



HAL
open science

PATRON DE POLYMORPHISME ET SIGNATURE MOLECULAIRE DE L'ADAPTATION AU MILIEU SALIN DE HELIANTHUS PARADOXUS

Cécile Edelist

► **To cite this version:**

Cécile Edelist. PATRON DE POLYMORPHISME ET SIGNATURE MOLECULAIRE DE L'ADAPTATION AU MILIEU SALIN DE HELIANTHUS PARADOXUS. Ecologie, Environnement. Université Paris Sud - Paris XI, 2007. Français. NNT: . tel-00196521

HAL Id: tel-00196521

<https://theses.hal.science/tel-00196521>

Submitted on 12 Dec 2007

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

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

N° Ordre : 8597

Université Paris XI
UFR Scientifique d'Orsay

THESE

Présentée pour obtenir le grade de
Docteur en Sciences de l'Université Paris XI Orsay

Spécialité : Sciences de la vie

par

Cécile EDELIST

**PATRON DE POLYMORPHISME ET SIGNATURE
MOLECULAIRE DE L'ADAPTATION AU MILIEU SALIN DE
*HELIANTHUS PARADOXUS***

Soutenue le 1^{er} février 2007 devant le jury :

Jérôme GOUDET, Professeur à l'Université de Lausanne	Président
Malika AINOUCHE, Maître de Conférences à l'Université de Rennes 1	Rapporteur
Xavier VEKEMANS, Professeur à l'Université de Lille 1	Rapporteur
Marie-Hélène MULLER, Chargée de Recherche à l'INRA de Mauguio	Examineur
Christine DILLMANN, Professeur à l'Université Paris XI	Directrice de thèse
Delphine SICARD, Maître de Conférences à l'Université Paris XI	Encadrante

PATRON DE POLYMORPHISME ET SIGNATURE MOLECULAIRE DE L'ADAPTATION AU MILIEU SALIN DE *HELIANTHUS PARADOXUS*

Deux espèces sauvages de tournesols *Helianthus annuus* et *H. petiolaris* ont donné naissance il y a plusieurs milliers d'années à une espèce hybride *H. paradoxus* inféodée à des marais salins. Ma thèse a consisté à caractériser les bases génétiques de l'adaptation de *H. paradoxus* à ce milieu extrême.

J'ai recherché des signatures de sélection à deux échelles : à l'échelle du génome et à une échelle plus fine comprise entre 20 et 90 cM. Pour cela j'ai analysé les patrons de diversité génétique pour des marqueurs microsatellites dans trois populations naturelles de *H. paradoxus* et des deux espèces parentales, autour de trois QTL de survie et ailleurs dans le génome. A l'échelle du génome, une baisse de diversité autour de ces QTL a été mise en évidence chez l'espèce hybride et non chez ses parents indiquant qu'il est possible de détecter une signature de la sélection à cette échelle chez cette espèce. A l'échelle plus fine, le patron de diversité est en mosaïque. L'ordre des marqueurs n'étant pas connu avec certitude chez l'espèce hybride, une méthode basée sur le déséquilibre de liaison entre des marqueurs génotypés dans des populations naturelles a été développée. Cette méthode permet de regrouper sur une carte des marqueurs en fort déséquilibre de liaison. La correspondance entre carte de déséquilibre de liaison et carte génétique varie selon les populations et l'échelle génétique.

Parallèlement une étude physiologique et d'expression génique visant à comprendre le mécanisme d'adaptation à la salinité de *H. paradoxus* a été conduite. L'espèce hybride est très plastique, mais se porte mieux en milieu salin, ses feuilles sont plus succulentes, avec une concentration en sodium et sulfate très élevée dans ses tissus. Plusieurs gènes candidats, impliqués dans différentes voies de réponse à la salinité, et dont certains sont cartographiés dans les régions des QTL, sont différentiellement exprimés chez l'espèce hybride et chez ses parents. Ces résultats confirment qu'ils sont des candidats potentiels importants dans l'adaptation de *H. paradoxus* aux marais salins et donc impliqués dans le processus de spéciation.

POLYMORPHISM PATTERN AND MOLECULAR SIGNATURE OF ECOLOGICAL SELECTION FOR SALT TOLERANCE IN *HELIANTHUS PARADOXUS*

The homoploid hybrid sunflower species, *Helianthus paradoxus*, is derived from two wild sunflower species *H. annuus* and *H. petiolaris*, and is adapted to salt marshes. My thesis work characterized the genetic basis of the natural selection that created the adaptation of *H. paradoxus* to this extreme habitat.

I searched for signatures of selection at the whole genome scale, as well as at a finer scale of 20 to 90 cM within individual chromosomes. Accordingly, I analyzed the genetic diversity of three natural populations of *H. paradoxus* and its parental species using microsatellite markers. For the analysis, I used microsatellite markers that are located near three survivorship QTLs, and compared their genetic diversity to markers from putative unselected regions. Genetic diversity was significantly lower around the survivorship QTL in the hybrid species but not in the parental species, signaling for the signature of selection in the *H. paradoxus* genome detectable at this scale. At the finer scale, I found a mosaic pattern of genetic diversity. To overcome unknown mapping locations of genetic markers in *H. paradoxus*, a method to group and order markers based on measure of linkage disequilibrium in natural populations was developed. Congruence of the order according to the sunflower genetic map and the linkage disequilibrium ordering depended on the populations and the genetic scale tested.

In addition, a physiological and gene expression study was developed to understand the mechanisms of *H. paradoxus* adaptation to salty habitat. The hybrid species exhibited a high plasticity response, and performed better than its parental species in a saline habitat. Leaves of *H. paradoxus* were more succulent and have a higher concentration of sodium and sulfate, compared to the parental species. Several candidate genes, implied in various salinity response pathways and located for some within the mapped QTL regions, were differentially expressed in the hybrid species and the two parental species. These results confirm that these genes are potential candidate genes for studying *H. paradoxus* adaptation to saline marshes, and probably played a major role in the process of speciation.

Remerciements

Cette aventure a commencé un jour d'automne 2002 – quand à la recherche d'un stage de DEA – Delphine Sicard, Frédéric Hospital et Christine Dillmann ont compris mieux que moi-même mes attirances pour la génétique évolutive et m'ont orienté vers Loren Rieseberg. Loren a accepté de relever le défi et la machine était lancée. Je tiens à vous remercier très vivement tous les quatre, sans vous mon sujet de thèse n'aurait jamais existé, ou du moins n'aurait pas été mien.

Merci Christine et Delphine d'avoir accepté de continuer l'aventure pour trois années supplémentaires sur un sujet qui n'était pas le votre. Tout au long de ces trois années, vous m'avez encadrée, aiguillée, consacrée de votre temps – oh combien précieux, je le sais – tout en me laissant une liberté dont je vous suis reconnaissante.

Je ne sais comment remercier Loren pour son accueil très chaleureux au cours de mes deux séjours à Bloomington. Merci Loren, également pour la rapidité et l'efficacité à laquelle tu règles les problèmes, réponds aux emails... c'est toujours très impressionnant.

Je tiens à remercier chaleureusement les membres de mon jury, Malika Ainouche et Xavier Vekemans d'avoir accepté d'être rapporteurs de ma thèse, Marie-Hélène Muller et Jérôme Goudet d'en être les examinateurs, malgré le surcroît de travail que cela occasionne.

