validation of species-climate impact
353 models under climate change. Global Change Biol 11:1504-1513

354 Brown WL, Wilson EO (1956) Character displacement. Syst Zool 5:49-64

355 Buermann W, Saatchi S, Smith TB et al (2008) Predicting species distributions across the
356 Amazonian and Andean regions using remote sensing data. J Biogeogr 35:1160-1176

357 Bull CM (1991) Ecology of parapatric distributions. Annu Rev Ecol Syst 22:19-36

358 Cicero C (2004) Barriers to sympatry between avian sibling species (Paridae : Baeolophus) in
359 local secondary contact. Evolution 58:1573-1587

360 Connor EF, Bowers MA (1987) The spatial consequences of interspecific competition. Ann.
361 Zool. Fennici 24: 213-226

362 Currat M, Ruedi M, Petit RJ, Excoffier L (2008) The hidden side of invasions: massive
363 introgression by local genes. Evolution 62:1908-1920

364 Dayan T, Simberloff D (2005) Ecological and community-wide character displacement: the
365 next generation. Ecol Lett 8:875-894

366 Dobzhansky T (1951) Genetics and the Origin of Species, Columbia University Press, New
367 York.

368 Elith J, Graham CH, Anderson RP et al (2006) Novel methods improve prediction of species'
369 distributions from occurrence data. Ecography 29:129-151

370 Fielding AH, Bell JF (1997) A review of methods for the assessment of prediction errors in
371 conservation presence/absence models. Environ Conserv 24:38-49
372 Funk D.J. (1998) Isolating a role for natural selection in speciation:host adaptation and sexual
373 isolation in Neochlamisus bebbianae leaf beetles. Evolution 52:1744–1759
374 Garcia-Ramos G, Sanchez-Garduno F, Maini PK (2000) Dispersal can sharpen parapatric
375 boundaries on a spatially varying environment. Ecology 81:749-760
376 Graham CH, Ron SR, Santos JC, Schneider CJ, Moritz C (2004) Integrating phylogenetics
377 and environmental niche models to explore speciation mechanisms in dendrobatid frogs.
378 Evolution 58:1781-1793
379 Grenouillet G, Buisson L, Casajus N, Lek S (2010) Ensemble modelling of species
380 distribution: the effects of geographical and environmental ranges. Ecography 34:9-17
381 Guisan A, Thuiller W (2005) Predicting species distribution: offering more than simple
382 habitat models. Ecol Lett 8:993-1009
383 Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution
384 interpolated climate surfaces for global land areas. Int J Climatol 25:1965-1978
385 Holt RD (2003) On the evolutionary ecology of species' ranges. Evol Ecol Res 5:159-178
386 Hutchinson GE (1953) The concept of pattern of ecology. Proc. Acad. Nat. Sci., Philadelphia
387 105:1-12
388 Justice CO, Vermote E, Townshend JRG et al (1998) The Moderate Resolution Imaging
389 Spectroradiometer (MODIS): Land remote sensing for global change research. Ieee T Geosci
390 Remote 36:1228-1249
391 Khimoun A, Burrus M, Andalo C et al (2011) Locally asymmetric introgressions between
392 subspecies suggest circular range expansion at the *Antirrhinum majus* global scale. J Evol
393 Biol 24:1433-1441

King M (1993) Species evolution. The role of chromosome change. Cambridge University Press, Cambridge

Kozak KH, Graham CH, Wiens JJ (2008) Integrating GIS-based environmental data into evolutionary biology. Trends Ecol Evol 23:141-148

Kremen C, Cameron A, Moilanen A et al (2008) Aligning conservation priorities across taxa in Madagascar with high-resolution planning tools. Science 320:222-226

Losos JB (2000) Ecological character displacement and the study of adaptation. Proc. Natl. Acad. Sci. USA 97:5693–5695

McCormack JE, Zellmer AJ, Knowles LL (2010) Does niche divergence accompany allopatric divergence in *Aphelocoma* jays as predicted under ecological speciation?: Insights from tests with niche models. Evolution 64:1231-1244

Miller RS (1967) Pattern and process in competition. Adv Ecol Res 4:1-74

Nakazato T, Bogonovich M, Moyle LC (2008) Environmental factors predict adaptive phenotypic differentiation within and between two wild andean tomatoes Evolution 62:774–792.

Pearson RG, Dawson TP, Liu C (2004) Modelling species distributions in Britain: a hierarchical integration of climate and land-cover data. Ecography 27:285-298

Pearson RG, Raxworthy CJ, Nakamura M, Peterson AT (2007) Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. J Biogeogr 34:102-117

Peterson AT, Soberon J, Sanchez-Cordero V (1999) Conservatism of ecological niches in evolutionary time. Science 285:1265-1267

Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. Ecol Model 190:231-259

Rausher MD (2008) Evolutionary Transitions in Floral Color. Int. J. Plant Sci. 169: 7–21

419 Ricklefs RE (2010) Evolutionary diversification, coevolution between populations and their
420 antagonists, and the filling of niche space. Proc Natl Acad Sci U S A 107:1265-1272
421 Rundle HD, Nosil P (2005) Ecological speciation. Ecology Letters 8:336-352
422 Schluter D (2001) Ecology and the origin of species. Trends in Ecol Evol 16:372-380
423 Schluter D (2009) Evidence for Ecological Speciation and Its Alternative. Science 323:737-
424 741
425 Schoener TW (1968) *Anolis* lizards of BIMINI - Resource partitioning in a complex fauna.
426 Ecology 49:704-&
427 Sillero N (2011) What does ecological modelling model? A proposed classification of
428 ecological niche models based on their underlying methods. Ecol Model 222:1343-1346
429 Warren DL, Glor RE, Turelli M (2008) Environmental niche equivalency *versus*
430 conservatism: Quantitative approaches to niche evolution. Evolution 62:2868-2883
431 Wiens JJ (2004) Speciation and ecology revisited: Phylogenetic niche conservatism and the
432 origin of species. Evolution 58:193-197
433 Wiens JJ, Graham CH (2005) Niche conservatism: Integrating evolution, ecology, and
434 conservation biology. Annu Rev Ecol Evol Syst 36:519-539
435 Wiens JA, Stralberg D, Jongsomjit D, Howell CA, Snyder MA (2009) Niches, models, and
436 climate change: Assessing the assumptions and uncertainties. Proc Natl Acad Sci U S A
437 106:19729-19736
438
439
440

List of appendices

Appendix 1: Test of model accuracy.

Appendix 2: Background test

Appendix 3: Pearson correlation coefficients between environmental variables and the first two Principal Component axes.

Table legend

Table 1: Summary of environmental variables used in the study

Table 2: Results of the GLM analysing the effects of environmental variables on the distribution of *A. m. pseudomajus* and *A.m. striatum* in contact zones.

*** = p<0.001

454 **Khimoun et al._Table 1**

Environmental variable	**Abbreviation**	**Resolution**	**Source**
Mean Normalized Vegetative Index	NDVI	30 x 30 s	MODIS (Justice et al. 1998) http://modis.gsfc.nasa.gov/
Productivity	PRO	30 x 30 s	MODIS
Annual Mean Temperature	BIO1	30 x 30 s	WorldClim (Hijmans et al. 2005) (http://www.worldclim.org/)
Mean Diurnal Temperature Range	BIO2	30 x 30 s	WorldClim
Isothermality	BIO3	30 x 30 s	WorldClim
Temperature Seasonality (Coefficient of Variation)	BIO4	30 x 30 s	WorldClim
Maximal Temperature of the Warmest Month	BIO5	30 x 30 s	WorldClim
Minimal Temperature of the Coldest Month	BIO6	30 x 30 s	WorldClim
Temperature Annual Range	BIO7	30 x 30 s	WorldClim
Mean Temperature of the Wettest Quarter	BIO8	30 x 30 s	WorldClim
Mean Temperature of the Driest Quarter	BIO9	30 x 30 s	WorldClim

Mean Temperature of the Warmest Quarter	BIO10	30 x 30 s	WorldClim
Mean Temperature of the Coldest Quarter	BIO11	30 x 30 s	WorldClim
Annual Precipitation	BIO12	30 x 30 s	WorldClim
Precipitation of the Wettest Month	BIO13	30 x 30 s	WorldClim
Precipitation of the Driest Month	BIO14	30 x 30 s	WorldClim
Precipitation Seasonality (Coefficient of Variation)	BIO15	30 x 30 s	WorldClim
Precipitation of the Wettest Quarter	BIO16	30 x 30 s	WorldClim
Precipitation of the Driest Quarter	BIO17	30 x 30 s	WorldClim
Precipitation of the Warmest Quarter	BIO18	30 x 30 s	WorldClim
Precipitation of the Coldest Quarter	BIO19	30 x 30 s	WorldClim
Forest Land	FOR	5 x 5 min	Harmonized World Soil Database, Land Use and Land Cover (http://www.iiasa.ac.at/Research/LUC/External-World-soil database/HTML/LandUseShares.html?sb=9)
Grass,Scrub or Woodland	GRA	5 x 5 min	Harmonized World Soil Database, Land Use and Land Cover

Rooting Conditions	ROO	5 x 5 min	Harmonized World Soil Database, Soil Qualities For Crop Production http://www.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/SoilQualityData.html?sb=11
Bulk Density Data	BUL	5 x 5 min	Distributed Active Archive Center for Biogeochemical Dynamics http://daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id=569
Field Capacity Data	FIE	5 x 5 min	Distributed Active Archive Center for Biogeochemical Dynamics
Profile Available Water Capacity Data	WAT	5 x 5 min	Distributed Active Archive Center for Biogeochemical Dynamics
Soil Carbon Density Data	SOI	5 x 5 min	Distributed Active Archive Center for Biogeochemical Dynamics
Thermal Capacity Data	THE	5 x 5 min	Distributed Active Archive Center for Biogeochemical Dynamics

Total Nitrogen Density Data	NIT	5 x 5 min	Distributed Active Archive Center for Biogeochemical Dynamics
Wilting Point Data	WIL	5 x 5 min	Distributed Active Archive Center for Biogeochemical Dynamics

Khimoun et al._Table 2

Variables	Estimates	Standard error	Df	P
PC1	-2.75	0.20	1	*P<0.001*
PC2	-15.52	0.03	1	*P<0.001*
PC3	38.03	0.06	1	*P<0.001*
PC1 x PC2	-0.48	0.18	1	*P<0.001*
PC1 x PC3	0.81	0.02	1	*P<0.001*
PC2 x PC3	-1.67	0.04	1	*P<0.001*

Figure legends

Figure 1: Sampled localities and predicted potential niches of *A. m. pseudomajus* and *A. m. striatum*

Black symbols represent allopatric populations and white symbols represent introgressed populations of the contact zones. Squares and triangles represent respectively *A. m. pseudomajus* and *A. m. striatum* populations. Blue and orange regions represent the potential niches of *A. m. pseudomajus* and *A. m. striatum* respectively, as predicted using only allopatric populations. Areas of niche overlap are represented in green.

Figure 2: Niche identity test of *A. m. pseudomajus* and *A. m. striatum*

The histograms represent the null distributions of niche overlap values from 1,000 pseudoreplicates. The arrow indicates the observed value of the potential niche overlap of *A. m. pseudomajus* and *A. m. striatum*.

Figure 3: Test of niche divergence between *A. m. pseudomajus* and *A. m. striatum* (background test).

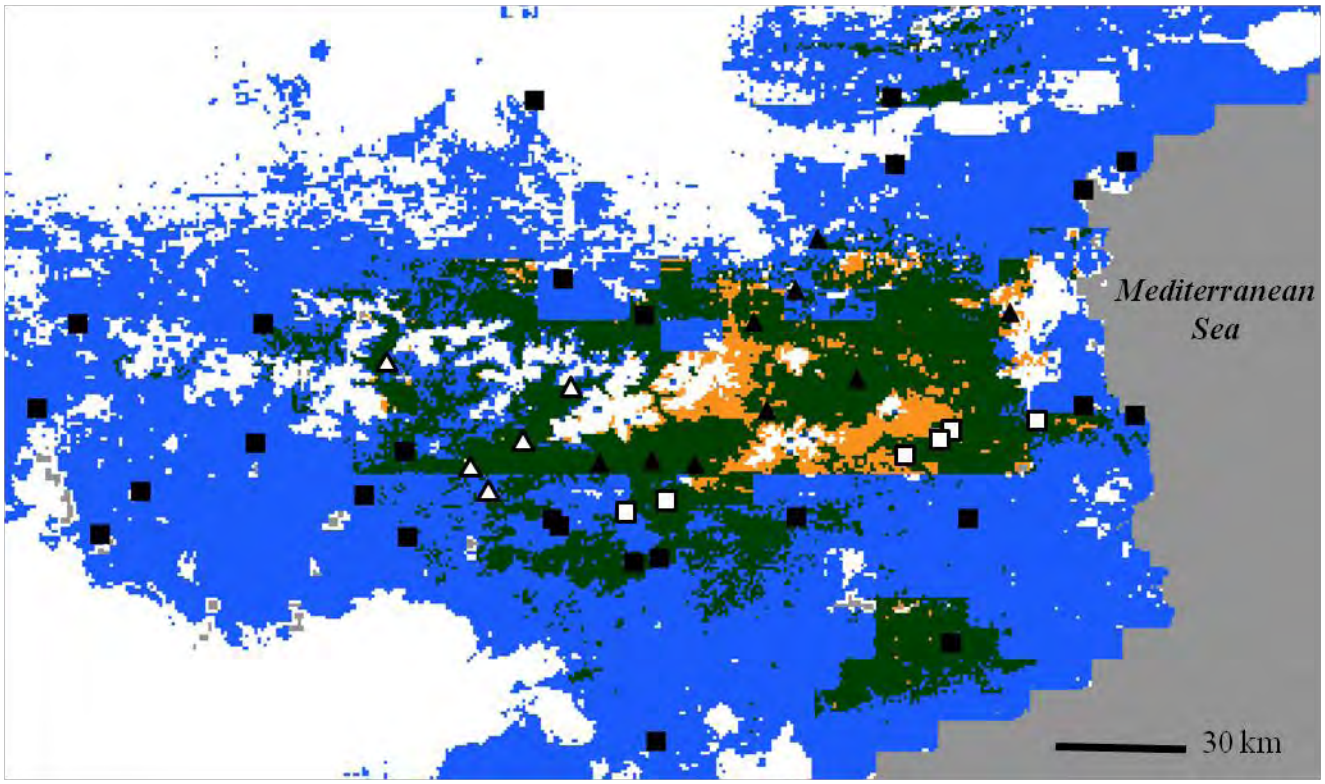
The blue histogram represents the null hypothesis when occurrences of *A. m. pseudomajus* are randomized within its background and the yellow histogram represents the null hypothesis when occurrences of *A. m. striatum* are randomized within its background. The arrow indicates the observed overlap values between *A. m. pseudomajus* and *A. m. striatum*.

Figure 4: 3D-Display of introgressed and allopatric *A. m. pseudomajus* and *A. m. striatum* populations along Principal Components 1, 2 and 3.

Blue and yellow spheres represent allopatric populations of *A. m. pseudomajus* and *A. m. striatum* respectively. Black spheres represent *A. m. pseudomajus* introgressed populations and grey spheres represent introgressed *A. m. striatum* populations. Squares within point clouds represent the barycentre positions.