Une thèse, c'est aussi le parcours du combattant administratif et au Moulon ce parcours est grandement simplifié par l'efficacité de Marie-Marguerite (j'ai la chance de t'avoir connue avant ton départ à la retraite), Elisabeth, Sylvie, Dominique, Valérie et Evelyne. Merci à toutes les six pour votre patience et bonne humeur qui font les petits plus du quotidien.

Merci à Matthieu Falque, Céline Ridet et Xavier Raffoux qui ont accepté l'arrivée d'une nouvelle « bestiole », le tournesol, au laboratoire et qui ont pris en charge, avec beaucoup de sérieux, une grande partie des manips. Merci à tous les trois pour votre disponibilité et vos éclaircissements. Merci Xavier de m'avoir montré l'application de mes cours de biophysique, le point critique n'est pas une invention, il existe, il est spectaculaire, magique, vive le lyophilisateur !

Un grand merci à Sylvie Coursol, Matthieu Falque, Olivier Martin, Adrienne Ressayre et bien sûr Christine et Delphine pour leurs relectures et leurs multiples suggestions. Vous aviez beaucoup d'autres choses à faire que de relire des morceaux de ma prose. Quant à toi

Aymé – maître absolu de la bibliographie – tu m’as économisé de nombreuses heures de sommeil et je t’en suis infiniment reconnaissante.

Au Moulon, si vous montez les escaliers du bâtiment principal, que vous tournez à droite deux fois, puis à gauche et que vous empruntez le couloir, traversez la cafet’, descendez quelques marches avant de prendre l’escalier sur votre droite, vous arrivez à l’étage de l’équipe GQF. GQF c’est actuellement Dominique, Olivier, Wang, Aurélie, Delphine, Christine, Adrienne, Aymé et deux plants d’aubergine aux œufs d’or. Si le quotidien au jour le jour, dans ce cul-de-sac du labyrinthe moulonais (mouloniens, moulonois !?!), est très agréable, c’est grâce à vous. Je n’oublierai notamment pas les parties de ping-pong, les repas libanais ou marocains avant de longues nuits de cinétiques pour certains, ainsi que les débats animés de la cantine – souvent politiques, ou sur la condition et le droit des femmes – (Adrienne, à la maison, c’est bien les femmes qui doivent passer l’aspirateur, ou je n’ai rien compris ?).

Ce ne serait pas fair-play de ne pas remercier l’ensemble des Rieseberger passés et présents. L’ambiance dans le labo de Loren, et Loren n’y est pas étranger, a fait que je me suis tout de suite sentie à ma place. Merci notamment à Sophie Karrenberg, Zhao Lai et Christian Lexer qui m’ont été d’un grand soutien. Merci également à Yuval Sapir pour l’impression de dernières minutes de mes posters, m’évitant à deux reprises les rouages français un peu trop rigides et complexes.

A Letizia, compagne-thésarde de galère, merci ; c’était bien agréable de pouvoir se reconforter quand en fin de thèse on pense qu’on n’y arrivera jamais. Bénédicte je ne te remercierai jamais assez pour ton organisation mémorable lors de notre déplacement à Dijon. Je me demande encore a posteriori si nous n’aurions pas été plus tranquilles dans l’hôtel de passes, si joliment dénommé ainsi par le gros colosse de la gare. Cela étant dit tu m’as vraiment impressionnée par ton calme. Ne t’en fais pas, même si cet épisode m’a marquée, je retiendrais également notre escapade très agréable dans le marais poitevin et j’espère bien que l’on aura d’autres occasions de se balader ensemble. Merci Bénédicte et Yves pour votre prestation de représentants de feuilles - couvertures cartonnées et colorées, les nerfs avaient lâché et votre réconfort n’a pas été inutile. Bon courage à vous Bénédicte, Aymé, Yves, Jonathan... pour la suite et / ou fin de vos thèses.

Enfin, outre ma famille qui a toujours été d’un très grand soutien au cours de ces trois années et de toutes les autres années il va sans dire, je tiens à souligner l’influence qu’a eu Grégoire Loïs sur mes choix. Grégoire, mon tout premier maître de stage, tu as été le premier à me parler de recherche, de thèse... Tu as toujours cru, plus que moi-même, que c’était la

voie à suivre, et celle que je suivrais. Depuis j'ai apprécié nos discussions, souvent par emails, ta façon très anarchique et ultra-lucide de voir, d'analyser l'actualité présente et passée. C'est sans compter la découverte que je te dois d'Albert Londres, un journaliste hors pair et de Steve Reich et de la musique minimaliste en général. Surtout ne change pas et continue d'assurer : la familia Loïs est vraiment extra.

Liste d'abréviations

ABA : ABscisic Acid

ACP : Analyse en Composantes Principales

AFLP : Amplified Fragment Length Polymorphism

BC : BackCross

CDPK : Calcium Dependant Protein Kinase

DL : Déséquilibre de Liaison

eQTL : expressed Quantitative Trait Loci

EST : Expressed Sequence Tag

LD : Linkage Disequilibrium

LEA : Late Embryogenesis Abundant

LOD : Logarithm of ODds

QTL : Quantitative Trait Loci

QTN : Quantitative Trait Nucleotide

RFLP : Random Fragment Length Polymorphism

NIL : Near Isogenic Line

ppm :partie par million

RIL : Recombinant Inbred Line

ROS : Reactive Oxygen Species

SOS : Salt Overly Sensitive

Table des matières

Introduction.....	1
I. Le processus de spéciation.....	1
1. Un bref historique.....	1
2. Les modes majeurs de spéciation.....	2
II. D'où viennent les différences entre espèces ?.....	7
1. Bases génétiques de la spéciation.....	7
2. Bases écologiques de la spéciation.....	13
III. Quels outils pour mettre en évidence de l'adaptation ?.....	18
1. Détecter de la sélection à partir de données moléculaires.....	18
2. Cartographie de gènes d'adaptation.....	20
3. Analyse des variations d'expression.....	22
IV. La spéciation au sein du groupe <i>Helianthus</i> - le cas de <i>Helianthus paradoxus</i>	23
1. <i>Helianthus paradoxus</i> dans le genre <i>Helianthus</i>	23
2. Les trois hybrides homoploïdes : <i>H. anomalus</i> , <i>H. deserticola</i> et <i>H. paradoxus</i>	25
3. La biologie et l'histoire de l'espèce <i>H. paradoxus</i>	27
4. Les outils dont on dispose et les expériences réalisées autour du complexe <i>H. annuus</i> et <i>H. petiolaris</i>	28
5. Quelles méthodes pour détecter et comprendre les bases génétiques de l'adaptation au milieu salin de <i>H. paradoxus</i> ?.....	29
Chapitre 1 : Signature génétique de la sélection autour de 3 QTLs de survie dans le génome de <i>Helianthus paradoxus</i>	31
Microsatellite signature of ecological selection for salt tolerance in a wild sunflower hybrid species, <i>Helianthus paradoxus</i>	34
Introduction.....	34
Materials and Methods.....	35
Results.....	38
Discussion.....	41

Chapitre 2 : Cartographie génétique sur des populations naturelles et signature de la sélection à l'échelle d'un groupe de liaison.....	46
I. Intensification du marquage sur les groupes de liaisons contenant les QTL de survie.....	46
II. Using linkage disequilibrium patterns for ordering genetic markers in natural populations.....	50
Introduction.....	52
Material and Methods.....	53
Results.....	58
Discussion.....	69
III. Signature de sélection à l'échelle d'un groupe de liaison chez <i>H. paradoxus</i>	74
Résultats.....	74
Discussion.....	77
 Chapitre 3 : Expression différentielle de gènes candidats potentiellement impliqués dans la réponse à la salinité chez trois espèces de tournesol : l'halophyte <i>Helianthus paradoxus</i> et ses parents glycophytes <i>H. annuus</i> et <i>H. petiolaris</i>	79
Introduction.....	79
Matériel et Méthodes.....	84
Résultats.....	96
Discussion.....	101
 Discussion et Perspectives.....	108
 Références.....	112
 Annexe.....	122