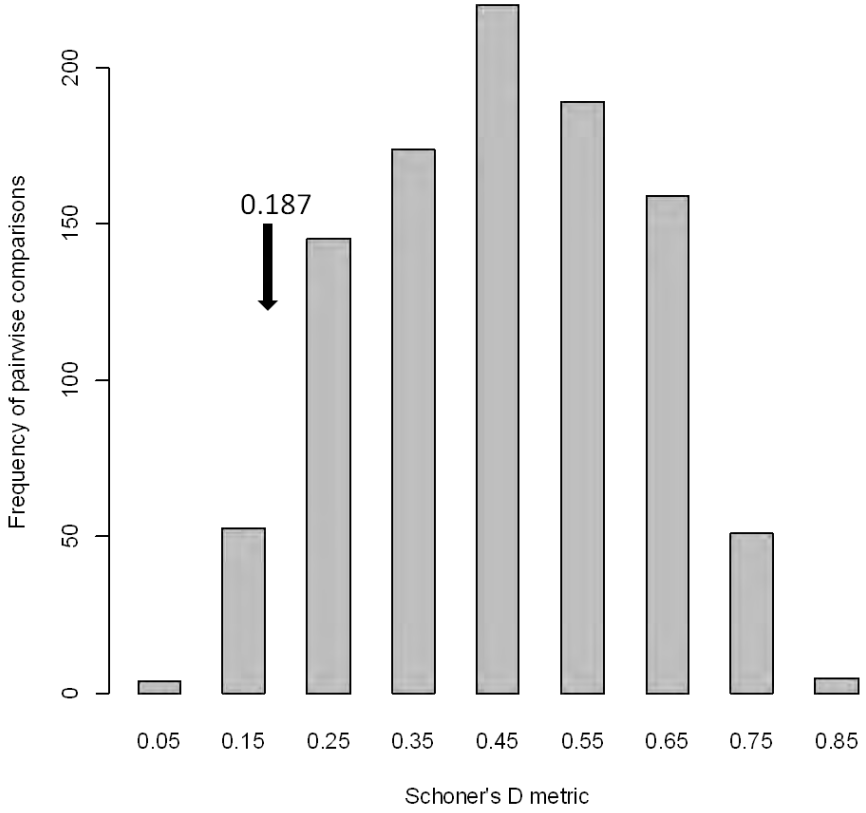
487 **Khimoun et al_Figure 1**

488



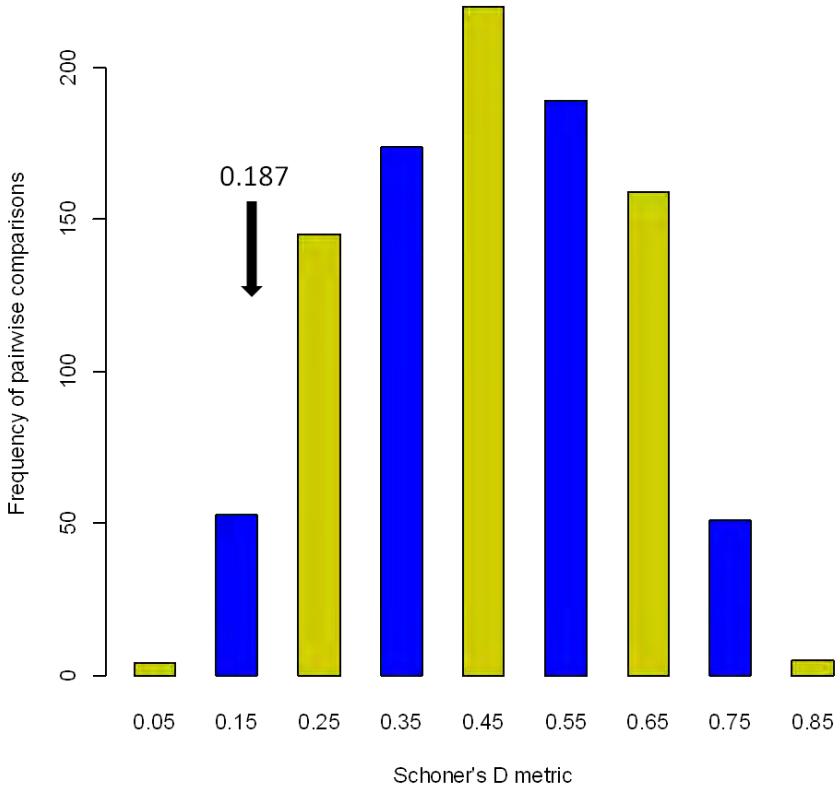
489

490 **Khimoun et al_Figure 2**



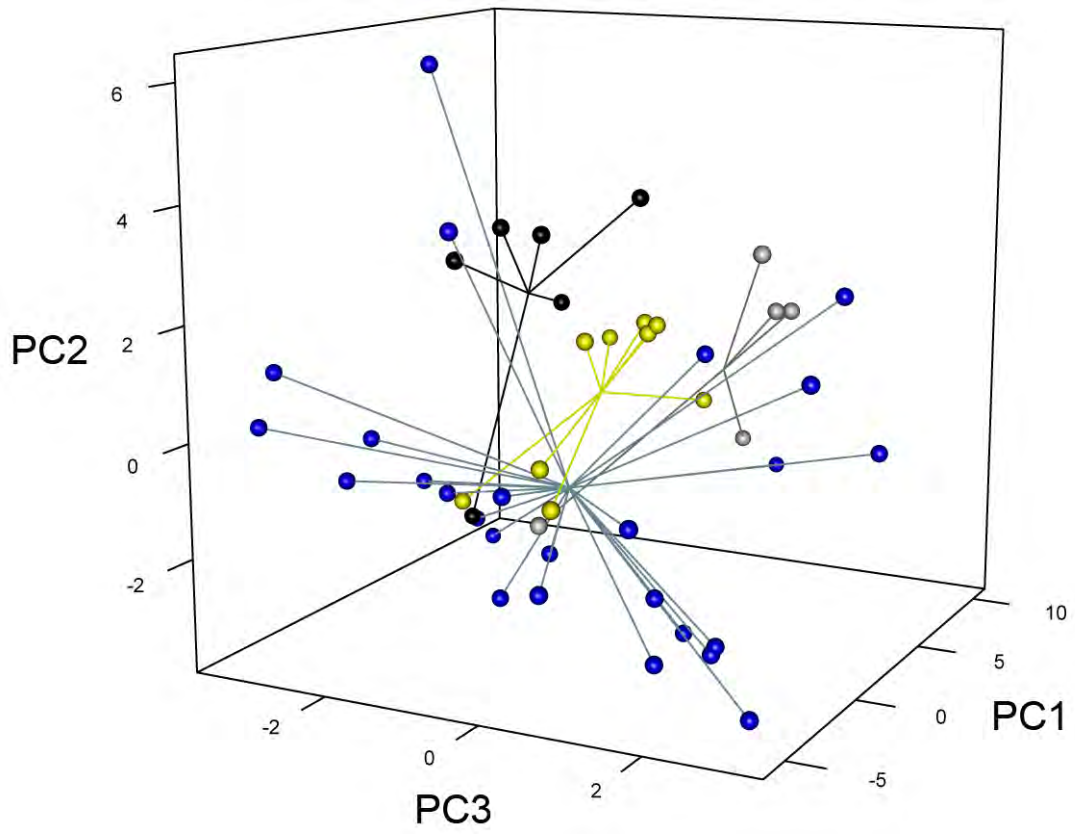
491

492 **Khimoun et al_Figure 3**



493

494 **Khimoun et al_Figure 4**



495

496

497

Online supplementary material

Appendix 1: Test of model accuracy.

To build our predictive performance statistic, we attributed a score of 1 for well predicted presence at the removed locality and a score of 0 for a failure. We subtracted from these scores the probability of a successful prediction by chance (that is the subspecies prevalence), obtaining quantity Q. This quantity is maximum (1) when the model correctly predicts a presence while the prevalence is low. In contrast, it is minimum (-1) when the model fails to predict a presence while the prevalence is high. We then statistically tested whether the accuracy of our predictions is greater than expected by chance taking into account the predicted prevalence of subspecies. The sum of the values of Q for the whole set of jackknife trials was then compared to the value of the same statistic obtained under a null hypothesis of random assignment of success/failure (1 or 0 values) over each jackknife trial.

Appendix 2: Background test

To test whether potential environmental niches of *A. m. pseudomajus* and *A. m. striatum* differ from what is expect given the environmental conditions of the geographic regions they occupy, we implemented an R language version of the background test implemented in ENM tools (Warren et al. 2008). A null distribution of 1000 overlap values was generated by comparing the average SDM of one subspecies to the averaged SDM built from random points drawn within the background of the other subspecies. This process was repeated for both subspecies. Observed measures of niche overlap were then compared to these null distributions (Warren et al. 2008). To perform this test, we first needed to define the spatial background from which each subspecies is supposed to select a particular habitat. We estimated both subspecies backgrounds using the minimum convex polygon algorithm (which

corresponds to the smallest polygon containing all presence sites in which no internal angle exceeds 180 degrees).

Appendix 3: Pearson correlation coefficients between environmental variables and the first two Principal Component axes.

Environmental variables	PCA 1	PCA2	PCA3
BIO1	**-0.933 *****	0.319 *	-0.057 ns
BIO2	-0.477 ***	-0.421 **	**0.730 *****
BIO3	-0.291 *	-0.483 ***	**0.733 *****
BIO4	-0.472 ***	-0.230 ns	**0.529 *****
BIO5	**-0.962 *****	0.216 ns	0.085 ns
BIO6	**-0.865 *****	0.413 ns	-0.221 ns
BIO7	**-0.595*****	-0.324 *	**0.676 *****
BIO8	**-0.657 *****	0.383 **	-0.193 ns
BIO9	**-0.647 *****	0.259 ns	0.021 ns
BIO10	**-0.941 *****	0.302 *	-0.042 ns
BIO11	**-0.907 *****	0.360 *	-0.134 ns
BIO12	**0.977 *****	-0.143 ns	-0.060 ns
BIO13	**0.916 *****	0.017 ns	-0.200 ns
BIO14	**0.951 *****	-0.211 ns	-0.016 ns
BIO15	**-0.658 *****	0.353 *	-0.136 ns
BIO16	**0.967 *****	-0.106 ns	-0.016 ns

BIO17	**-0.940 *****	-0.155 ns	-0.055 ns
BIO18	**0.895 *****	-0.083 ns	-0.009 ns
BIO19	**0.781 *****	-0.171 ns	0.046 ns
BUL	-0.450 **	**-0.793 *****	-0.358 *
CAR	0.449 **	**0.748 *****	0.321 *
FIE	0.521 ***	**0.667 *****	0.287 ns
FOR	0.338 *	0.287 ns	**-0.368 ***
GRA	0.248 ns	**-0.311 ***	**0.318 ***
NDVI	-0.025 ns	0.290 ns	**-0.334 ***
NIT	0.385 **	**0.730 *****	0.349 *
PRO	-0.110 ns	**0.397 ****	-0.164 ns
ROO	**0.704 *****	-0.041 ns	-0.011 ns
THE	-0.449 **	**-0.793 *****	-0.358 *
WAT	**0.513 *****	**0.633 *****	0.299 *
WIL	**0.515 *****	**0.624 *****	0.284 ns

*** = p<0.001 ; ** = p<0.01; * = p<0.05; ns = non-significant

Chapitre 4

Cline discordance and cytonuclear disequilibrium in an *Antirrhinum* hybrid zone

Running title: Cline discordance in an *Antirrhinum* hybrid zone

Cline discordance and cytonuclear disequilibrium in an *Antirrhinum* hybrid zone

---

Khimoun A [1,2], Palero F [3], Andalo C [1,2] , Burrus M [1,2], Thebaud C [1,2],

[1] Université de Toulouse; UPS; EDB (Laboratoire Evolution et Diversité Biologique); UMR5174; 118 route de Narbonne, F-31062 Toulouse cedex 9, France.

[2] CNRS; EDB (Laboratoire Evolution et Diversité Biologique); UMR5174; F-31062 Toulouse, France

[3] Unitat Mixta Genòmica i Salut CSISP-UV/Institut Cavanilles Universitat de Valencia, C/ Catedrático Jose Beltran 2, 46980 Paterna, Spain

Author for correspondence: khimoun@cict.fr

Phone number: 00 33 (0)5 61 55 74 86

Fax number: 00 33 (0)5 61 55 73 27

Abstract

Cytonuclear disequilibrium is expected to arise in hybrid populations when there is non-random mating among the hybridizing taxa, high levels of migration of individuals with different cytonuclear genotypes from nearby populations, epistatic interactions between the nuclear and cytoplasmic genomes affecting hybrid fitness, or a postmating barrier to reproduction. In the present study, a stable pattern of discordant clines is found among a chloroplast marker and a nuclear gene related to flower color in a contact zone among populations of the perennial plant *Antirrhinum majus*. A binomial logistic regression model allowed us to estimate cline shape parameters and cline position at two different time points. The results obtained in our study indicate that clines for both the flower-color coding gene (*ROS1*) and the chloroplast marker have remained at the same position during the last decade. Furthermore, the separation between both clines seems to have remained constant as well. The similarities shared among clines at different and non-linked markers points out to non-neutral processes (e.g. epistatic selection) being responsible for the maintenance of the hybrid zone.

Keywords: *Antirrhinum majus*, cytonuclear disequilibrium, Pyrennees, speciation, epistasis, assortative mating.

INTRODUCTION

Zones of secondary contact are of great interest to evolutionary biologists because they provide unique opportunities to observe the evolutionary interactions between divergent but related taxa (Barton, Hewitt, 1985; Harrison, 1993). When divergent taxa meet in secondary contact, a number of outcomes are possible. First, if strong reproductive isolation has evolved in allopatry as an incidental by-product of natural selection, sexual selection, or genetic drift, the two taxa can remain distinct (Coyne, Orr, 2004; Price, 2008). However, reproductive barriers are often incomplete, and taxa in secondary contact may be able to interbreed to some extent. If reproductive isolating mechanisms are nonexistent, widespread hybridization and introgression will lead towards neutral diffusion of alleles between the two parental populations and mixing of the two parental gene pools (Barton, Gale, 1993; Endler, 1977). A third potential outcome is secondary contact with partial reproductive isolation between two taxa. This can lead to the formation of a stable hybrid zone between the two taxa if some form of selection maintains the zone (Barton, Hewitt, 1981). Despite creating taxonomic confusion (Cicero, Johnson, 1998; Hubbard, 1969), zones of secondary contact and hybridization provide unique opportunities to examine the factors that contribute to evolutionary divergence and reproductive isolation (Price, 2008).

Within this context, molecular genetic data can assist in distinguishing among the possible outcomes of secondary contact. If reproductive isolation between two taxa is complete, genetically distinct individuals should coexist in sympatry in a contact zone with no genetically intermediate individuals present. If neutral diffusion of alleles is occurring, a gradual, clinal transition in genetic characters will form between the two taxa, with genetically intermediate individuals occurring over a geographic area that increases with time. If a stable hybrid zone has formed, genetically intermediate individuals will be present in a more restricted geographic area between the two parental populations, and one of the two

main types of hybrid zones may form. In the first case, the hybrid zone is maintained by the differential fitness among hybrid and parental populations, and does not necessarily correspond to or track specific geographic or environmental features (so-called "tension zones" (Barton, Gale, 1993; Barton, Hewitt, 1981). On the second case, under a spatially-dependent selection model, the two parental populations would occur in distinct environments so that hybrids are restricted to a transition area that intermediate between the two parental environments (Moore, 1977; Moore, Price, 1993). The width and placement of this area is generally determined by the extent of the intermediate habitat.