INTRODUCTION

I. Le processus de spéciation

1. Un bref historique

La spéciation est le processus évolutif par lequel de nouvelles espèces sont formées. Darwin dans son livre « l'origine des espèces » (1859) est le premier à reconnaître que les espèces évoluent et se « séparent » pour donner naissance à de nouvelles espèces. Cette séparation est la conséquence selon lui de l'adaptation à de nouvelles niches écologiques. Il considère que la spéciation est un phénomène progressif et continu (Figure 1), dont le moteur est la sélection naturelle. Dans sa théorie, Darwin ne fait pas de réelle distinction entre spéciation et adaptation : il voit l'origine des espèces comme une conséquence directe de la lutte pour la vie entre individus.

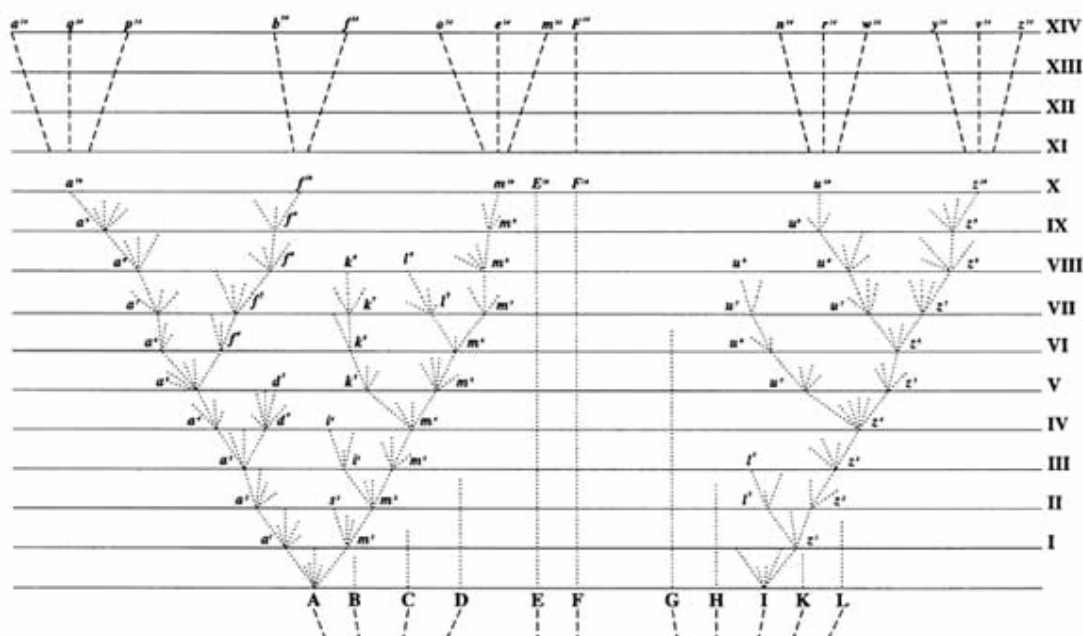


Figure 1 : Les lettres de A à L représentent des espèces d'un genre ; elles sont séparées par une plus ou moins grande distance selon leur degré d'apparement. Une espèce peut s'éteindre plus (B, C, D, H, K, L) ou moins (E et G) vite, se transformer au cours du temps (E, F), ou donner naissance à de nouvelles espèces (A et I) (Darwin 1859).

Les « mutationnistes », De Vries (1906), Bateson (1922), eux, pensent que l'évolution se fait par sauts mutationnels. Il leur paraît impossible d'expliquer les entités discontinues que sont les « espèces » dans la nature par un phénomène évolutif continu.

C'est Dobzhansky (1935) qui réconciliera les deux écoles. La spéciation est bien un processus continu, et les changements de fréquences alléliques peuvent produire des groupes d'individus génétiquement et morphologiquement différents dans un même habitat. Cependant la notion de spéciation de Darwin est incomplète, car il est nécessaire pour que des

formes écologiquement distinctes coexistent, que des barrières géniques soient mises en place. La spéciation n'est plus uniquement un problème de colonisation et d'occupation de nouvelles niches écologiques, mais également de déploiement de barrières d'isolement reproducteur entre espèces (Dobzhansky 1937). Si Dobzhansky et Mayr s'entendent pour dire que la sélection naturelle joue un rôle prépondérant dans la spéciation, Mayr (1963) affirme qu'il ne faut pas pour autant négliger le rôle de la dérive génétique.

Il est intéressant de noter que c'est Mayr en 1942 qui pour la première fois codifie le sens du mot « espèce » avec le « biological species concept » (BSC). Selon ce concept, une espèce est un groupe de populations naturellement interfécondes et isolées reproductivement de tout autre groupe. L'espèce de ce fait représente, selon lui, l'unité fondamentale de l'évolution. Une autre grande définition d'espèce utilisée en évolution est le « phylogenetic species concept » (PSC) où les groupes d'espèces sont formés à partir de la description de gènes devenus monophylétiques. La notion d'espèce a largement été discutée (dernièrement chez les plantes par Rieseberg *et al.* 2006) et sort du cadre de ce travail, c'est donc le BSC que j'ai choisi d'adopter ici faute de consensus (Hey 2001, dénombre pas moins de 12 définitions différentes, dont la plupart n'ont été définies que dans les dernières décades).

2. Les modes majeurs de spéciation

S'il existe différents modes de spéciation, plusieurs conditions sont toujours nécessaires pour que la spéciation puisse aboutir :

- 1- diminution et/ou arrêt des échanges génétiques (flux de gènes réduits ou inexistantes),
- 2- accumulation des différences (comportementales, morphologiques...).
- 3- mise en place de mécanismes d'isolement reproducteur,

Classer les différents modes de spéciation n'est pas évident, et peut varier selon le critère retenu.

2.1. Selon l'arrangement géographique des populations

Traditionnellement, les modes de spéciation sont classés selon l'organisation géographique des populations au cours du processus : populations allopatriques, sympatriques ou parapatriques.

- Il peut y avoir spéciation allopatrique si des populations interfécondes se retrouvent isolées géographiquement (rivières, montagnes..., ou migration) et évoluent indépendamment pour

aboutir à la formation d'espèces distinctes. Si une population périphérique se sépare de la population mère on parle de spéciation péripatrique.

- Lors de la spéciation sympatrique les populations divergent bien que n'étant pas isolées géographiquement. En général, on considère la spéciation sympatrique comme l'aboutissement d'une sélection disruptive, c'est-à-dire que la sélection naturelle favorise deux phénotypes souvent extrêmes, d'une population ; par exemple une population où les individus petits et grands auraient un avantage par rapport aux intermédiaires. On distingue deux classes de modèles : les modèles écologiques avec une compétition pour les ressources (Dieckmann et Doebeli 1999 ; Drossel et McKane 2000) ; les modèles de sélection sexuelle basés sur la compétition pour trouver un partenaire (Higashi *et al.* 1999). Dans les deux cas la rencontre non aléatoire (« assortative mating ») entre individus est primordiale : par exemple si l'adaptation à différentes niches écologiques chez des poissons, *Gasterosteus* spp, produit des différences dans la taille du corps, les poissons se ressemblant morphologiquement auront tendance à plus se reproduire ensemble (Nagel et Schluter 1998). Gavrilets (communication personnelle) a montré théoriquement que la gamme des paramètres pour qu'il y ait spéciation sympatrique est très précise : il est nécessaire qu'il y ait une forte préférence au cours de l'accouplement, un nombre élevé de populations, un petit nombre de locus impliqués, une force de sélection pour l'adaptation locale intermédiaire et une capacité biotique du milieu élevée.

- La spéciation parapatrique est un mélange des deux modèles précédents, puisqu'il concerne des populations adjacentes qui n'ont qu'un faible recouvrement géographique et des échanges géniques limités, dans des zones dites de tension. Le résultat peut être soit la spéciation, soit un « modèle en gradients » où une espèce unique est distribuée de façon continue le long d'un gradient environnemental, comme l'ont décrit théoriquement Fisher (1930, pp. 125 – 129), puis Doebeli et Dieckmann (2003). Empiriquement une espèce unique peut montrer des signes de différences morphologiques et / ou génétique au niveau de transitions d'habitats, comme c'est le cas des lézards, *Anolis roquet*, à la Martinique (Ogden et Thorpe 2002).

2.2. Selon le mécanisme d'isolement reproducteur

Une autre classification consiste à décrire le mécanisme d'isolement reproducteur mis en place lors du processus de spéciation.

L'isolement reproducteur peut être pré-zygotique. Il empêche l'accouplement, par isolement éthologique (chants, signaux lumineux...), anatomique (incompatibilité des organes reproducteurs), phénologique (décalage dans le temps de la période d'accouplement), ou

géographique (milieux fragmentés). Si deux partenaires se rencontrent, l'isolement reproducteur peut aussi avoir lieu au niveau des gamètes qui ne se rencontrent pas ou ne survivent pas.

L'isolement reproducteur peut être post-zygotique dans le cas où deux individus se rencontrent et donnent naissance à des individus hybrides non viables ou stériles. Les mécanismes impliqués dans la barrière post-zygotique sont probablement nombreux. Les réarrangements chromosomiques qui ont eu lieu lors de la spéciation pourraient conduire à un isolement reproducteur. En effet, ils sont à l'origine d'une fécondité réduite des hétérozygotes due à des problèmes de disjonction et de ségrégation des chromosomes lors de la méiose. La barrière post-zygotique à la reproduction entre espèces, aussi appelée incompatibilité hybride, peut être le résultat d'interactions épistatiques entre allèles qui diffèrent entre espèces et se retrouvent chez l'hybride ensemble. Cette idée a été proposée par Dobzhansky et Muller (Dobzhansky 1937 ; Muller 1940) et est connue sous le nom de modèle de Dobzhansky-Muller : Lors de la spéciation, des combinaisons d'allèles différentes apparaissent et augmentent en fréquence dans chaque population isolée, conduisant à deux espèces filles. Quand les populations se rencontrent et les individus des deux espèces filles se croisent, les hybrides sont stériles ou létaux parce qu'il y a incompatibilité entre certains allèles. Les gènes d'incompatibilité hybride doivent avoir trois propriétés pour suivre le modèle de Dobzhansky-Muller. Chaque gène doit i) diminuer la fitness de l'hybride, ii) avoir divergé fonctionnellement entre les deux espèces qui s'hybrident et iii) rentrer en interaction avec le gène « partenaire » de l'autre espèce pour causer l'incompatibilité hybride.

Modèle de Dobzhansky - Muller

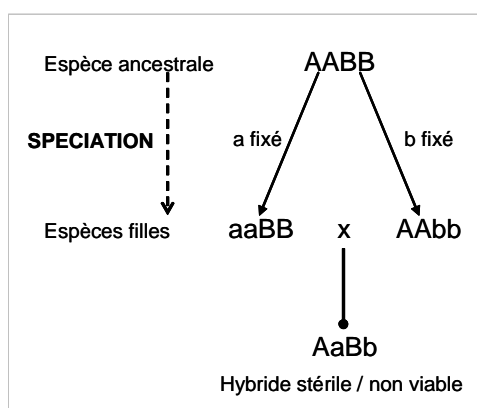


Figure 2 : Deux gènes évoluent indépendamment dans deux populations qui divergent en deux espèces filles. Au cours de l'hybridation entre ces deux espèces filles, de nouvelles interactions épistatiques apparaissent chez les hybrides et conduisent à la mort et/ou une stérilité de l'hybride.

Des gènes d'incompatibilité hybride ont été découverts et leur divergence fonctionnelle entre espèces a été montrée directement ou indirectement (Ting *et al.* 1998, Barbash *et al.* 2004, Presgraves *et al.* 2003). Des systèmes d'incompatibilité hybride composés de plusieurs facteurs, ou des interactions épistatiques chez des hybrides interspécifiques ont déjà été décrits (voir par exemple Sweigart *et al.* 2006). Ce n'est que récemment (Brideau *et al.* 2006) qu'il a été montré que les interactions de deux gènes suffisaient pour expliquer la létalité hybride chez les drosophiles.

2.3. Selon les processus évolutifs majoritairement impliqués

Enfin, une classification probablement plus intéressante, consiste à regarder les processus évolutifs principaux à l'origine de la spéciation. On distingue notamment :

- la « spéciation écologique » par analogie au terme anglais « ecological speciation » qui se fait par sélection naturelle divergente. On parle de radiation adaptative quand il y a diversification rapide d'un groupe monophylétique. Un bel exemple est celui des Cichlidés, des poissons, qui vivent dans les grands lacs est-africains. Les trois principaux lacs (Victoria, Malawi et Tanganyika) contiennent probablement plus de 1 500 espèces, dont beaucoup n'ont jamais encore été décrites.

- la spéciation en allopatrie par accumulation de mutations sous sélection uniforme : la sélection naturelle avantage différentes mutations qui apparaissent dans des grandes populations séparées, mais subissent des pressions de sélection similaires, dans des environnements similaires. On peut alors observer des phénomènes de convergence phénotypique dus à la même mutation. C'est le cas de la mutation du gène *EDA* (Ectodyplasin), un gène impliqué dans le nombre de plaques latérales osseuses défensives chez des épinoches *Gasterosteus aculeatus*. Les populations d'épinoches d'eau douce des différentes parties du globe ont toutes sélectionné le gène conduisant à un faible développement des plaques latérales du fait d'une réduction de la prédation en eau douce (Colosimo *et al.* 2005).

- la spéciation par dérive génétique. La dérive génétique peut avoir un rôle important dans certaines circonstances : petites populations isolées géographiquement, ou subissant des goulots d'étranglement, ou populations en bordure de l'aire de répartition d'une espèce. Si l'on ne connaît pas d'exemples documentés de spéciation par dérive génétique, on peut faire le lien avec le rôle de la dérive dans l'apparition par hasard de nouvelles combinaisons de gènes co-adaptatifs et dans la facilitation de transition entre deux pas dans un paysage adaptatif (Wright, 1932).