Given the complexity of these processes, accurate inference of evolutionary dynamics in secondary contact zones generally requires examination of multiple types of genetic evidence (Edwards *et al.*, 2005). Indeed, within a single hybrid zone, character clines for genetic markers or phenotypic traits can display similar (concordant) or different (discordant) patterns. One particular type of character cline that has traditionally received much attention within hybrid zones is cytonuclear disequilibria, which corresponds to a non-random association between alleles or genotypes at a nuclear locus with haplotypes at a cytoplasmic locus. Because cytonuclear organelles are often uniparentally inherited, analysis of cytonuclear disequilibria can identify the direction of hybridization and whether certain species are the maternal or paternal contributors to hybrid progeny. A substantial body of theoretical work also exists that predicts the nature of disequilibria that are expected to arise from a variety of evolutionary forces, including assortative mating, hybridization, and selection (Asmussen *et al.*, 1987; Asmussen *et al.*, 1989; Asmussen, Basten, 1994).

Cytoplasmic introgression (*i.e.* the transfer of chloroplastic or mitochondrial material of one species to another species) is rendered possible by fertile hybrids that occupy the contact zones and act as "bridges to gene flow", therefore allowing gene exchange between species to occur (Broyles, 2002). In the case of plant species, asymmetric chloroplast

introgression (*i.e.* unidirectional introgression the chloroplast of one species into the gene pool of the other species) may be generated and maintained either by selection or dispersal (Currat *et al.,* 2008). Selection might explain the local asymmetry in the introgression pattern, even in the presence of bidirectional gene flow. In such case, we would expect one cytonuclear discordant association to provide a selective advantage over the other. This may result from intrinsic attributes of species such as pre-zygotic asymmetric barriers (e.g. asymmetric pollen-style incompatibilities (Cruzan, Arnold, 1994), sex-biased dispersal (Petit *et al.,* 2003), or post-zygotic asymmetric barriers (e.g. partial hybrid sterility; (Shuker *et al.,* 2005), which are commonly attributed to cytonuclear epistatic interactions (Levin, 1971; Tiffin *et al.,* 2001)). Selection driven by extrinsic factors might also explain pattern of asymmetric introgression between species. Environmental-based selection, may act on the epistatic interaction between a nuclear and a chloroplastic gene leading to the higher selective value of only one cytonuclear association over the other. Finally, although selection may explain asymmetric chloroplast introgressions, neutral processes such as seed/pollen differential dispersal abilities or very recent range expansions may also explain such pattern. Asymmetric chloroplast introgression is a common outcome when invading populations can spread their nuclear genes at a long distance by means of pollen flow and seed dispersal is limited (Petit *et al.,* 2003). Such hypothesis is not exclusive because asymmetric chloroplast introgression could also be explained by demographic expansion through seed dispersal. In such case, the demographic imbalance between invaders and residents would result in the asymmetric introgression of genes from the resident species genome into the invader genome (Currat *et al.,* 2008).

Antirrhinum majus pseudomajus and *A. m. striatum* are two interfertile subspecies that occupy parapatric ranges. Recent studies, indicated that floral trait segregation, in

combination with pollinator behavior, could affect the maintenance of flower color polymorphism in one particularly narrow hybrid zone between *A. m. pseudomajus* and *A. m. striatum* subspecies (Tastard *et al.*, 2008; Tastard *et al.*, 2011). A previous analysis on this same hybrid zone using three nuclear genes has revealed an abrupt cline for the *ROS1* gene (which is responsible for most of the flower color differentiation between the yellow-flowered *A. m. striatum* and magenta-flowered *A. m. pseudomajus*) whereas no clinal pattern was observed for the other two genes, *Pallida* and *Dichotoma*, responsible for the flower color and symmetry (Whibley *et al.*, 2006). Cytonuclear associations of *ROS1* alleles (flower color) and chloroplast haplotypes have been recently analyzed at a global scale, indicating that asymmetric patterns of chloroplast introgression could result from range expansion of each subspecies in opposite directions (Khimoun *et al.*, 2011).

Although selection on the *ROS1* gene may explain the maintenance of a narrow hybrid zone, the stability of this hybrid zone has never been tested and no detailed description of cytonuclear disequilibria has been provided so far. In the present study, we narrow down the magnitude and pattern of chloroplast introgression between the two subspecies by focusing on the previously studied *ROS1* hybrid zone (Whibley *et al.*, 2006). This will allow us to test locally the unidirectional pattern of chloroplast introgression previously observed at a global scale and to assess the geographic extent of chloroplast introgression along the hybrid zone. Finally, the inclusion of samples from 2002 and 2010 will allow us to compare the shape and position of the chloroplast and *ROS1* clines and infer any temporal displacement of the two clines.

MATERIAL AND METHODS

Plant material sampling strategy

A total of 598 hybrid plants were sampled during two field sessions (with n=403 in 2002 and n=195 in 2010) from the hybrid zone described by (Whibley *et al.*, 2006) (Figure 1). Despite having a larger sample size, the 2002 sample included populations scattered around the actual transition area, so that sampling was intensified in that region during 2010. In both years, geographic coordinates of populations were recorded by using a GPS device (Garmin, Olathe, Kansas, USA). For each individual, young leaves and shoot tips were collected and stored at -20°C until DNA was extracted by using the DNeasy Plant Mini kit (Qiagen, Hilden, Germany).

ROSEA genotyping

The *ROSEA* locus is made of 2 MYB genes, out of which *ROS1* has the main role in flower colour variation (Schwinn *et al.*, 2006). *ROS1* sequences can be grouped in 3 main haplotypes *ROS1-Ma*, *ROS1-Mb* and *ROS1-Y* (Whibley, 2004). *ROS1-Ma* and *ROS1-Mb* haplotypes are diagnostic of *A. m. pseudomajus* and are grouped under the name of *ROS1-M* whereas the *ROS1-Y* haplotype is diagnostic of *A. m. striatum* (Whibley, 2004). *ROS1* genotypic data were available for 496 hybrids that were previously examined by (Whibley *et al.*, 2006). We obtained *ROS1* genotypic data for the remaining 256 hybrids using the RG4/RR21, RG6/RR21 and RG1/RR21 primers in a single PCR reaction, following the protocol established by (Whibley, 2004).

PCR-RFLP analysis of chloroplast DNA

Maternal lineages were determined for the 752 hybrids by genotyping the 1.6kb *psbC [psII 44-kDa protein] - trnS [tRNA-Ser(UGA)]* intergenic region, using the CS universal primers (Demesure *et al.*, 1995). Sequencing of this chloroplast region revealed 2 haplotypes that differed at 2 SNP loci, one of which was included in a *MseI* restriction site. We therefore obtained two different haplotypes after digestion of the *psbC-trnS* fragment by the *Mse I* enzyme. Haplotype I was characterised by eight *Mse I* restriction sites that generated a nine-band profile on agarose gel. Haplotype II was characterised by a 10-band profile. We followed the protocol established by Khimoun *et al.* (2011).

Cline fitting

There is an increasing variety of approaches we can rely on in order to analyze hybrid zones. For example, we could use general purpose computer programs such as Structure (Pritchard *et al.*, 2000) that seek patterns in ancestries of individuals without reference to any explicit model of hybrid zones. However, the most generalized approach depends on hybrid zone models that predict patterns of allele frequencies and fit corresponding parametric curves (Analyse program (Barton, Baird, 1995)). The main advantage of such models is that the estimated parameter values can be directly related to the processes of interest (e.g. natural selection).

In the present paper, we focused on a simple binomial logistic regression model of the form $\text{Log}[p/1-p] = 1/1+e^{-a+bx}$ that fits the allele frequencies for successive values of x (spatial location). This is among the most widely used methods (Barton, Gale, 1993) and corresponds to the assumption that populations are distributed along a one-dimensional habitat. Simply stated, the model predicts the probability of finding one type of allele as we move from one parental population in one extreme of the distribution (e.g. a parental population of *A. m.*

striatum) to another. Under this model, the cline center can be defined as the spatial position at which the allele frequency is *P*= 0.5. Estimates for the cline center observed in different markers and different years will inform us about the staggering (spatial displacement) of clines at different loci and the displacement of clines with time.

As for the cytonuclear disequilibrium, there are a number of ways by which the statistical association between a nuclear and cytoplasmic locus could be measured (Weir, Wilson, 1986). We followed Asmussen *et al.* (1987), whom introduced several measures which arise naturally in a wide class of biological models. If we consider a diploid population with two alleles 'A' and 'a' at a nuclear locus and two other alleles M and m at, for example, at a chloroplastic locus, then there are six possible genotypes. At the level of genotypes, nuclear-cytoplasmic disequilibria can be measured by the departures of genotypic frequencies from expectations under random association (see Asmussen *et al.* (1987) for details). Finally, the allelic association between cytoplasmic and nuclear markers can be described by the gametic disequilibrium parameter D, which measures the departure of gametic frequencies from expectations under random association (Hedrick, 1983).

Estimates of cline parameters and their confidence intervals (CI), together with estimates of the position of the cline center for the different years were obtained through several Mathematica functions as implemented in v8.0.4 (e.g. LogitModelFit, Wolfram 2008). Mathematica code for estimating cytonuclear-disequilibria is available from the authors upon request.

RESULTS

A clinal pattern was recovered for both the nuclear gene responsible for flower color (*ROS1*) and the chloroplast marker. Allele frequencies for the 'typical' *A. majus striatum* populations

(the most common allele in *A. m. striatum* parental populations) decay gradually as one moves south east and towards the red-flowered *A. m. pseudomajus* populations (placed down the valley) (Figure 2). Parameter estimates obtained for the cline-model fitting of the data obtained in the years 2002 and 2010 are summarized in Table 2. The dataset for 2002 included too few samples for the chloroplast transition area and therefore estimated confidence intervals are extremely large. Nevertheless, the 95% confidence intervals indicate that the clinal pattern is significant in all cases except for the 2002 chloroplast cline (only case for which the parameter values include zero). It is worth pointing out that parameter b would be proportional to the cline width, and represents the intensity of the balance between migration and selection.

Our cline-fitting approach also supports the stability of the discordance and the cline position for the chloroplast and *ROS1* alleles among different years. From our genetic dataset, the cline center for the nuclear marker would be placed closer to the parental *A. m. striatum* populations (that is, further to the north-western area of the hybrid zone) whereas the opposite would be the case for the cline center of the chloroplast marker. The estimated value for the staggering or separation between both clines would be around 6 km in 2002 and 4 km in 2010 (Figure 3). It should be pointed out that this difference in value may be related to the higher intensity of sampling along the chloroplast transition area during the year 2010.

Finally, the cytonuclear disequilibrium estimates along the cline indicates that there is a clear statistical association between the nuclear and cytoplasmic loci within the hybrid zone during 2010. In the figure 4 it can be seen that the value of D shows a peak (D1 = 0.08) around an area 20 km apart from the extreme *A. m. striatum* populations. Again, the value obtained for this parameter when using the data collected during the year 2002 should be taken with caution due to the extreme values for the chloroplast cline.

DISCUSSION

In cases where reproductive isolation between two closely related species is not complete, secondary contact leads to hybridization and often to stable hybrid zones (Arnold, 1997). In some cases, such as the *Antirrhinum* hybrid zone analyzed in our study, differential introgression of loci across a contact zone can be observed. The fact that genes encoding factors directly involved in reproductive isolation and/or environmental adaptation are often subject to adaptive selection, may lead to restricted gene flow and clinal allele distribution across hybrid zones. Indeed, the strength of such selection is often reflected in the slope of the allele frequency cline (Barton, Hewitt, 1985). In contrast, gene flow can be less restricted for those loci that do not decrease hybrid survival, reproductive performance or local adaptation, and are not closely linked to selected loci.

Distinct selection regimes can lead to differential introgression among different genes across the same hybrid zone. Where such discordances have been reported, this has been commonly attributed to natural selection (directly or indirectly through genetic hitchhiking) on at least one of the loci (McDonald, 1994). However, discordant differentiation patterns among genes can also result from factors others than selection (Bierne *et al.*, 2003). For instance, random genetic drift introduces a large amount of stochastic variation in the evolution of unlinked neutral loci (Bierne *et al.*, 2003) and this heterogeneity may be inflated in secondary contact zones by the combined effect of hybridization and recombination. Furthermore, sharp clinal variation can also be observed for neutral markers as a transitory stage immediately after secondary contact (Edwards, Skibinski, 1987). As even a large heterogeneity in patterns of genetic differentiation among different genetic markers is often consistent with genetic drift expectations (Bierne *et al.*, 2003), selection can only be inferred if its effects exceed the expectations from these stochastic effects and all factors should be taken into account.

Non- coincidence between *ROS1* and chloroplast cline centres

The results obtained in the present study indicate that cline centers for the *ROS1* gene and for the *psbC-trnS* intergenic region did not coincide either in 2002 nor in 2010 (Figure 3). In both years, the chloroplast cline center is displaced around 5 Km eastern compared to the cline center for *ROS1*. The most western populations (two for the 2002 sampling and five for the 2010 sampling) were fixed for the chloroplast Haplotype II whereas they remain polymorphic for *ROS1* gene (Figure 2). as previously stated, such pattern might simply result from the neutral action of genetic drift. Because chloroplast effective size is smaller than nuclear effective size, genetic drift might have fixed chloroplast Haplotype II more rapidly than nuclear *ROS1* gene in these localities, especially as they include a smaller number of individuals (field observation). Nonetheless, under this hypothesis of neutral fixation of chloroplast haplotypes and persistence of *ROS1* polymorphism we do not expect to observe width concordance between *ROS1* and chloroplast clines. However, the distribution of chloroplast haplotypes along the hybrid zone was clinal, with a narrow slope similar in shape with the one of *ROS1* cline. It seems rather unlikely that genetic drift could cause patterns so similar among independent loci, so this result would indicate that genetic drift is not the only factor needed to explain the observed distribution of chloroplast haplotypes and *ROS1* alleles in the hybrid zone.

Another explanation for why the center of the chloroplast cline is displaced compared to the center of *ROS1* cline would be associated with sex-biased dispersal. Chloroplasts are maternally inherited and dispersed by seeds whereas nuclear genes are biparentally inherited and dispersed by both seeds and pollen. Indeed, dispersal characteristics of *A. majus* could support this hypothesis of differential dispersal ability of seeds and pollen. *A. majus* seeds are very small and light (<15 mg (Andalo *et al.*, 2010)) and can mostly be dispersed at a short

distance of the maternal plant by gravity. In contrast, *A. majus* pollen is transported by bumblebees (several *Bombus* species) and carpenter bees (*Xylocopa* sp.) and is therefore likely to migrate across long distances (Whibley, 2004) . The estimated distance covered by carpenter bees of the species *Xylocopa violacea* can reach up to 1.2 km (Molitor, 1937) whereas bumblebees of the species *Bombus terrestris* can cover up to 2.8 km (Chapman *et al.*, 2003; Darvill *et al.*, 2004). From east to west, the hybrid zone is positioned along a gradient of elevation (Figure 1) that seems difficult to cross for seeds (which disperse by gravity) whereas pollinators and consequently pollen could cross. Again, on the basis of this hypothesis alone, we do not necessarily expect to observe similar cline widths between the nuclear and chloroplast clines. Although probably playing an important role, pollen/seeds differential dispersal ability would not seem to explain, on its own, the distribution of chloroplast haplotypes and *ROS1* alleles in the hybrid zone.

As stated above, the hybrid zone matches with an elevation gradient whose extremes correspond to the two parental subspecies (Figure 2). We thus expect environmental conditions (e.g. temperature, precipitations, soil characteristics...) to vary along the hybrid zone according to the altitude. Heterogeneous environmental conditions along the hybrid zone could drive differential local selection. A previous work conducted at the global scale of *A. majus* subspecies ranges revealed the correlation of *A. m. striatum* populations (yellow flowers) with higher precipitations, lower temperatures, higher thermal amplitudes and more compacted and wetter soils poor in nutrients compared to *A. m. pseudomajus* populations (magenta flowers) (Khimoun *et al.*, Accepted). Along the elevation gradient, the clinal distribution of *ROS1* alleles could be explained by local selection on flower color driven by environmental conditions. Although the hypothesis of differential local adaptation of the two subspecies to different environmental conditions could explain the cline of *ROS1* gene, it does not explain the cline observed for chloroplast haplotypes. Indeed, under such hypothesis of

selection on flower color, we do not expect concordance of cline width because only the *ROS1* gene should be prevented to flow between the two subspecies, whereas the chloroplast genes should freely move between the two subspecies. An alternative hypothesis to explain the concordance of *ROS1* and chloroplast cline width is to invoke some form of epistatic selection. Under such hypothesis, individuals with the *ROS1-Y* allele and chloroplast Haplotype II would be the more adapted to the environmental conditions of the west extremity of the hybrid zone (higher altitude conditions) whereas *ROS1-M* and chloroplast Haplotype I would be the more adapted to the environmental conditions of the east extremity of the hybrid zone (lower altitude conditions). In the core of the hybrid zone, with intermediate environmental conditions, we would expect the association of *ROS1-M* alleles with chloroplast Haplotype II to be selectively advantageous whereas the association of *ROS1-Y* alleles with chloroplast Haplotype I would be maladapted to these environmental conditions. Our results call for experimentally testing the adaptive significance of the different associations of *ROS1* alleles and chloroplast haplotypes under contrasting environmental conditions.

However, as a final note, we stress that the spatial regression of admixture proportions does not capture all the complexity of hybrid zones: their semi-permeable nature, the fine scale discordance of clines and the interplay of various component of reproductive isolation. Admixture proportions and cline width are only a rough summary of how genomes intermix in hybrid zones and hybrid zones cannot simply be summarised by logistic variation of admixture proportions. We think the present study will be of great help as a complementary procedure to estimate the position of hybrid zone centres and intensity of selection.

CONCLUSION

Although probably playing an important role, genetic drift and seed/pollen differential dispersal ability, alone, cannot explain together the non-coincidence of *ROS1* and chloroplast cline center and the concordance of *ROS1* and chloroplast cline width. A selective hypothesis of epistatic interaction between the nuclear *ROS1* gene and the chloroplast could be invoked to explain the observed distribution of chloroplast haplotypes and *ROS1* alleles in the hybrid zone. Finally, the stability of the nuclear and chloroplast cline over 8 years could be the result of low seeds and pollen dispersal abilities and/or high selection on the interaction *ROS1* alleles and chloroplast haplotypes.

Bibliography

Andalo C, Cruzan MB, Cazettes C, *et al.* (2010) Post-pollination barriers do not explain the persistence of two distinct *Antirrhinum* subspecies with parapatric distribution. *Plant Systematics and Evolution* **286**, 223-234.

Arnold ML (1997) *Natural hybridization and evolution.* Oxford University Press, Oxford.

Asmussen MA, Arnold J, Avise JC (1987) Definition and properties of disequilibrium statistics for associations between nuclear and cytoplasmic genotypes. *Genetics* **115**, 755-768.

Asmussen MA, Arnold J, Avise JC (1989) The effects of assortative mating and migration on cytonuclear associations in hybrid zones. *Genetics* **122**, 923-934.

Asmussen MA, Basten CJ (1994) Sampling theory for cytonuclear disequilibria. *Genetics* **138**, 1351-1363.

Barton NH, Baird SJE (1995) Analyse - An application for analyzing hybrid zones (Freeware, available at request from Stuart JE Baird), Edinburgh.

Barton NH, Gale KS (1993) Genetic analysis of hybrid zones. In: *Hybrid zones and the evolutionary process* (ed. Harrison RG), pp. 13-45. Oxford University Press, Oxford.

Barton NH, Hewitt GM (1981) The genetic basis of hybrid inviability in the grasshopper *Podisma Pedestris*. *Heredity* **47**, 367-383.

Barton NH, Hewitt GM (1985) Analysis of hybrid zones. *Annual Review of Ecology and Systematics* **16**, 113-148.

Bierne N, Daguin C, Bonhomme F, David P, Borsa P (2003) Direct selection on allozymes is not required to explain heterogeneity among marker loci across a *Mytilus* hybrid zone. *Molecular Ecology* **12**, 2505-2510.

Broyles SB (2002) Hybrid bridges to gene flow: A case study in milkweeds (*Asclepias*). *Evolution* **56**, 1943-1953.

Chapman RE, Wang J, Bourke AFG (2003) Genetic analysis of spatial foraging patterns and resource sharing in bumble bee pollinators. *Molecular Ecology* **12**, 2801-2808.

Cicero C, Johnson NK (1998) Molecular phylogeny and ecological diversification in a clade of New World songbirds (genus *Vireo*). *Molecular Ecology* **7**, 1359-1370.

Coyne JA, Orr HA (2004) *Speciation*, Sunderland.

Cruzan MB, Arnold ML (1994) Assortative mating and natural selection in an iris hybrid zone. *Evolution* **48**, 1946-1958.

Currat M, Ruedi M, Petit RJ, Excoffier L (2008) The hidden side of invasions: Massive introgression by local genes. *Evolution* **62**, 1908-1920.

Darvill B, Knight ME, Goulson D (2004) Use of genetic markers to quantify bumblebee foraging range and nest density. *Oikos* **107**, 471-478.

Demesure B, Sodzi N, Petit RJ (1995) A set of universal primers for amplification of polymorphic noncoding regions of mitochondrial and chloroplast DNA in plants. *Molecular Ecology* **4**, 129-131.

Edwards CA, Skibinski DOF (1987) Genetic variation of mitochondrial DNA in mussel (*Mytilus edulis* and *Mytilus* galloprovincialis) populations from south west England and South-Wales. *Marine Biology* **94**, 547-556.

Edwards SV, Kingan SB, Calkins JD, *et al.* (2005) Speciation in birds: genes, geography, and sexual selection. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 6550-6557.

Endler JA (1977) *Geographic variation, speciation and clines*, Princeton.

Harrison RG (1993) *Hybrid zones and the evolutionary process* Oxford University Press, Oxford.

Hedrick PW (1983) Recombinaison and directional selection. *Nature* **302**, 727-727.

Hubbard JP (1969) The relationships and evolution of the *Dendroica coronata* complex. *Auk* **86**, 393-432.

Khimoun A, Burrus M, Andalo C, *et al.* (2011) Locally asymmetric introgressions between subspecies suggest circular range expansion at the *Antirrhinum majus* global scale. *Journal of Evolutionary Biology* **24**, 1433-1441.

Khimoun A, Cornuault J, Burrus M, *et al.* (Accepted) Ecology predicts parapatric distributions in two closely related *Antirrhinum majus* subspecies. *Evolutionary Ecology*.

Levin DA (1971) The origin of reproductive isolating mechanisms in flowering plants. *Taxon* **20**, 91-113.

McDonald JH (1994) Detecting neutral selection by comparing geographic variation in protein and DNA polymorphisms. In: *Non-neutral Evolution: Theories and Molecular Data* (ed. Golding B), pp. 88-100. Chapman and Hall, New York.

Molitor A (1937) Zur vergleichenden Psychobiologie der akuleaten Hymenopteren auf experimenteller Grundlage. *Biologia Generalis* **13**, 294-333.

Moore WS (1977) Evaluation of narrow hybrid zones in vertebrates. *Quarterly Review of Biology* **52**, 263-277.

Moore WS, Price JT (1993) The nature of selection in the Northern Flicker hybrid zone and its implications for speciation theory. In: *Hybrid zones and the evolutionary process* (ed. Harrison RG). Oxford University Press, Oxford.

Petit RJ, Bodenes C, Ducousso A, Roussel G, Kremer A (2003) Hybridization as a mechanism of invasion in oaks. *New Phytologist* **161**, 151-164.

Price TD (2008) *Speciation in birds*, Greenwood Village.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-959.

Schwinn K, Venail J, Shang YJ, *et al.* (2006) A small family of MYB-regulatory genes controls floral pigmentation intensity and patterning in the genus *Antirrhinum*. *Plant Cell* **18**, 831-851.

Shuker DM, Underwood K, King TM, Butlin RK (2005) Patterns of male sterility in a grasshopper hybrid zone imply accumulation of hybrid incompatibilities without selection. *Proceedings of the Royal Society B-Biological Sciences* **272**, 2491-2497.

Tastard E, Andalo C, Giurfa M, Burrus M, Thebaud C (2008) Flower colour variation across a hybrid zone in *Antirrhinum* as perceived by bumblebee pollinators. *Arthropod-Plant Interactions* **2**, 237-246.

Tastard E, Ferdy JB, Burrus M, Thebaud C, Andalo C (2011) Patterns of floral colour neighbourhood and their effects on female reproductive success in an *Antirrhinum* hybrid zone. *Journal of Evolutionary Biology* **25**, 388-399.

Tiffin P, Olson MS, Moyle LC (2001) Asymmetrical crossing barriers in angiosperms. *Proceedings of the Royal Society of London Series B-Biological Sciences* **268**, 861-867.

Weir BS, Wilson SR (1986) Log-linear models for linked loci. *Biometrics* **42**, 665-670.

Whibley AC (2004) *Molecular and genetic variation underlying the evolution of flower colour in Antirrhinum* University of East Anglia

Whibley AC, Langlade NB, Andalo C, *et al.* (2006) Evolutionary paths underlying flower color variation in *Antirrhinum*. *Science* **313**, 963-966.

List of Tables

Table 1. Geographic location and altitude from sea level for the *A. majus* populations sampled along the hybrid zone during the years 2002 and 2010.

n = sample size.

Population	Year	n	Latitude	Longitude	Altitude
HZ2002_01	2002	46	42.3600	1.9267	1492
HZ2002_02	2002	14	42.3297	2.0550	1515
HZ2002_03	2002	20	42.3246	2.0598	1259
HZ2002_04	2002	30	42.3265	2.0685	1418
HZ2002_05	2002	20	42.3308	2.0705	1448
HZ2002_06	2002	13	42.3323	2.0714	1379
HZ2002_07	2002	8	42.3260	2.0750	1382
HZ2002_08	2002	14	42.3274	2.0768	1370
HZ2002_09	2002	67	42.3263	2.0778	1368
HZ2002_10	2002	12	42.3254	2.0804	1353
HZ2002_11	2002	69	42.3225	2.0840	1232
HZ2002_12	2002	90	42.3142	2.1386	1111
HZ2010_01	2010	29	42.3219	2.0710	1155
HZ2010_02	2010	20	42.3203	2.0970	1207
HZ2010_03	2010	19	42.3179	2.1262	1207
HZ2010_04	2010	25	42.3127	2.1312	1062
HZ2010_05	2010	14	42.3181	2.1313	1301
HZ2010_06	2010	20	42.3119	2.1354	1034
HZ2010_07	2010	16	42.3163	2.1374	1249
HZ2010_08	2010	11	42.3150	2.1392	1190
HZ2010_09	2010	20	42.3116	2.1413	1019
HZ2010_10	2010	21	42.3123	2.1487	984

Table 2: Parameter estimates obtained for the cline-model fitting of the data obtained in the years 2002 and 2010. The dataset for 2002 included too few samples for the chloroplast transition area and therefore estimated confidence intervals are extremely large.

		parameter a {95% CI}		parameter b {95% CI}		Cline center
ROS1	**2002**	6.523	{7.49914;13.5113}	-0.426	{-0.867196;-0.494592}	15.429
	2010	15.667	{7.341;26.2012}	-1.055	{-1.7621;-0.52189}	14.686
Chloroplast	**2002**	91.206	{-14861.6;15058.3}	-4.186	{-690.401;681.309}	21.637
	2010	23.422	{12.057;21.8556}	-1.265	{-1.15598;-0.658064}	18.695

List of Figures

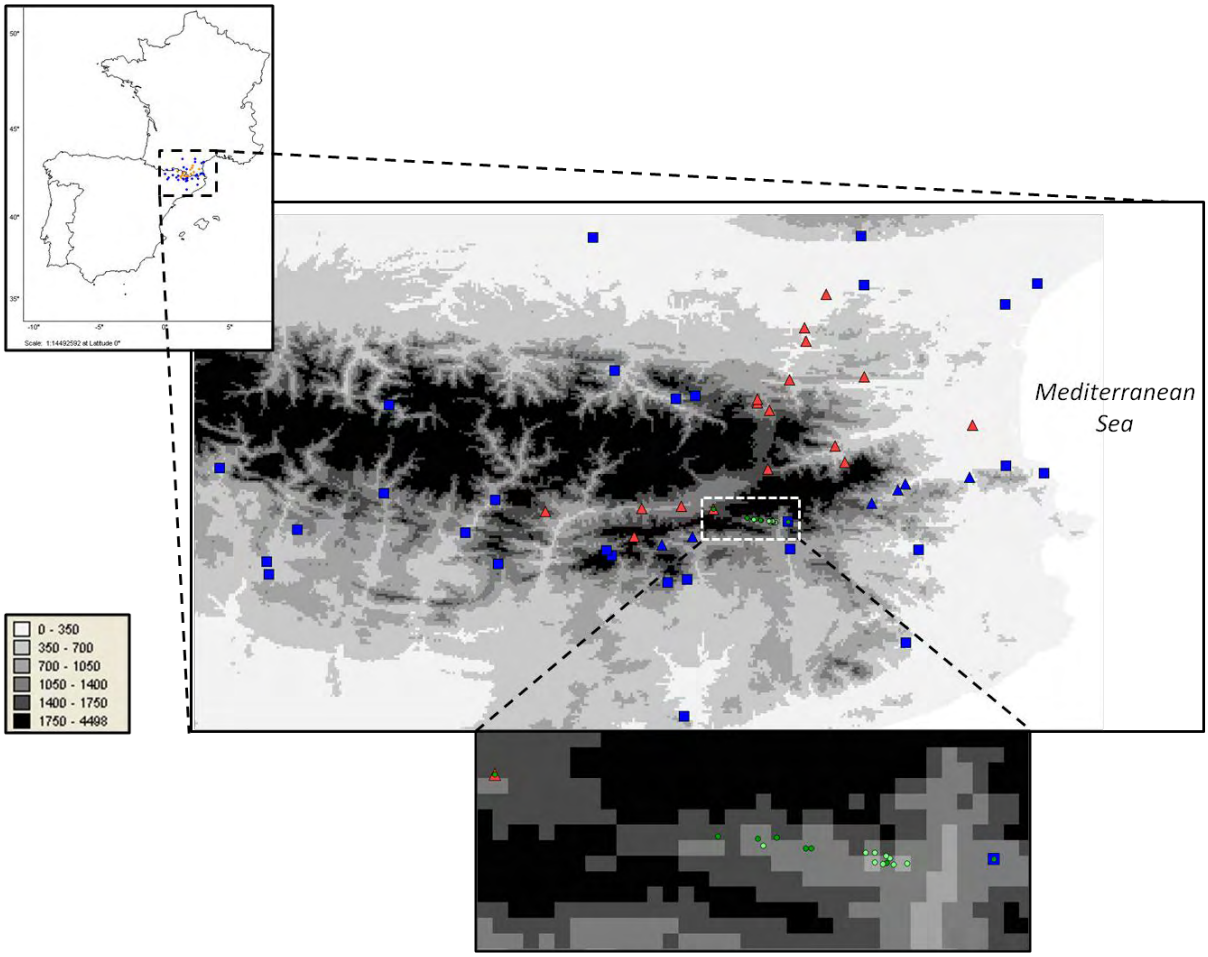


Figure 1: Geographic distribution of *A. majus pseudomajus* and *A. m. striatum* populations and the hybrid zone

A.m. pseudomajus and *A. m. striatum* populations are represented in blue and orange symbols respectively. Squares and triangle represent chloroplast Haplotype I and Haplotype II respectively. Hybrid samples from 2002 are represented in dark green and from 2010 in light green.