- la spéciation par hybridation. On distingue l'hybridation homoploïde - l'hybride possède le même nombre de chromosomes que les deux espèces parentales -, et hybridation alloploïde - l'hybride hérite du génome de ses deux parents-. Contrairement aux types de spéciation précédents, où une espèce donne naissance au minimum à deux espèces, dans le cas de spéciation par hybridation, deux espèces donnent naissance à une troisième espèce. Les espèces polyploïdes sont caractérisées par des origines multiples consécutives à des événements se produisant de façon récurrente (Soltis et Soltis, 1999). On trouve beaucoup plus d'espèces polyploïdes chez les plantes (entre 30% et 70% des espèces de plantes auraient une origine polyploïde, Otto et Whitton, 2000), en particulier les angiospermes et les mousses, que chez les animaux. Une hypothèse pour expliquer cette différence entre règnes serait que la polyploïdisation interfère avec le déterminisme du sexe (Müller, 1925). Effectivement, on trouve plus d'espèces polyploïdes chez les animaux à reproduction parthénogénétique. De même, chez les amphibiens et les reptiles, où la polyploïdie existe dans des groupes à reproduction sexuée, les chromosomes sexuels hétéromorphes sont rares (Otto et Whitton, 2000).

Notons que l'ensemble des forces évolutives est évidemment impliqué dans le processus de spéciation et qu'il s'agit ici seulement de classer en fonction de la force évolutive majoritairement impliquée ou détectée.

II. D'où viennent les différences entre espèces ?

Les différences entre espèces viennent de variations génétiques et environnementales. Les bases génétiques de la spéciation ont été étudiées par des approches de recherches de QTL impliqués dans les différences pour des caractères liés à la fitness ou de gènes à l'origine de l'isolement reproducteur (dites approches génétiques ici). Le rôle de l'environnement dans la spéciation est analysé par des approches de génétique évolutive et d'écologie (ou approches écologiques ici).

1. Bases génétiques de la spéciation

1.1. Nombre, effets et type de gènes

Depuis Darwin (1859), beaucoup d'études ont permis de comprendre le rôle clé de l'adaptation dans l'évolution. Néanmoins, saisir comment la sélection sur des phénotypes influence les changements génétiques et génomiques est loin d'être résolu. La nature précise du déterminisme génétique de la spéciation reste méconnue : le nombre et les effets des changements des caractères adaptatifs sont difficiles à estimer théoriquement et empiriquement.

Théoriquement, Fisher (1930) a proposé un modèle des changements phénotypiques et de l'adaptation (Figure 3). Dans son modèle, un système complexe (l'organisme) doit se conformer à un autre système complexe (l'environnement) qui peut être représenté par un système géométrique simple dans lequel un organisme essaie d'améliorer simultanément de nombreux caractères en produisant des mutations aléatoires dans toutes les directions phénotypiques. Il a montré que, alors que des mutations de faible effet ont 50% de chance d'être avantageuses, les mutations d'effet plus important ont une probabilité décroissante d'être favorables. Ainsi, les mutations à effet majeur ne peuvent pas jouer un rôle important dans l'adaptation. L'adaptation se produit par accumulation de nombreuses mutations bénéfiques de faible effet (Fisher 1930, Orr 2005a et b). Pour Kimura (1983) les mutations de faible effet ont plus de chance d'être bénéfiques mais ont moins de chance de se fixer dans une population, ce qui revient à avantager les mutations intermédiaires. Orr (1998) reprend le modèle géométrique de Fisher (1930) et calcule la distribution des effets fixés tout au long du processus adaptatif. Il démontre qu'il s'agit d'un processus « auto-similaire ». A chaque pas de temps, la distance à l'optimum se réduit grâce à la fixation d'une mutation favorable. La distribution résultante des effets des mutations fixées au cours de l'adaptation est une

distribution de type exponentiel, avec une forte fréquence des mutations de faible effet. Cette distribution dépend surtout de la distance initiale à l'optimum, et non du nombre de caractères impliqués dans la valeur sélective. Dans une population polymorphe Gillepsie (1991) montre que la mutation qui a le plus de chances de se fixer est celle qui a le coefficient de sélection le plus fort. Avec une sélection forte et un environnement constant Kim et Orr (2005) ont montré que ce sont toujours les mutations de plus fort effet qui se fixent en premier. Par contre, dans de grandes populations soumises à des changements environnementaux graduels, les mutations de faible effet ont tendance à se fixer avant les mutations d'effet fort (Bello et Waxman 2006).

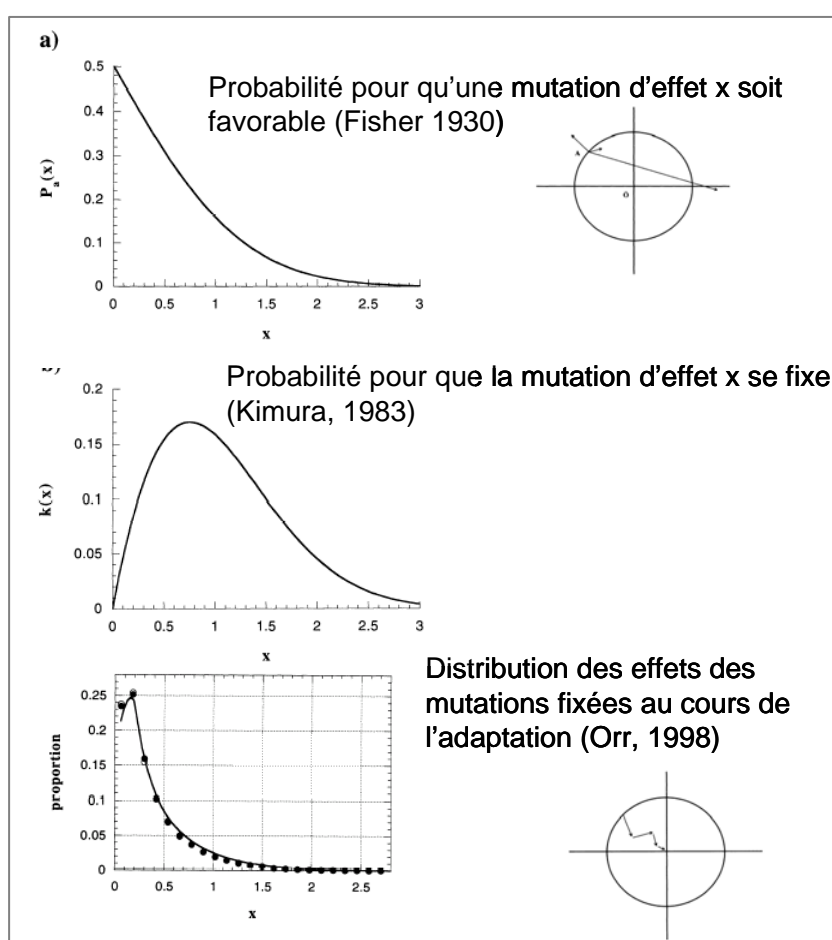


Figure 3 : Modèle géométrique de Fisher (1930) repris par Orr (1998), à gauche les distributions de probabilité et à droite des organismes schématisés en deux dimensions, chaque coordonnée correspondant à un caractère et le phénotype optimum se situe à l'origine θ . Tous les phénotypes à l'intérieur du cercle sont mieux adaptés que ne l'est l'actuelle population. Les flèches représentent des mutations aléatoires de différentes magnitudes et directions

Empiriquement, deux types d'étude ont été développés pour caractériser le nombre et le type de gènes impliqués dans la spéciation :

1- des études de recherche de QTL de caractères qui diffèrent entre espèces proches

2- des études s'appuyant sur la létalité ou la stérilité d'hybrides pour comprendre l'isolement reproducteur.