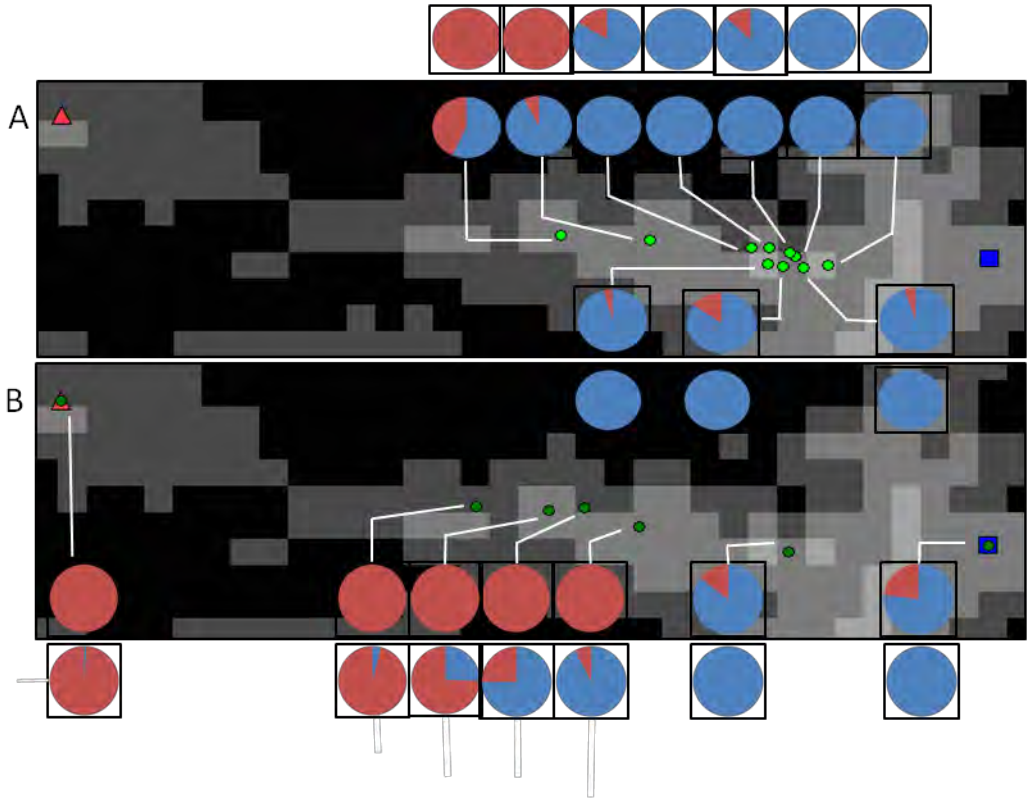


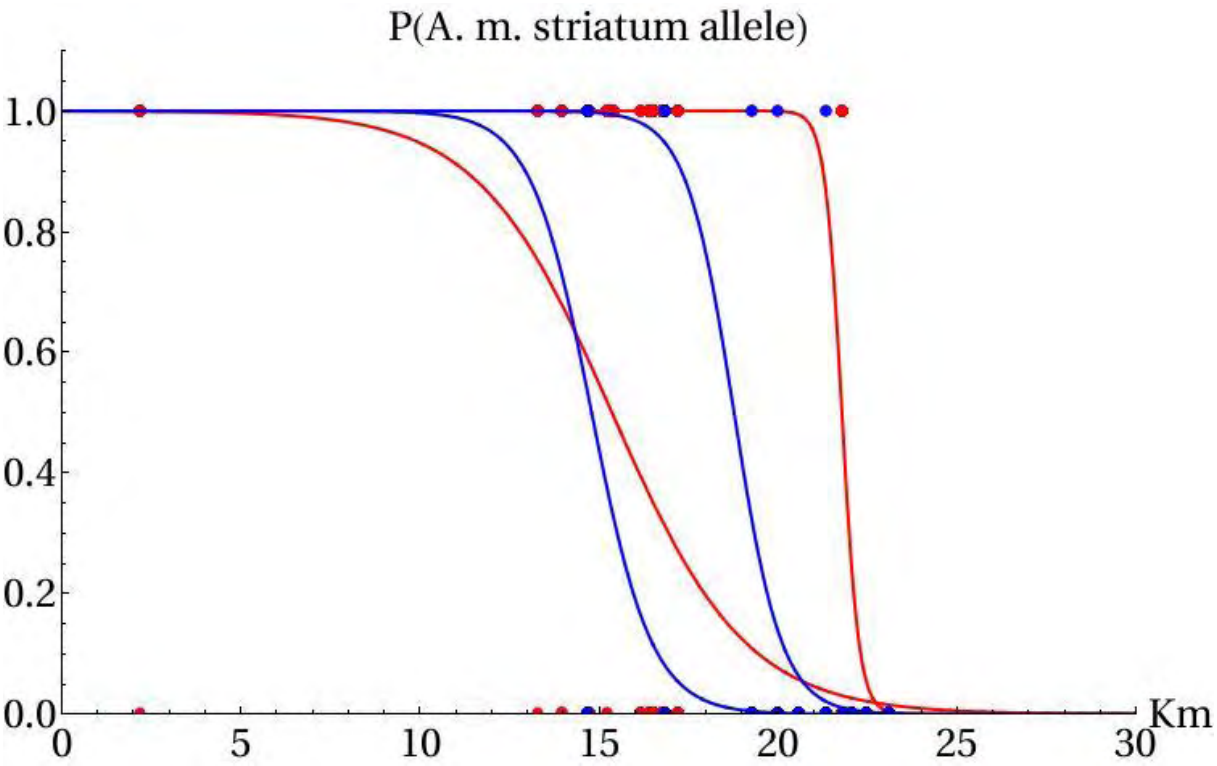
Figure 2: Distribution of chloroplast haplotypes and *ROS1* alleles through the hybrid zone

A/ 2010 sampling locations in light green

B/ 2002 sampling locations in dark green

For each sampled locality, the upper pie chart represents the frequency of chloroplast haplotypes, Haplotype I are represented in blue and Haplotype II are represented in orange. The lower pie chart represents the frequency of *ROS1* alleles, *ROS1-M* alleles are represented in blue and *ROS1-Y* alleles are represented in orange.

Figure 3: Cline fitting for typical *A. m. striatum* chloroplast haplotype and *ROS1-Y* alleles through the hybrid zone



Clines in 2002 are represented in red and clines in 2010 are represented in blue. The two clines that are on the left side are *ROS1-Y* allele and the two clines that are on the right side are chloroplast haplotype clines. Note that the clines for both the chloroplast and the *ROS1-Y* allele in 2002 (in red) are not so well defined due to heterogeneity among sampling locations. In both years, clines at different loci were non-overlapping.

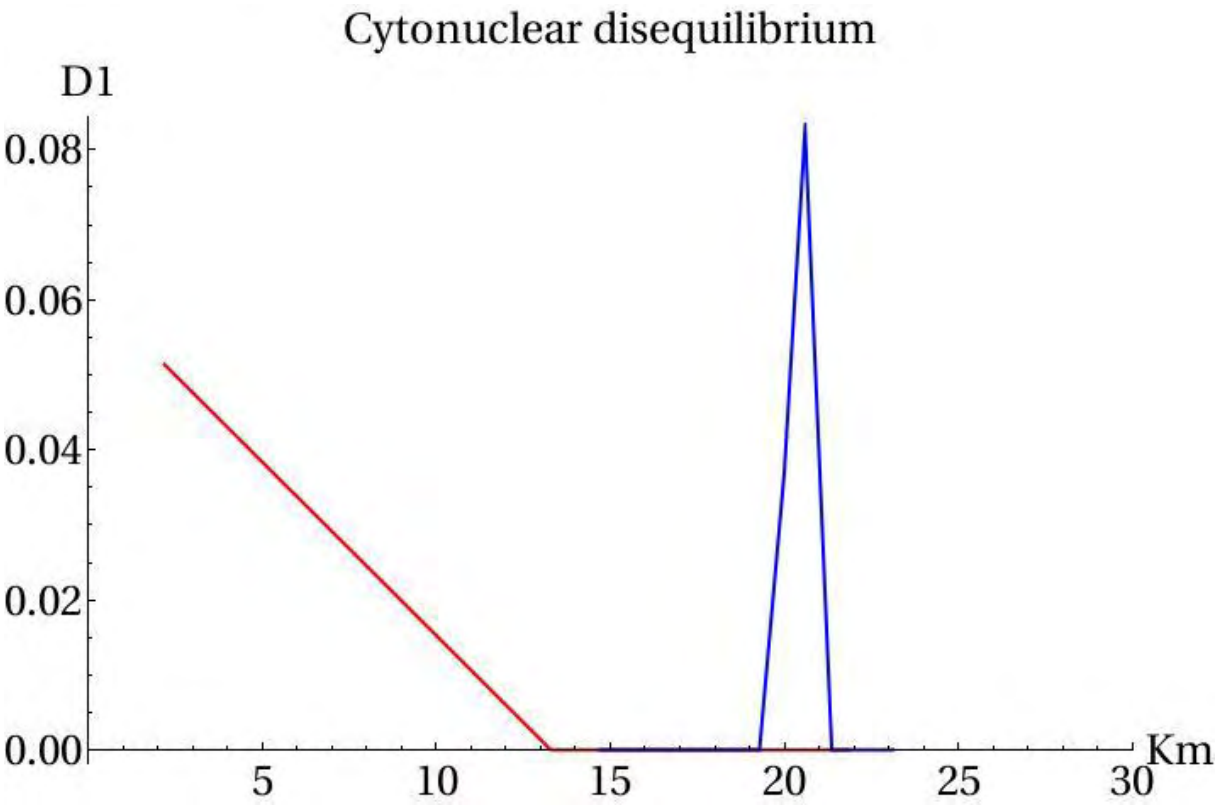


Figure 4: Distribution of the cytonuclear disequilibrium along the hybrid zone during the years 2002 (in red) and 2010 (in blue). It can be seen that the value of D1 shows a peak (D1 = 0.08) around an area 20 km apart from the extreme *A. m. striatum* populations. Again, the value obtained for the year 2002 is not so reliable due to the extreme values for the chloroplast cline.

Synthèse générale

L'objectif de cette thèse était d'étudier les différents processus impliqués dans la divergence en cours de deux sous-espèces de muflier, *Antirrhinum majus pseudomajus* et *A. m. striatum,* et d'apporter des éléments de reconstruction de leur « route de spéciation ». Afin de reconstruire cette route évolutive, nous avons replacé le processus de spéciation dans un cadre dynamique. Nous avons étudié les mécanismes impliqués, comment ils interagissent et à quelle(s) phase(s) du processus ils sont intervenus.

Afin de satisfaire aux mieux ces objectifs, notre démarche a combiné une approche de génétique des populations et une approche d'écologie à différentes échelles spatiale : de i) l'échelle globale de l'aire de distribution *d'A. majus* à ii) l'échelle fine d'une zone hybride particulière, en passant par iii) l'échelle des différentes zone de contact entre les deux sous-espèces. En guise de synthèse, nous combinerons les résultats des différents chapitres pour apporter des éléments de réponse quand à i) l'histoire de la colonisation de l'aire de distribution actuelle d'*Antirrhinum majus* et l'impact potentiel des dernières glaciations, ii) le contexte géographique et le rôle de l'écologie dans l'initiation de la divergence phénotypique entre les deux sous-espèces, iii) le rôle potentiel de l'écologie et des mécanismes de renforcement dans le maintien de la distribution parapatrique des deux sous-espèces et enfin, nous discuterons iv) de quelques éléments préliminaires sur la génétique de la divergence phénotypique entre les deux sous-espèces.

Quand « l'habit ne fait pas le moine »

Structuration génétique cryptique des populations *d'A. majus* principalement indépendante des deux sous-espèces

L'étude de la diversité génétique neutre (microsatellites nucléaires) à l'échelle globale de l'aire de distribution d'*Antirrhinum majus* a permis de révéler l'influence des processus historiques et démographiques sur la distribution actuelle des populations (chapitre 1). Les résultats, obtenus à partir de différentes méthodes, ont mis en évidence une structuration de la diversité génétique en deux groupes distincts. Ces deux groupes sont distribués indépendamment du critère taxonomique majeur d'identification des deux sous-espèces qui est la couleur des fleurs (Rothmaler, 1956; Sutton, 1988). Le premier groupe génétique est constitué principalement des populations d'*A. m. pseudomajus* et *A. m .striatum* situées au nord-est de l'aire de distribution d'*A. majus,* alors que les populations du sud-ouest constituent le deuxième groupe génétique. Cette structuration nord/sud est concordante avec deux des principaux scénarios de colonisation post-glaciaire connus à ce jour. Selon ces deux scénarios, les Pyrénées ont agit en tant que barrière physique à la colonisation post-glaciaire de certains lignages à partir des refuges glaciaires. Le premier scénario correspond à la route de colonisation inférée chez le criquet des pâtures, *Chorthippus parallelus* (Butlin, 1998; Hewitt, 1993). L'étude de la variation génétique neutre corrobore l'existence d'un premier refuge glaciaire dans les Balkans et un deuxième en Espagne, à partir desquels les populations auraient recolonisé l'Europe et seraient entrées en contact secondaire dans les Pyrénées. Dans ce scénario, les Pyrénées ont joué le rôle de barrière à la recolonisation du sud de l'Europe par les populations originaires des Balkans et à la recolonisation du nord de l'Europe par les populations originaires d'Espagne. Le deuxième scénario de colonisation qui pourrait expliquer une telle structuration nord/sud des populations correspond à celui mis en évidence chez le hérisson *Erinaceus europaeus* qui a re-colonisé le nord de l'Europe à partir d'un refuge Ibérique, en contournant la chaine principale des Pyrénées par l'est et l'ouest où les altitudes sont moindres (Seddon *et al.*, 2001). La plausibilité de ces deux scénarios alternatifs a été testée chez *Antirrhinum*

majus mais aucun des deux scenarios n'est apparu moins probable que l'autre sur la base de la variabilité génétique des populations dont nous disposions.

Les scenarios criquet et hérisson peuvent potentiellement expliquer la différenciation génétique des populations observée chez *A. majus*. Le premier implique la colonisation suivant une route est-ouest de l'aire de distribution d'*A.majus* à partir d'un seul refuge glaciaire en Espagne. Il est ici important de noter que le climat à l'est de la chaine Pyrénéenne est soumis à une influence Méditerranéenne qui crée, à l'est, des conditions climatiques très différentes de celles retrouvées à l'ouest de la chaine Pyrénéenne. Cette hétérogénéité climatique le long d'un gradient ouest/est est souvent invoquée pour expliquer les glaciations plus tardives et de moindre importance à l'est des Pyrénées en comparaison avec la partie ouest. Pour cette raison, et parce qu'aucune populations d'*A. majus* n'est trouvée dans l'ouest des Pyrénées, la colonisation d'*A. majus* vers le nord après le contournement des Pyrénées n'aurait vraisemblablement eu lieu que par l'est (Figure 11 A). A partir de leur refuge glaciaire, les populations d'*A.majus* auraient alors progressé vers le nord tout en colonisant, vers l'ouest, les Pyrénées centrales. Les deux groupes génétiques se seraient ensuite différenciés par vicariance. L'avancée des glaciers durant la période glaciaire a pu isoler physiquement les populations distribuées sur les versants nord (Pyrénées françaises) et sud des Pyrénées (Pyrénées espagnoles). Alors spatialement isolées et donc reproductivement isolées, les populations auraient évolué en accumulant des différences génétiques. Le deuxième scénario pouvant expliquer la distribution de la diversité neutre d'*A. majus* est celui d'une double colonisation à partir de deux refuges différents (Figure 11 B). La colonisation du nord de l'aire de distribution d'*A. majus* aurait eu lieu à partir d'un refuge balkanique alors que le versant le sud ouest de l'aire de distribution d'*A. majus* aurait une origine espagnole.

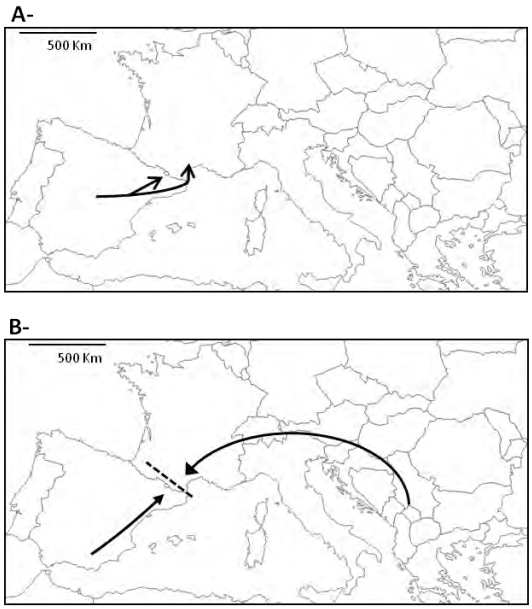


Figure 11 : Scenarios possibles de route de colonisation d'*A. majus*

La Figure A représente la colonisation de l'aire de distribution d'*A. majus* à partir d'un seul refuge alors que la Figure B représente un évènement de double colonisations indépendantes de l'aire de distribution d'*A. majus* à partir de refuges différents.

Cette étude a révélé des patrons discordants de structuration entre les marqueurs neutres et le gène *ROSEA*. Les marqueurs neutres reflètent généralement des processus historiques et démographiques alors que les gènes codant pour des traits phénotypiques, comme c'est le cas pour *ROSEA* peuvent refléter également une réponse microévolutive à des pressions de sélection engendrées par des facteurs écologiques. De tels patrons discordants de structuration sont souvent observés lorsque l'origine de la divergence phénotypique est plus récente et adaptative (Mila *et al.*). Des travaux précédents ont suggéré le rôle adaptatif de *ROSEA* (Whibley *et al.*, 2006), via le comportement des pollinisateurs (Tastard *et al.*, 2008; Tastard *et al.*, 2011). Parce qu'elle met en évidence des histoires évolutives différentes, la comparaison de la structuration génétique des microsatellites et de *ROSEA* renforce l'hypothèse du rôle adaptatif de la couleur des fleurs.

Origine de la divergence phénotypique des deux sous-espèces : contexte spatial et écologique des populations

L'origine de la structuration génétique observée entre les deux sous-espèces a été estimée à une date plus récente que celle observée entre le nord-est et le sud-ouest. La reconstruction des niches environnementales des deux sous-espèces a mis en évidence un patron de divergence de niche qui peut potentiellement être impliqué dans la divergence des deux sous-espèces (Chapitre 3). A l'échelle globale, la distribution spatiale des populations des deux sous-espèces est expliquée par une combinaison de facteurs environnementaux climatiques et édaphiques. Deux scenarios de divergence écologique peuvent expliquer ces résultats. Dans le premier, les deux sous-espèces ont été isolées reproductivement par des barrières physiques et se sont adaptées à l'environnement dans lequel chacune se trouvait. Dans un tel scénario, la divergence des deux sous-espèces et leur différenciation écologique sont les sous-produits de leur isolement géographique. Le second scenario met en scène la divergence des deux sous-espèces en l'absence de barrière géographique. Suivant ce scénario, les deux sous-espèces auraient subi des pressions de sélection divergente qui les ont conduites à s'adapter à des environnements différents, à se différencier phénotypiquement et, dans un second temps, à occuper des régions géographiques différentes lors de l'expansion de leur aire de distribution. Dans ce scenario, la divergence phénotypique des deux sous-espèces est le sous-produit de leur adaptation à des environnements différents. Ces deux scenarios offrent des visions différentes du rôle de l'écologie et du contexte spatial des populations dans la divergence des deux sous-espèces. Dans le premier scenario, l'isolement spatial des populations est à l'origine de leur adaptation à des environnements différents et à leur divergence, alors que dans le deuxième scenario, c'est l'adaptation des deux sous-espèces à des conditions environnementales différentes qui a entrainé leur isolement spatial et reproducteur. Qu'elle soit à l'origine de leur divergence phénotypique ou qu'elle ne soit qu'une conséquence de leur isolement

géographique, l'adaptation des deux sous-espèces à des environnements différents semble avoir joué un rôle dans leur divergence phénotypique.

Chez les plantes, les traits floraux tels que la couleur des fleurs peuvent être influencés par des facteurs abiotiques. Chez de nombreuses espèces, les individus dont les parties florales ou végétatives contiennent des anthocyanes (pigment entrainant une couleur violette ou rose dans les tissus où il est exprimé) ont de manière générale une meilleure tolérance au stress hydrique et thermique que les individus qui n'en contiennent pas (généralement de couleur blanche ou jaune). Il a été montré par sélection artificielle que les individus à fleurs violettes, rouge ou roses contenant des anthocyanes, ont une valeur sélective supérieure aux individus blancs ou jaunes dans des environnements secs (e.g. *Holcus lanatus*, *Polygonum persicaria* et *Vicia sepium*, (Warren, Mackenzie, 2001)) ou face à de fortes températures (e.g. *Ipomoea purpurea*, (Coberly, Rausher, 2003)). Ainsi la présence d'*A. m. pseudomajus* dans des conditions de températures plus élevées et de plus faibles précipitations comparées à *A. m. striatum* pourrait être associée à une meilleure survie due aux anthocyanes qu'elle contient.

Bien que les deux sous-espèces aient des niches environnementales différentes, les facteurs environnementaux, seuls, ne permettent pas d'expliquer le maintien de leur distribution parapatrique. En effet, l'isolement écologique entre les deux sous-espèces n'est pas complet, nous avons montré que les facteurs environnementaux sont favorables à l'établissement des deux sous-espèces dans les zones de contact. Alors, comment expliquer que les deux sous espèces ne sont jamais trouvées en sympatrie malgré des habitats qui leurs sont favorables dans les zones de contact ?

Rôle des processus écologiques dans le maintien de la distribution parapatrique des deux sous-espèces

L'un des facteurs couramment invoqué pour expliquer pourquoi une espèce n'occupe pas la totalité de sa niche environnementale potentielle est la limitation par la dispersion. Les deux sous-espèces pourraient être isolées géographiquement dans les zones de contact simplement parce qu'elles n'ont pas encore eu le temps d'atteindre ces régions du fait de leur capacité limité de dispersion, maintenant ainsi leur distribution parapatrique. Pour tester cette hypothèse nous nous sommes focalisés sur la zone de contact entre les deux sous-espèces au sein du cluster génétique sud, pour laquelle nous avons l'échantillonnage le plus important. L'étude conjointe de la structuration spatiale d'un marqueur chloroplastique et du gène nucléaire de la couleur des fleurs, ROS1, a révélé des traces d'introgression et donc de flux de gènes entre les deux sous-espèces (Chapitre 2). Ceci atteste de l'absence d'isolement reproducteur complet entre les deux sous-espèces et la présence des deux sous espèces à un moment donné dans ces zones de contact. Bien que la dispersion limité puisse être une hypothèse pour expliquer la l'isolement spatial des deux sous-espèces à l'échelle de leur aires de distribution, elle ne permet pas d'expliquer le maintien de leur distribution parapatrique (*i.e.* isolement géographique partiel) dans les zones de contact malgré leur isolement écologique incomplet.

L'étude de différentes zones de contact a mis en évidence des évènements d'introgression asymétriques entre les deux sous-espèces et dont la direction est opposée entre la partie ouest et est de la zone de contact (Chapitre 2). Dans la région ouest, les populations d'*A. m. striatum* possèdent l'haplotype chloroplastique associé aux populations d'*A. m. pseudomajus* alors que dans les régions centre et est de la zone de contact, les populations allopatriques d'*A.m.pseudomajus* possèdent l'haplotype chloroplastique associé aux populations d'*A. m. striatum*. Bien qu'ils soient structurés spatialement, ces deux types d'introgression sont trouvés à l'échelle de la zone de

contact globale entre les deux sous espèces. Ces résultats confirment l'absence de barrière intrinsèque prézygotique (e.g. incompatibilités asymétriques entre pollen et style (Cruzan, Arnold, 1994) ou postzygotique (stérilité partielle des hybrides (Shuker *et al.*, 2005) mise en évidence par (Andalo *et al.*, 2010). Entre les deux sous-espèces étudiées, il n'y a probablement pas d'incompatibilité cytonuléaire intrinsèque (Levin, 1971; Tiffin *et al.*, 2001).

Les patrons locaux d'introgression asymétrique peuvent être dus à des flux de gènes unidirectionnels récurrents. La présence de barrière physique, telles que la présence de hauts reliefs, peut contraindre la direction du flux de gène. Nous avons testé si la localisation des populations dans différentes vallées pouvait contraindre la direction des flux de gènes et entrainer leur asymétrie dans les zones de contact. Nos résultats illustrent que la structuration génétique neutre des populations n'est pas expliquée par leur localisation dans les différentes vallées. Des populations de deux vallées différentes semblent partager plus de gènes entre elles que certaines populations au sein de la même vallée. Si la topologie ne semble pas avoir d'influence sur les flux de gène entre populations, le rôle potentiel d'autres facteurs physiques ne peut pas être écarté. Ainsi, pour poursuivre dans cette direction, il serait intéressant de tester le rôle d'autres facteurs physiques, tels que la direction des vents dominants qui pourrait avoir un impact plus direct sur l'orientation du vol des pollinisateurs et de la dissémination des graines.

Enfin, les patrons locaux d'introgression asymétrique peuvent être le résultat de l'adaptation locale malgré un flux de gène bidirectionnel. Sous cette hypothèse, les associations inter-sous-espèces entre haplotype chloroplastique et allèle de *ROSEA* auraient un avantage sélectif différent dans l'environnement réciproque et natif de chaque sous-espèce. Nous nous attendrions dans ce cas à trouver les combinaisons génétiques différentes dans des localités caractérisées par des environnements différents. Cet attendu a été vérifié par la comparaison des niches réalisées des deux sous-espèces dans les zones de contact. La présence de l'une ou de l'autre sous-espèce était corrélée à des conditions environnementales particulières. L'association de fortes précipitations avec des faibles températures, une forte amplitude thermique et des sols pauvres en nutriments avec une forte disponibilité en eau augmentait significativement la probabilité d'observer *A. m. striatum* plutôt qu'*A. m. pseudomajus* dans les zones de contact. L'adaptation locale à des conditions environnementales différentes pourrait

donc potentiellement expliquer l'asymétrie locale du remplacement d'une sous-espèce par l'autre mais des études plus directes d'un tel phénomène seraient nécessaires pour le confirmer ou l'infirmer. La distribution parapatrique des deux sous-espèces pourrait être maintenue, malgré le flux de gènes, par des capacités de survie ou de reproduction variables des deux sous-espèces dans ces différents environnements. Des capacités supérieures à utiliser les ressources en eau ou en nutriments dans les sols pourraient être un exemple d'adaptation différentielle des deux sous-espèces dans des environnements contrastés.

Il est intéressant de noter que la niche réalisée de chaque sous-espèce dans les zones de contact comprend une gamme plus restreinte de conditions environnementales qui semble déplacée par rapport à la gamme de conditions environnementales que chaque sous-espèce occupe lorsque leurs populations sont en allopatrie (i.e. absence d'interaction). Même si nous n'en avons pas de preuve directe, le déplacement de niche des deux sous-espèces lorsqu'elles sont en contact suggère que la compétition entre les deux sous-espèces pourrait jouer un rôle important dans le maintien de leur isolement géographique (Servedio, Noor, 2003). Une analyse directe de la capacité fonctionnelle des deux sous-espèces à exploiter différentes ressources dans différents environnements pourrait éclaircir ce point.

Génétique de l'adaptation locale des deux sous-espèces : éléments préliminaires et futures investigations

L'analyse de la variation des fréquences alléliques du gène nucléaire *ROSEA* (responsable de la synthèse des anthocyanes et donc de la couleur des fleurs) le long de la zone hybride étudiée depuis plus d'une dizaine d'année dans les Pyrénées a montré la stabilité temporelle du cline abrupt de la fréquence des allèles de *ROSEA*. Nos résultats confirment donc l'hypothèse de sélection sur *ROSEA* émise par Whibley et al (2006). D'un point de vue mécanistique, la couleur des fleurs fournit des indices visuels aux polinisateurs et stimule leur système sensoriel, les attirant ainsi de manière sélective. Les variations de couleur des fleurs entre *A. m. pseudomajus*, *A. m. striatum* et leurs hybrides sont discriminées par les pollinisateurs qui peuvent ensuite exercer un choix sur la fleur à visiter (Tastard *et al.*, 2008). Des études menées à l'échelle d'une zone hybride de quelques kilomètres à l'intérieur de laquelle une grande diversité phénotypique est maintenue (Figure 10), ont montré que les phénotypes hybrides étaient moins visités que les phénotypes parentaux (Tastard *et al.*, 2011). Les pollinisateurs incriminés semblent avoir une préférence innée pour les phénotypes jaunes et ils exercent un choix olfactif vis-à-vis des composés volatils émis par les deux sous-espèces (Suchet *et al.*, 2011). Le choix des polinisateurs est influencé par l'acétophénone qui est un composé de la famille des benzénoïdes émis exclusivement par *A. m. pseudomajus*. Des tests de choix sur les bourdons ont montré que ce composé avait un effet aversif sur les bourdons qui choisissent principalement l'odeur alternative proposée (Suchet *et al.*, 2011). Les voix de biosynthèse des anthocyanes et de l'acétophénone sont liées, ce qui peut entrainer une sélection indirecte sur les anthocyanes via la sélection exercée à l'encontre de l'acétophénone si toutefois son effet se confirme dans le long terme de l'interaction avec les pollinisateurs. Néanmoins, les pollinisateurs ne sont pas les seuls agents biotiques interagissant avec *A. majus* et le charançon prédateur de graines *Rhinusa vestita* ainsi que le papillon *Melitea deione* dont les chenilles consomment les feuilles pourraient bien influencer l'équilibre des pressions

de sélections liés à la chaine de biosynthèse des anthocyanes et de l'acétophénone. Ainsi, une piste intéressante à poursuivre serait de mesurer directement le caractère adaptatif des anthocyanes florales ou végétatives dans la reproduction et le cout de la prédation chez les deux sous-espèces.