Déterminisme génétique des différences entre espèces pour des caractères adaptatifs

Comme le note Orr (2001) les différences génétiques qui existent entre espèces peuvent être 'simples' ou 'complexes' et ce de plusieurs manières :

- 1- le nombre de changements génétiques sous-jacents à un phénotype est variable,
- 2- des gènes à large effet peuvent être présents ou absents,
- 3- des interactions épistatiques peuvent exister ou non.

Les études mises en place sont généralement des expériences de cartographie de QTL sur des caractères très variés – comme la résistance à des toxines, les phéromones, ou la taille de certains membre chez des espèces proches de drosophiles, ou la forme, la taille d'organes floraux chez des espèces cousines de mimules - (voir la revue de Orr 2001). Ce qui ressort de l'ensemble de ces études est qu'il est difficile de faire des généralisations sur les bases génétiques des différences entre espèces. Le nombre de QTL pour chaque caractère étudié, qui est à la base de différences entre espèces, est très variable, mais reste faible – d'un à une dizaine selon les caractères. Les différents caractères étudiés ont indéniablement eu des histoires évolutives différentes et ce n'est pas étonnant que le nombre de QTLs soit si variable. L'effet des QTLs diffère aussi selon les caractères : ils expliquent un pourcentage variable de la différence entre espèces. Dans certaines études un gène majeur a été découvert. Par exemple, Sucena et Stern (2000) ont trouvé un gène expliquant la différence morphologique entre des larves de deux espèces de drosophiles, Takahashi *et al.* (2001) ont montré qu'une insertion / délétion dans un gène explique complètement la différence du profil d'hydrocarbures de la cuticules de deux populations, africaine et caribéenne de *Drosophila melanogaster*.

Il est intéressant de noter que moins de QTL sont en moyenne détectés sur des études inter-spécifiques qu'intra-spécifiques probablement pour des raisons de méthodologie. Entre espèces, le nombre de QTL détectés varie entre un et une vingtaine avec une moyenne autour de cinq (Figure 4a), alors que pour les études intra-spécifiques le nombre de QTL détectés varie entre un et 30 avec une moyenne autour de 7 (Figure4b) (Chardon *et al.* 2004 'supplementary material').

Les limites des études de QTL inter-spécifiques en termes de puissance sont que la taille des populations F2 ou backcross est généralement petite. Cela réduit la probabilité de

trouver l'ensemble des QTL d'un caractère donné car l'estimation des effets des QTL est affectée. L'héritabilité du caractère étudié et la densité de marqueurs génotypés peuvent également affecter la qualité des résultats, sans compter que le polymorphisme dans les populations F2, obtenues à partir d'espèces parentales différentes, est très important.

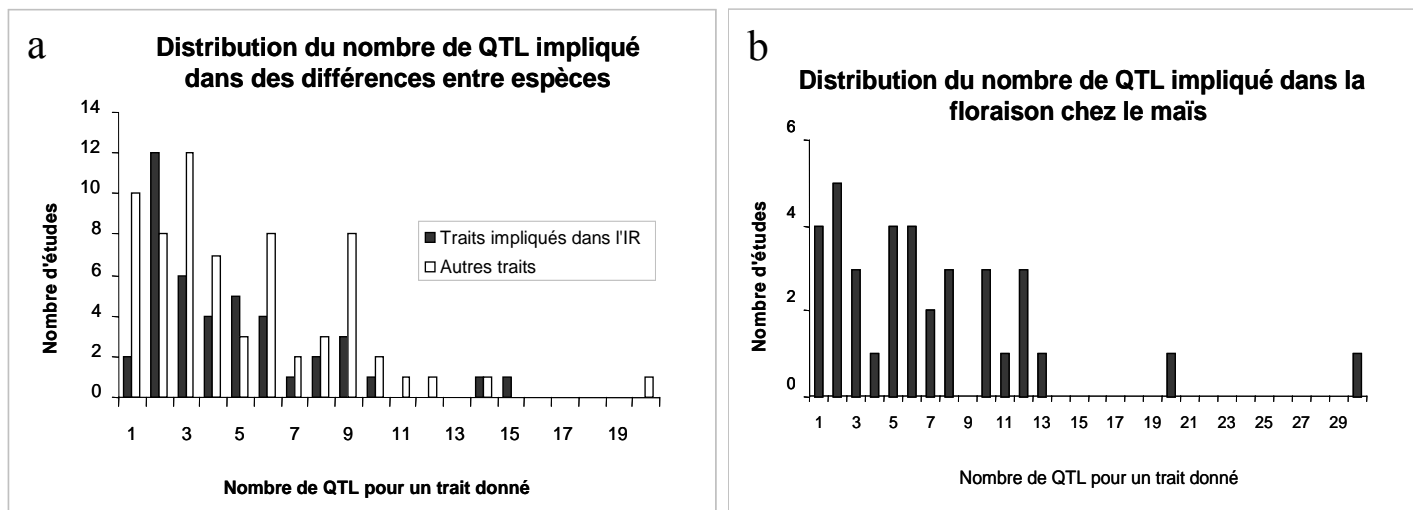


Figure 4 : a) Distribution de QTL dans des croisements inter-spécifiques : le nombre de QTL détectés est généralement faible quelque soit le type de caractère étudié entre deux espèces voisines. En noir, distribution du nombre de QTL dans des études de détection de QTL de caractères directement impliqués dans l'isolement reproducteur (IR) (non viabilité et stérilité des hybrides), données tirées de Coyne et Orr 1998. En blanc, distribution du nombre de QTL dans des études portant sur des caractères différenciant des espèces, mais qui ne semblent pas directement impliqués dans l'isolement reproducteur, données de Coyne et Orr 2004 et Orr 2001. b) Distribution de QTL dans des croisements intra-spécifiques, ici il s'agit de 20 études portant sur des données de floraison chez le maïs (Chardon *et al.* 2004 supplementary material)

La létalité ou la stérilité des hybrides (Incompatibilité hybride)

Lorsque les barrières à la reproduction sont post-zygotiques, elles conduisent à l'incompatibilité hybride, qui s'exprime par la létalité des hybrides F1, ou leur stérilité, partielle ou totale. Une constatation empirique est que la stérilité hybride n'affecte généralement qu'un seul sexe, le sexe hétérogamétique, c'est la règle de Haldane (1922).

On n'a pas d'explication unique pour cette règle, bien qu'elle ait généré une littérature abondante. Le modèle le plus séduisant est la théorie de la dominance (Orr 1993 ; Turelli et Orr 1995), qui suppose que les allèles affectant la fitness des hybrides sont récessifs et ne s'expriment pas chez l'hybride, sauf s'ils sont portés par le sexe hétérogamétique (revue dans Coyne et Orr 2004). De nombreuses études expérimentales, surtout basées sur des croisements entre espèces proches de drosophiles, ont confirmé la récessivité des allèles d'incompatibilités (Presgraves 2003; Tao *et al.* 2003a, b; Tao et Hartl 2003), ainsi que le rôle

majeur du chromosome X dans la stérilité hybride (Turelli et Orr 2000). Cet effet majeur du chromosome X pourrait aussi s'expliquer par une asymétrie des contributions génétiques de chaque espèce à l'hybride entre le chromosome X (espèce utilisée comme femelle) et les autosomes (contribution égale des deux espèces) (Muller 1940), impliquant des interactions épistatiques.

Des études plus récentes ont effectivement montrées que :

- 1- la stérilité du sexe hétérogamétique chez les hybrides est polygénique et complexe (Davis et Wu 1996; Tao *et al.* 2003b),
- 2- l'effet majeur du chromosome X pourrait être un biais, car il disparaît lorsque l'on compare l'effet du chromosome X à celui des autosomes à l'état homozygote (Hollocher et Wu 1996; True *et al.* 1996; Sawamura *et al.* 2000; Tao *et al.* 2003a).

Chez la drosophile, c'est l'interaction entre le gène autosomal *Lhr* présent chez *D. simulans* et le gène *Hmr* présent sur le chromosome X de *D. melanogaster* qui provoque la létalité des hybrides entre ces deux espèces (Brideau *et al.* 2006).

Le modèle de Dobzhansky et Muller quant à lui met en avant le rôle de l'épistasie dans l'incompatibilité hybride. L'épistasie est plutôt rare et faible pour les caractères phénotypiques qui diffèrent entre espèces. Par contre, comme le propose Dobzhansky et Muller, elle est importante pour expliquer l'isolement reproducteur. De nombreuses études expérimentales ont confirmé le modèle de Dobzhansky et Muller. Par exemple, chez deux espèces de mimules, *Mimulus guttatus* et *M. nasutus*, vivant partiellement en sympatrie sur la côte ouest d'Amérique du Nord, deux locus de stérilité mâle hybride ont pu être identifiés et localisés dans chacune des deux espèces, l'allèle d'incompatibilité de *M. guttatus* est dominant au locus 1 et celui de *M. nasutus* récessif au locus 2. Les génotypes d'incompatibilité aux deux locus réduisent également la fertilité femelle, laissant supposer que la fertilité mâle et femelle chez les hybrides issus de ces deux espèces est gouvernée par le même couple de gènes (Sweigart *et al.* 2006). Aucune généralité ne peut donc être tirée quant aux nombres de gènes impliqués dans la stérilité et la récessivité, ou dominance de ceux-ci.

1.2. Mécanismes chromosomiques à l'origine d'isolement reproducteur

Plusieurs travaux expérimentaux révèlent l'importance des variations chromosomiques dans l'isolement reproducteur. Les espèces de drosophile, *Drosophila pseudoobscura* et *D. persimilis*, ont largement servi de modèle depuis les années 1930 pour étudier la spéciation. Noor et son équipe qui étudient les bases génétiques des barrières aux échanges géniques, à

travers la stérilité hybride ou la discrimination dans les rencontres, ont cartographié des caractères des mécanismes d'isolement reproducteur entre ces espèces. Ils ont trouvé que des régions où il y a des inversions chromosomiques entre les deux espèces sont très fortement associées avec les caractères d'isolement reproducteur (Noor *et al.* 2001a et b). Or les régions avec inversions sont moins sujettes à introgressions que les régions colinéaires chez ces espèces (Wang et Hey 1996 ; Machado *et al.* 2002).

Rieseberg *et al.* (1995) ont montré que dans une population backcross d'hybrides synthétiques entre deux espèces de tournesol, *Helianthus annuus* et *H. petiolaris*, l'introgression était moins importante dans les régions chromosomiques réarrangées que colinéaires. De la même manière le flux de gènes entre les espèces sauvages *Helianthus annuus* et *H. petiolaris* dans des zones où les deux espèces coexistent, a été trouvé réduit au niveau des régions de forts réarrangements chromosomiques par rapport aux régions colinéaires (Rieseberg *et al.* 1999)

Ces travaux suggèrent que les inversions chromosomiques entre espèces contribuent à limiter le flux de gènes et donc peut-être à favoriser la spéciation en renforçant l'isolement reproducteur (Rieseberg 2001). Navarro et Barton (2003) ajoutent que des arrangements chromosomiques différents se fixent dans deux populations adjacentes du fait qu'ils contiennent des gènes adaptés aux environnements respectifs des populations. Après la différenciation, il peut y avoir accumulation de nouveaux allèles adaptatifs dans les régions de réarrangements. Les flux de gènes étant réduits dans ces régions, les nouveaux allèles avantageux dans une population mais incompatibles avec les allèles présents dans l'autre population peuvent facilement s'accumuler. Les régions réarrangées deviennent alors des refuges pour des gènes impliqués dans l'isolement reproducteur et constituent d'incroyables barrières à la reproduction entre espèces.

1.3. Les génomes comme une mosaïque génétiques de la spéciation

La spéciation se produit (Figure 5) :

- Dans le temps, l'idéal serait de faire des études entre les différentes étapes car les évènements clés peuvent devenir invisibles au cours du temps.
- Dans l'espace, d'où l'intérêt d'étudier des populations géographiquement distinctes
- Dans le génome, les régions génomiques divergent à des taux différents. Certains locus sont affectés par de la sélection divergente, alors que la majeure partie du génome évolue par dérive. Les deux groupes de locus doivent pouvoir révéler différents aspects du

processus de spéciation : les génomes peuvent ainsi être vus comme des mosaïques génétiques de la spéciation.

La spéciation se produit dans le temps, dans l'espace et dans le génome

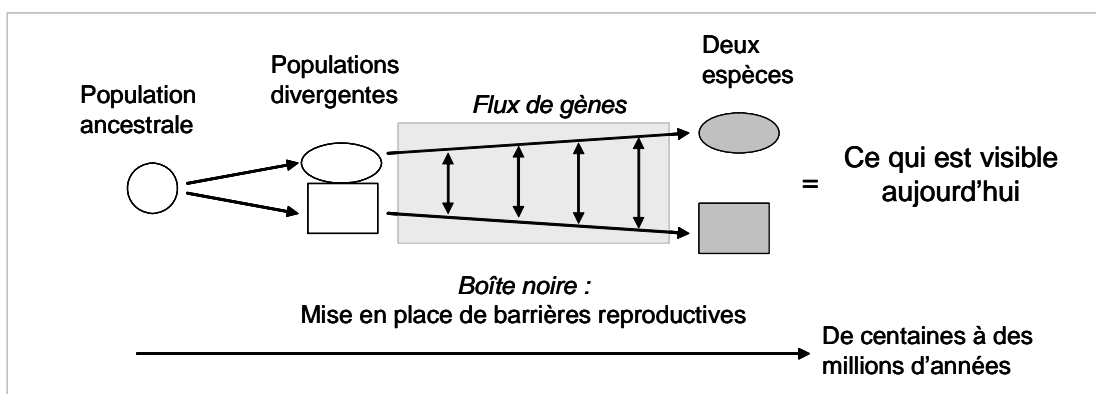


Figure 5 : la divergence conduisant à la spéciation est un continuum avec certains éléments clés qui disparaissent ou deviennent invisibles au cours du temps.

Via et son équipe travaillent sur des populations de pucerons *Acyrtosiphon pisum* vivant en sympatrie mais sur des espèces de plantes différentes, trèfle ou luzerne. Les populations divergent donc pour leur choix d'hôte, ce qui réduit les flux de gènes inter-populations. Une étude de cartographie de QTL de fécondité et de choix d'hôtes a été menée sur une population d'individus F₂ rétro-croisée à chacune des populations initiales. Les QTL trouvés dans chaque population backcross ont été cartographiés dans les mêmes régions, et certains des QTL de fécondité et de choix de l'hôte sont co-localisés, laissant subodorer l'importance de ces régions dans la sélection pour la divergence de choix d'habitat (Hawthorne et Via 2001). Si l'on reprend la théorie de mosaïque génétique le long du génome, on s'attend à ce qu'il y ait moins d'échanges géniques autour des QTL et donc qu'il y ait une divergence entre les deux populations plus importante dans ces régions. C'est effectivement ce que Via et ses collègues observent sur une dizaine de centiMorgans autour des QTL (Via, communication personnelle).