L'analyse de la variation des fréquences d'haplotypes chloroplastiques le long de la zone hybride a mis en évidence un cline abrupt de même largeur que celui observé pour *ROSEA*. Ce résultat suggère que la sélection modèle également la diversité des gènes chloroplastiques. La similarité des patrons de variation abrupts des allèles de *ROSEA* et des haplotypes chloroplastiques pourrait potentiellement s'expliquer par une relation épistatique entre génome et plastome. L'incompatibilité nucleo-plastidique chez les hybrides est un phénomène très répandu chez les plantes qui peut jouer un rôle dans la spéciation (Greiner *et al.*, 2012). Ce type d'incompatibilités se comporte de manière analogue aux incompatibilités nucléaires du modèle de Dobzhansky-Muller sauf qu'elles impliquent dans ce cas la co-évolution entre loci nucléaire et chloroplastique (Bomblies, Weigel, 2007; Burke, Arnold, 2001; Presgraves, 2010; Turelli, Orr, 2000). En causant principalement la stérilité de certains types d'hybrides, ces incompatibilités entrainent généralement des barrières postzygotiques asymétriques. Elles peuvent aussi entrainer des effets plus subtils sur la survie des hybrides. Par exemple, des performances photosynthétiques différentes ont été observées entre individus introgressés avec des lignées cytoplasmiques différentes (Glick, Sears, 1994; Iwanaga *et al.*, 1978; Wu, Campbell, 2007). Pour laisser libre cours à la spéculation, il se pourrait que la photosynthèse qui est très fortement influencée par des facteurs abiotiques tels que la disponibilité en eau, la luminosité ou les températures entraine un avantage adaptatif de certaines combinaisons nucléo-plastidiques chez les hybrides dans certaines conditions abiotiques locales. Quoi qu'il en soit, il semblerait intéressant d'étudier plus en détail la mise en place de barrières post-zygotiques asymétriques entre les deux sous espèces d'*A. majus* dépendantes de l'environnement. Ici encore, il serait nécessaire de mesurer directement la valeur sélective des différentes combinaisons d'haplotypes chloroplastique et d'allèles de *ROSEA* dans différents environnements abiotiques.

Conclusions et perspectives

L'ensemble des résultats de cette thèse nous a permis de proposer une hypothèse de route de spéciation en cours des deux sous-espèces, *Antirrhinum majus pseudomajus* et *Antirrhinum majus striatum.* Cette hypothèse établit l'adaptation à des niches environnementales différentes comme possible origine de la divergence des phénotypes floraux des deux sous-espèces, fleurs jaunes pour *A. m. striatum* et fleurs magenta pour *A. m. pseudomajus* (Figure 14, traits en tirets). Suivant cette hypothèse, l'adaptation des deux sous-espèces à des environnements biotiques et/ou abiotiques différents aurait conduit à l'isolement géographique et reproducteur des populations des deux sous-espèces. Néanmoins, l'isolement écologique et reproducteur n'étant pas complet, les deux sous-espèces ont échangé et échangent aujourd'hui encore, de manière récurrente, des gènes dans des zones de contacts (Figure 14, trait en pointillés). Le maintien de la parapatrie des deux sous-espèces et le maintien de leur intégrité phénotypique, malgré le flux de gènes, peut probablement être imputé à des mécanismes de renforcement faisant intervenir un réseau complexe d'interactions biotiques et abiotiques (Figure 14, trait plein).

En conclusion, la divergence des deux sous-espèces semble résulter de la combinaison dans le temps de processus historiques neutres et démographiques actuels avec des processus sélectifs associés à l'écologie des sous espèces qui modèlent non indépendamment leur diversité.

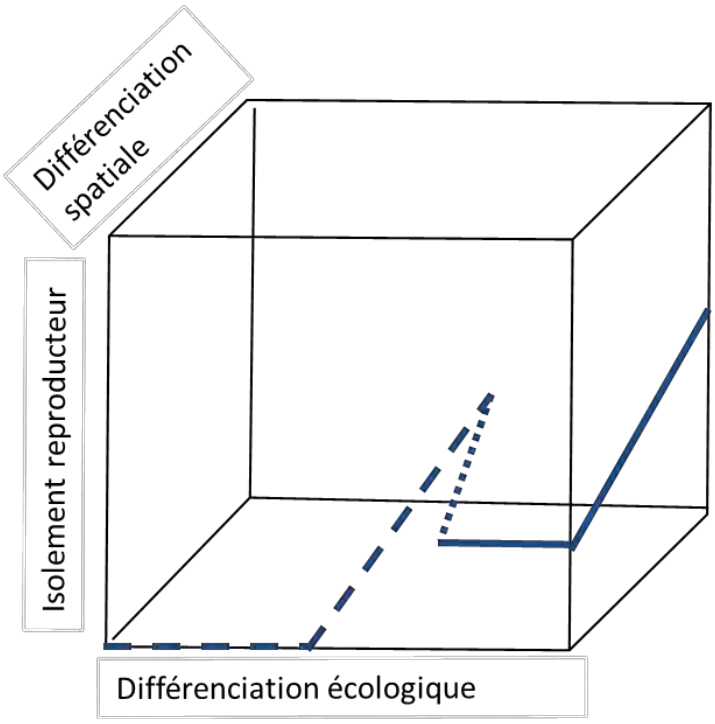


Figure 14 : Hypothèse de « route de spéciation » en cours chez *A. m. Pseudomajus et A. m. striatum*

Bien que les résultats de cette thèse soulignent le rôle prépondérant des facteurs écologiques et de leur interaction dans l'origine et le maintien des deux sous-espèces, nos études ont bien entendu des limites. Les niches environnementales des deux sous-espèces ont été étudiées par une méthode indirecte de modélisation de niche qui est une alternative aux expériences plus couramment utilisées de transplantation mais non envisageables chez *Antirrhinum* pour des raisons techniques. La divergence de niche à l'origine de la divergence des deux sous espèces ne reste toutefois qu'une hypothèse inférée à partir de résultats indirects. L'adaptation des deux sous-espèces à leur niche écologique demande à être testée de manière directe par des mesures en conditions naturelles ou expérimentales. De plus si les deux sous-espèces s'avéraient être actuellement adaptées à des facteurs environnementaux différents, ces mêmes facteurs ne sont pas nécessairement ceux qui sont historiquement à l'origine de leur divergence. Le paysage adaptatif des espèces est dynamique dans le temps. Ainsi, la lecture que nous en faisons à partir d'observations actuelles peut ne pas être fidèle à ce qu'il était au moment de la divergence des deux sous-espèces (Herrera, 1996). Une perspective intéressante pourrait être de projeter la niche environnementale potentielle actuelle des deux sous-espèces dans l'espace des conditions environnementales disponibles durant

la dernière période post-glaciaire afin de voir si les conditions actuellement favorables à l'établissement de chaque sous-espèce étaient présentes durant cette période.

Enfin, les résultats de cette thèse soutiennent le rôle adaptatif de la couleur des fleurs et plus particulièrement de la présence/absence des anthocyanes. Il semble maintenant nécessaire d'étudier plus en détail l'influence, directe (visuelle) ou indirecte (composés de défense), des anthocyanes sur le réseau d'interaction des espèces liées à la pollinisation et la prédation chez *Antirrhinum majus.* En accord avec nos résultats, il semble également pertinent de s'intéresser à la stabilité de ces interactions et de l'isolement tel qu'il est aujourd'hui entre les deux sous espèces. La prédiction des réponses du réseau d'interaction biotiques face aux changements environnementaux et climatiques pourrait permettre prédire l'évolution et le devenir des deux sous-espèces.

Références bibliographiques

Andalo C, Cruzan MB, Cazettes C*, et al.* (2010) Post-pollination barriers do not explain the persistence of two distinct *Antirrhinum* subspecies with parapatric distribution. *Plant Systematics and Evolution* **286**, 223-234.

Anderson E (1948) Hybridization of the habitat. *Evolution* **2**, 1-9.

Anderson E (1949) *Introgressive hybridization* Whiley & Sons, New York.

Armbruster WS, Edwards ME, Debevec EM (1994) Floral character displacement generates assemblage structure of western australian triggerplants (*Stylidium*). *Ecology* **75**, 315-329.

Avise JC (2000) *Some genetic consequences of ice ages, and their role in divergence and speciation*, Cambridge.

Barton NH, Charlesworth B (1984) Genetic revolutions, founder effects, and speciation. *Annual Review of Ecology and Systematics* **15**, 133-164.

Barton NH, Hewitt GM (1985) Analysis of hybrid zones. *Annual Review of Ecology and Systematics* **16**, 113-148.

Bateson W (1909) Heredity and variation in modern light In: *Darwin and modern science* (ed. Seward AC), pp. 85-101. Cambridge University Press, Cambridge.

Blount ZD, Borland CZ, Lenski RE (2008) Historical contingency and the evolution of a key innovation in an experimental population of Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 7899-7906.

Bolnick DI (2004) Can intraspecific competition drive disruptive selection? An experimental test in natural populations of sticklebacks. *Evolution* **58**, 608-618.

Bomblies K, Weigel D (2007) Hybrid necrosis: autoimmunity as a potential gene-flow barrier in plant species. *Nature Reviews Genetics* **8**, 382-393.

Boughman JW (2002) How sensory drive can promote speciation. *Trends in Ecology & Evolution* **17**, 571-577.

Brown JH, Stevens GC, Kaufman DM (1996) The geographic range: Size, shape, boundaries, and internal structure. *Annual Review of Ecology and Systematics* **27**, 597-623.

Burke JM, Arnold ML (2001) Genetics and the fitness of hybrids. *Annual Review of Genetics* **35**, 31-52.

Butlin RK (1998) What do hybrid zones in general, and the *Chorthippus parallelus* zone in particular, tell us about speciation? In: *Endless forms: species, speciation* (eds. Howard DJ, Berlocher SH), pp. 367–378. Oxford University Press, New York

Butlin RK, Galindo J, Grahame JW (2008) Sympatric, parapatric or allopatric: the most important way to classify speciation? *Philosophical Transactions of the Royal Society B-Biological Sciences* **363**, 2997-3007.

Cai L, Giraud T, Zhang N*, et al.* (2011) The evolution of species concepts and species recognition criteria in plant pathogenic fungi. *Fungal Diversity* **50**, 121-133.

Case AL, Willis JH (2008) Hybrid male sterility in *Mimulus* (phrymaceae) is associated with a geographically restricted mitochondrial rearrangement. *Evolution* **62**, 1026-1039.

Coberly LC, Rausher MD (2003) Analysis of a chalcone synthase mutant in *Ipomoea purpurea* reveals a novel function for flavonoids: amelioration of heat stress. *Molecular Ecology* **12**, 1113-1124.

Coyne JA, Orr HA (2004) *Speciation*, Sunderland.

Cruzan MB, Arnold ML (1994) Assortative mating and natural selection in an iris hybrid zone. *Evolution* **48**, 1946-1958.

Davis MB, Shaw RG (2001) Range shifts and adaptive responses to Quaternary climate change. *Science* **292**, 673-679.

de Queiroz K (1998) The general lineage concept of species, species criteria, and the process of speciation: a conceptual unification and terminological recommendations. In: *Endless forms: species and speciation* (eds. DJ H, SH B), pp. 57-75. Oxford University Press, New York.

de Queiroz K (2007) Species concepts and species delimitation. *Systematic Biology* **56**, 879-886.

Dieckmann U, Doebeli M, Metz JAJ, Tautz D (2004) *Adaptive speciation* Cambridge University Press, Cambridge.

Dobzhansky T (1937) Genetics and the origin of species. In: *Colombia Biological Series* (eds. Dunn LC, Clarke HT, Detwiler SR), pp. 1-364. Columbia University Press, New York.

Dobzhansky T (1951) *Genetics and the Origin of Species*, New York.

Dres M, Mallet J (2002) Host races in plant-feeding insects and their importance in sympatric speciation. *Philosophical Transactions of the Royal Society B-Biological Sciences* **357**, 471-492.

Emms SK, Arnold ML (2000) Site-to-site differences in pollinator visitation patterns in a Louisiana iris hybrid zone. *Oikos* **91**, 568-578.

Endler JA (1977) *Geographic variation, speciation and clines*, Princeton.

Endler JA (1992) SignaIs, signal conditions, and the direction of evolution. *American Naturalist* **139**, S125-S153.

Friesen ML, Saxer G, Travisano M, Doebeli M (2004) Experimental evidence for sympatric ecological diversification due to frequency-dependent competition in *Escherichia coli*. *Evolution* **58**, 245-260.

Funk DJ (1998) Isolating a role for natural selection in speciation: Host adaptation and sexual isolation in *Neochlamisus bebbianae* leaf beetles. *Evolution* **52**, 1744-1759.

Funk DJ, Filchak KE, Feder JL (2002) Herbivorous insects: model systems for the comparative study of speciation ecology. *Genetica* **116**, 251-267.

Futuyma DJ, Mayer GC (1980) on-allopatric speciation in animals. *Systematic Zoology* **29**, 254-271.

Gaston KJ (2003) *The structure and dynamics of geographic ranges*, Oxford.

Gavrilets S (2003) Perspective: Models of speciation: What have we learned in 40 years? *Evolution* **57**, 2197-2215.

Giraud T, Refregier G, Le Gac M, de Vienne DM, Hood ME (2008) Speciation in fungi. *Fungal Genetics and Biology* **45**, 791-802.

Glick RE, Sears BB (1994) Genetically programmed chloroplast dedifferentiation as a consequence of plastome-genome incompatibility in *Oenothera*. *Plant Physiology* **106**, 367-373.

Greiner S, Rauwolf U, Meurer J, Herrmann RG (2012) The role of plastids in plant speciation. *Molecular Ecology* **20**, 671-691.

Gubitz T, Caldwell A, Hudson A (2003) Rapid molecular evolution of CYCLOIDEA-like genes in *Antirrhinum* and its relatives. *Molecular Biology and Evolution* **20**, 1537-1544.

Guillaume CP, Heulin B, Arrayago MJ, Bea A, Brana F (2000) Refuge areas and suture zones in the Pyrenean and Cantabrian regions: geographic variation of the female MPI sex-linked alleles among oviparous populations of the lizard *Lacerta* (*Zootoca*) *vivipara*. *Ecography* **23**, 3-10.

Hall MC, Willis JH (2006) Divergent selection on flowering time contributes to local adaptation in *Mimulus guttatus* populations. *Evolution* **60**, 2466-2477.

Herrera CM (1996) Floral traits and plant adaptation to insect pollinators: a devil's advocate approach. In: *Floral biology: studies on floral evolution in animal-pollinated plants* (eds. Lloyd DG, Barrett CH). Chapman & Hall, New York.

Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature* **405**, 907-913.

Hewitt G (2003) Ice Ages: their impact on species distributions and evolution. In: *Evolution of planet earth* (eds. Rothschild LJ, Lister AM), pp. 339–361. Academic Press.

Hewitt GM (1993) After the ice - *parallelus* meets *erythropus* in the Pyrenees. In: *Hybrid zones and the evolutionary process* (ed. Harrison RG), pp. 140–164. Oxford University Press, Oxford.

Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society* **58**, 247-276.

Hewitt GM (1999) Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society* **68**, 87-112.

Hewitt GM (2001) Speciation, hybrid zones and phylogeography - or seeing genes in space and time. *Molecular Ecology* **10**, 537-549.

Hewitt GM (2004) Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **359**, 183-195.

Hey J (2006) On the failure of modern species concepts. *Trends in Ecology & Evolution* **21**, 447-450.

Iwanaga M, Mukai Y, Panayotov I, Tsunewaki K (1978) Genetic diversity of the cytoplasm in *Triticum* and *Aegilops*: VII. Cytoplasmic effects on respiratory and photosynthetic rates. *Japanese Journal of Genetics* **53**, 387-396.

Jiggins CD, Naisbit RE, Coe RL, Mallet J (2001) Reproductive isolation caused by colour pattern mimicry. *Nature* **411**, 302-305.

Johnson PA, Hoppensteadt FC, Smith JJ, Bush GL (1996) Conditions for sympatric speciation: A diploid model incorporating habitat fidelity and non-habitat assortative mating. *Evolutionary Ecology* **10**, 187-205.

Kirkpatrick M, Barton NH (1997) The strength of indirect selection on female mating preferences. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 1282-1286.

Kropf M, Kadereit JW, Comes HP (2002) Late Quaternary distributional stasis in the submediterranean mountain plant *Anthyllis montana L.* (Fabaceae) inferred from ITS sequences and amplified fragment length polymorphism markers. *Molecular Ecology* **11**, 447-463.

Lamont BB, He T, Enright NJ, Krauss SL, Miller BP (2003) Anthropogenic disturbance promotes hybridization between *Banksia* species by altering their biology. *Journal of Evolutionary Biology* **16**, 551-557.

Langlade NB, Feng XZ, Dransfield T*, et al.* (2005) Evolution through genetically controlled allometry space. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 10221-10226.

Leal M, Fleishman LJ (2002) Evidence for habitat partitioning based on adaptation to environmental light in a pair of sympatric lizard species. *Proceedings of the Royal Society B-Biological Sciences* **269**, 351-359.

Levin DA (1971) The origin of reproductive isolating mechanisms in flowering plants. *Taxon* **20**, 91-113.

Levin DA, (1975) Minority cytotype exclusion in local plant populations. *Taxon* **24**, 35-43.

Lowry DB, Rockwood RC, Willis JH (2008) Ecological reproductive isolation of coast and inland races of *Mimulus guttatus*. *Evolution* **62**, 2196-2214.

Macnair MR, Christie P (1983) Reproductive isolation as a pleiotropic effect of copper tolerance in *Mimulus guttatus*. *Heredity* **50**, 295-302.

Mallet J (1989) The genetics of warning color in peruvian hybrid zones of *Heliconius erato* and *Heliconius melpomene*. *Proceedings of the Royal Society of London Series B-Biological Sciences* **236**, 163-&.

Mallet J, Barton NH (1989) Strong natural selection in a warning-color hybrid zone. *Evolution* **43**, 421-431.

Mayr E (1842) *Systematics and the origin of species* Columbia University Press, New York.

Mila B, Warren BH, Heeb P, Thebaud C The geographic scale of diversification on islands: genetic and morphological divergence at a very small spatial scale in the Mascarene grey white-eye (Aves: *Zosterops borbonicus*). *Bmc Evolutionary Biology* **10**.

Miller RS (1967) Pattern and process in competition. *Advance in Ecological Research* **4**, 1-74.

Muller HJ (1942) Isolating mechanisms, evolution and temperature. *Biological Symposia* **6**, 71-125.

Nei M, Chesser RK (1983) Estimation of fixation indexes and gene diversities. *Annals of Human Genetics* **47**, 253-259.

Neuffer B, Auge H, Mesch H, Amarell U, Brandl R (1999) Spread of violets in polluted pine forests: morphological and molecular evidence for the ecological importance of interspecific hybridization. *Molecular Ecology* **8**, 365-377.

Nosil P (2004) Reproductive isolation caused by visual predation on migrants between divergent environments. *Proceedings of the Royal Society of London Series B-Biological Sciences* **271**, 1521-1528.

Nosil P, Flaxman SM (2011) Conditions for mutation-order speciation. *Proceedings of the Royal Society B-Biological Sciences* **278**, 399-407.

Nosil P, Vines TH, Funk DJ (2005) Perspective: Reproductive isolation caused by natural selection against immigrants from divergent habitats. *Evolution* **59**, 705-719.

Pease CM, Lande R, Bull JJ (1989) A model of population-growth, dispersal and evolution in a changing environment. *Ecology* **70**, 1657-1664.

Presgraves DC (2010) The molecular evolutionary basis of species formation. *Nature Reviews Genetics* **11**, 175-180.

Presgraves DC, Balagopalan L, Abmayr SM, Orr HA (2003) Adaptive evolution drives divergence of a hybrid inviability gene between two species of *Drosophila*. *Nature* **423**, 715-719.

Ramsey J, Bradshaw HD, Schemske DW (2003) Components of reproductive isolation between the monkeyflowers *Mimulus lewisii* and *M-cardinalis* (Phrymaceae). *Evolution* **57**, 1520-1534.

Ramsey J, Schemske DW (1998) Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annual Review of Ecology and Systematics* **29**, 467-501.

Redenbach Z, Taylor EB (2002) Evidence for historical introgression along a contact zone between two species of char (*Pisces* : *Salmonidae*) in northwestern North America. *Evolution* **56**, 1021-1035.

Remington CL (1968) Suture-zones of hybrid interaction between recently joined biotas. In: *Evolutionary biology* (eds. Dobzhansky T, Hecht MK, Steere WC), pp. 321–428. Plenum, New Yorl.

Rice WR (1998) Requisite mutational load, pathway epistasis, and deterministic mutation accumulation in sexual versus asexual populations. *Genetica* **102-3**, 71-81.

Rice WR, Hostert EE (1993) Laboratory experiments on speication- What have we learned in 40 years. *Evolution* **47**, 1637-1653.

Rice WR, Salt GW (1988) Speciation via disruptive selection on habitat preference-experimental evidence. *American Naturalist* **131**, 911-917.

Rice WR, Salt GW (1990) The evolution of reproductive isolation as a correlated character under sympatric conditions- experimental evidence. *Evolution* **44**, 1140-1152.

Ridley M (1993) *Evolution*, Boston.

Rothmaler W (1956) Taxonomische monographie der Gattung *Antirrhinum* (ed. Akademie-Verlag), Berlin.

Rundle HD (2003) Divergent environments and population bottlenecks fail to generate premating isolation in *Drosophila pseudoobscura*. *Evolution* **57**, 2557-2565.

Rundle HD, Mooers AO, Whitlock MC (1998) Single founder-flush events and the evolution of reproductive isolation. *Evolution* **52**, 1850-1855.

Rundle HD, Nosil P (2005) Ecological speciation. *Ecology Letters* **8**, 336-352.

Savolainen V, Anstett MC, Lexer C*, et al.* (2006) Sympatric speciation in palms on an oceanic island. *Nature* **441**, 210-213.

Schemske DW, Bradshaw HD (1999) Pollinator preference and the evolution of floral traits in monkeyflowers (*Mimulus*). *Proceedings of the National Academy of Sciences of the United States of America* **96**, 11910-11915.

Schluter D (2000) *The Ecology of Adaptive Radiation.*, Oxford.

Schluter D (2001) Ecology and the origin of species. *Trends in Ecology & Evolution* **16**, 372-380.

Schluter D (2009) Evidence for ecological speciation and its alternative. *Science* **323**, 737-741.

Schwinn KE (1999) *Regulation of anthocyanin biosynthesis in Antirrhinum majus*, University of East Anglia.

Seddon JM, Santucci F, Reeve NJ, Hewitt GM (2001) DNA footprints of European hedgehogs, *Erinaceus europaeus* and *E. concolor*. Pleistocene refugia, postglacial expansion and colonization routes. *Molecular Ecology* **10**, 2187-2198.

Servedio MR, Noor MAF (2003) The role of reinforcement in speciation: Theory and data. *Annual Review of Ecology Evolution and Systematics* **34**, 339-364.

Shuker DM, Underwood K, King TM, Butlin RK (2005) Patterns of male sterility in a grasshopper hybrid zone imply accumulation of hybrid incompatibilities without selection. *Proceedings of the Royal Society B-Biological Sciences* **272**, 2491-2497.

Smith HM (1955) The perspective of species. *Turtox News* **33**, 74-77.

Sobel JM, Chen GF, Watt LR, Schemske DW (2010) The biology of speciation. *Evolution* **64**, 295-315.

Suchet C, Dormont L, Schatz B*, et al.* (2011) Floral scent variation in two *Antirrhinum majus* subspecies influences the choice of naïve bumblebees. *Behavioral Ecology and Sociobiology* **65**, 1015-1027.

Sutton DA (1988) *A revision of the tribe Antirrhineae*, London and New York.

Swenson NG, Howard DJ (2005) Clustering of contact zones, hybrid zones, and phylogeographic breaks in North America. *American Naturalist* **166**, 581-591.

Taberlet P, Fumagalli L, Wust-Saucy A-G, Cossons J-F (1998) Comparative phylogeography and post-glacial colonization routes in Europe. *Molecular Ecology* **7**, 453-464.

Tastard E, Andalo C, Giurfa M, Burrus M, Thebaud C (2008) Flower colour variation across a hybrid zone in *Antirrhinum* as perceived by bumblebee pollinators. *Arthropod-Plant Interactions* **2**, 237-246.

Tastard E, Ferdy JB, Burrus M, Thebaud C, Andalo C (2011) Patterns of floral colour neighbourhood and their effects on female reproductive success in an *Antirrhinum* hybrid zone. *Journal of Evolutionary Biology* **25**, 388-399.

Taylor JW, Jacobson DJ, Kroken S*, et al.* (2000) Phylogenetic species recognition and species concepts in fungi. *Fungal Genetics and Biology* **31**, 21-32.

Tiffin P, Olson MS, Moyle LC (2001) Asymmetrical crossing barriers in angiosperms. *Proceedings of the Royal Society of London Series B-Biological Sciences* **268**, 861-867.

Turelli M, Barton NH, Coyne JA (2001) Theory and speciation. *Trends in Ecology & Evolution* **16**, 330-343.

Turelli M, Orr HA (2000) Dominance, epistasis and the genetics of postzygotic isolation. *Genetics* **154**, 1663-1679.

Tzedakis PC, Hooghiemstra H, Palike H (2006) The last 1.35 million years at Tenaghi Philippon: revised chronostratigraphy and long-term vegetation trends. *Quaternary Science Reviews* **25**, 3416-3430.

Vargas P, Carrio E, Guzman B, Amat E, Guemes J (2009) A geographical pattern of *Antirrhinum* (Scrophulariaceae) speciation since the Pliocene based on plastid and nuclear DNA polymorphisms. *Journal of Biogeography* **36**, 1297-1312.

Via S (1999) Reproductive isolation between sympatric races of pea aphids. I. Gene flow restriction and habitat choice. *Evolution* **53**, 1446-1457.

Via S, Bouck AC, Skillman S (2000) Reproductive isolation between divergent races of pea aphids on two hosts. II. Selection against migrants and hybrids in the parental environments. *Evolution* **54**, 1626-1637.

Warren J, Mackenzie S (2001) Why are all colour combinations not equally represented as flower-colour polymorphisms? *New Phytologist* **151**, 237-241.

Webb T, Bartlein PJ (1992) Global change during the last 3 million years - Climatic controls and biotic responses. *Annual Review of Ecology and Systematics* **23**, 141-173.

Whibley AC (2004) *Molecular and genetic variation underlying the evolution of flower colour in Antirrhinum* University of East Anglia

Whibley AC, Langlade NB, Andalo C*, et al.* (2006) Evolutionary paths underlying flower color variation in *Antirrhinum*. *Science* **313**, 963-966.

Wilson Y, Hudson A (2011) The evolutionary history of Antirrhinum suggests that ancestral phenotype combinations survived repeated hybridizations. *Plant Journal* **66**, 1032-1043.

Wood TK, Keese MC (1990) Host-plant-induced assortative mating in *Enchenopa* treehoppers *Evolution* **44**, 619-628.

Wu CA, Campbell DR (2007) Leaf physiology reflects environmental differences and cytoplasmic background in *Ipomopsis* (*Polemoniaceae*) hybrids. *American Journal of Botany* **94**, 1804-1812.

Auteur : Aurélie Khimoun

Titre : Histoire évolutive, contexte spatial et écologique de la divergence de deux sous-espèces d'*A. majus*.

Directeurs de thèse : Monique Burrus et Christophe Andalo

Résumé : Identifier les mécanismes évolutifs impliqués dans la spéciation, *i.e.* formation des espèces, est crucial dans une époque de crise de la biodiversité. À ce jour, la classification dichotomique des mécanismes impliqués dans le processus de spéciation (*i.e.* allopatrie *versus* sympatrie, écologique *versus* non-écologique …) n'intègrent pas l'interaction dynamique qui les régit. Dans cette thèse, nous avons cherché à intégrer au mieux les différentes composantes de la spéciation – isolement spatial, écologique et reproducteur – dans un cadre temporel. Pour cela, nous avons étudié la divergence de deux sous-espèces de plantes à fleurs, *Antirrhinum majus pseudomajus* et *A. m. striatum* aux fleurs respectivement rouges et jaunes. Nous avons intégré durant cette thèse des approches d'écologie (modélisation de niche) de génétique des populations (analyses de structuration génétique, inférence basée sur le coalescent, ajustement de clines) à différentes échelles spatiales, de l'aire de distribution globale à une zone hybride localisée, en pensant par le détail des zones de contact. Nos résultats montrent comment les processus historiques et démographiques agissent conjointement sur la distribution actuelle des populations des deux sous espèces d'*A. majus*. L'ensemble de nos résultats soutient l'hypothèse du rôle adaptatif de la couleur des fleurs dans le maintien de l'isolement à l'échelle de l'espèce. De plus, la divergence de niche que nous avons détectée entre les deux sous-espèces pourrait être à l'origine de leur divergence phénotypique. Dans les zones de contact, nous avons mis en évidence des échanges de gènes récurrents entre sous-espèces associés à des gradients d'expansion géographique dans des directions opposées. L'ensemble de nos résultats soulève l'hypothèse d'un maintien de la parapatrie entre sous-espèces assuré par leur exclusion compétitive dans les zones de contact. En conclusion, la divergence des deux sous-espèces semble résulter de la combinaison dans le temps de processus historiques neutres et démographiques actuels avec des processus sélectifs associés à l'écologie des sous espèces qui modèlent non indépendamment leur diversité.

Discipline : Écologie, Biodiversité, Évolution

Thèse réalisée au Laboratoire EDB (Évolution et Diversité Biologique), UMR 5174 CNRS/UPS, 118 route de Narbonne, F-31062 Toulouse, France.