2. Bases écologiques de la spéciation

L'hypothèse de la spéciation écologique est que l'isolement reproducteur évolue comme une conséquence de la sélection naturelle divergente sur des caractères et entre environnements variés (voir revue de Schluter 2001). La spéciation écologique peut avoir lieu en allopatrie comme en sympatrie, elle résulte de combinaisons de processus adaptatifs. Nous illustrons ici différentes stratégies pouvant être mises en place.

2.1. Isolement reproducteur et changement d'habitat

Les variations d'habitat sont à l'origine de sélection divergente qui participe à l'isolement génétique des populations, la diminution des flux de gènes entre populations et l'isolement reproducteur pré-zygotique (par exemple par sélection sexuelle) ou post-zygotique (contre-sélection des hybrides).

Un des plus beaux exemples de spéciation écologique concerne les célèbres pinsons de Darwin, *Geospiza* spp, qui ont évolués sous sélection divergente, dans l'archipel des Galapagos. Ils ont des tailles de corps et de becs qui jouent sur leurs vocalises et se répercutent sur le choix des partenaires (Ratcliffe et Grant 1983 ; Podos 2001).

Un autre exemple classique de « spéciation écologique » qui conduit à un isolement reproducteur pré-zygotique concerne les poissons de lacs glaciaires. Par exemple, des écotypes nains et normaux de l'espèce de poissons, *Coregonus clupeaformis*, vivent en sympatrie dans plusieurs lacs alimentés par la rivière Saint-John au Canada. Une étude phylogéographique du complexe d'espèce a montré que les populations ayant colonisé ces lacs étaient plutôt adaptées à des modes de vie en eaux profondes, et que le phénotype nain a évolué dans chaque lac par un déplacement de niche écologique. Une étude portant sur la différenciation morphologique (19 caractères) et génétique (six microsatellites) entre les deux écotypes dans six lacs a montré l'existence d'une corrélation négative entre les flux de gènes entre les deux types de poissons et les différences morphologiques observées. L'isolement reproducteur entre les poissons nains et normaux semble être lié à leur potentiel à occuper différentes niches trophiques, dans lesquelles ils sont soumis à des pressions sélectives différentes (Lu et Bernatchez 1999). Un autre exemple est celui de populations indépendantes de poissons, *Gasterosteus* spp, ayant évolué dans le même lac et qui se reconnaissent moins bien que les populations provenant de différents lacs mais d'écotype similaire (Rundle *et al.* 2000). Cela illustre comment l'adaptation à des habitats distincts en sympatrie peut accélérer les différences entre populations et conduire à l'isolement reproducteur.

Un exemple de spéciation écologique avec isolement reproducteur post-zygotique concerne les poissons *Gasterosteus aculeatus* et *G. Wheatlandii*, Hatfield et Schluter (1999) ont étudié la fitness d'hybrides issus du croisement de ces deux espèces de poissons qui diffèrent morphologiquement et écologiquement. La fitness des hybrides, F1 et F2, est élevée en laboratoire, mais une fois transplantée dans le milieu naturel de chacun des deux parents, la croissance des hybrides est inférieure à celle de chacun des parents. La fitness moindre des hybrides dans la nature pourraient expliquer la mise en place de barrière de reproduction sans incompatibilité génétique intrinsèque totale, comme dans le modèle de Dobzhansky Muller.

2.2. Le cas de l'hybridation

Il y a deux mécanismes d'hybridation qui peuvent donner naissance à une nouvelle espèce : l'hybridation allopoloïde et homoploïde. Bien qu'ils ne soient pas exclusifs du règne animal, ils ont principalement été étudiés chez les plantes (Grant 1981). L'hybridation est toujours initiée en sympatrie et les deux types d'hybridation requièrent donc que des barrières d'isolement reproducteur soient mises en place, car les flux de gène sont toujours possibles.

Les botanistes reconnaissent depuis longtemps l'importance de l'hybridation allopoloïde dans l'évolution (Stebbins, 1950) et les modèles récents montrent la facilité qu'ont les nouvelles espèces allopoloïdes à apparaître et à s'établir (Rodriguez, 1996). L'allopoloïdie représente en outre le mode de spéciation le plus rapide : l'isolement reproducteur est instantané du fait du doublement du génome. Ce doublement de génome est souvent accompagné par des changements morphologiques, physiologiques ou d'histoire de vie (Levin 1983, 2002 ; Thompson et Lumaret, 1992). Les changements phénotypiques s'accompagnent parfois d'un changement de préférence de niche écologique, ce qui accroît la probabilité d'établissement de la néo-espèce polyploïde (Rodriguez 1996). Ainsi des plantes différant par leur niveau de polyploïdie ont des préférences ou tolérances écologiques différentes : prédilection de substrats, de températures... (Levin 1983, 2002).

L'hybridation homoploïde a été moins étudiée que l'hybridation polyploïde, peut-être parce que c'est un phénomène plus rare dans la nature, ou en tout cas plus difficile à détecter. Les lignées d'hybrides homoploïdes ne sont pas isolées reproductivement des deux espèces parentales, à l'inverse des hybrides allopoloïdes. Stebbins (1957), Grant (1958), Templeton (1980) ont suggéré que le réarrangement chromosomique, la divergence dans le choix d'habitat et les séparations spatiales étaient des facteurs importants d'isolement d'une nouvelle espèce hybride homoploïde. Ces résultats ont été confirmés récemment par des simulations (McCarthy *et al.* 1995 ; Buerkle *et al.* 2000). Les divergences caryotypiques et l'isolement spatial sont bien documentés chez les espèces hybrides naturelles (Gallez et Gottlieb 1982 ; Arnold *et al.* 1990 ; Rieseberg 1991 ; Rieseberg *et al.* 1995 ; Hirai *et al.* 2000) et la plupart des hybrides occupent des habitats différents, voire extrêmes comparés à celui des espèces parentales (Abbott 1992 ; Arnold 1997 ; Rieseberg 1997 ; Schwarzbach *et al.* 2001 ; Gross et Rieseberg 2005). Les modèles de simulation supposent que la différenciation d'habitat doit se faire tôt dans le processus de spéciation (Buerkle *et al.* 2000). Ainsi, le seul cas où l'hybridation homoploïde puisse conduire à une spéciation implique une différenciation entre les espèces parentales et le néo-hybride très rapide.

2.3. Les bases écologiques de la sélection divergente

Chez les plantes il est possible de mesurer directement la valeur sélective de variants alléliques dans leur milieu naturel, ce qui est une stratégie intéressante pour comprendre les bases génétiques de l'adaptation et de la divergence entre populations.

Ainsi Bradshaw et Schemske (2003) ont montré qu'une mutation, impliquée dans le changement de la couleur des fleurs de rose à orangé, a un large effet sur la barrière reproductive entre deux espèces de mimules, *M. cardinalis* et *M. lewisii*. L'allèle de *M. cardinalis* dans une population NIL (Near Isogenic Line) de *M. lewisii* conduit à un changement de pollinisateurs : les plantes mutantes sont majoritairement pollinisées par des oiseaux mouches alors que leurs congénères sauvages sont principalement visitées par des bourdons. Inversement les individus NIL *M. cardinalis* avec le gène de *M. lewisii* ont la faveur des bourdons et non plus des oiseaux mouches. Cela suggère fortement que la couleur de la fleur qui distingue ces deux espèces a évolué en allopatrie, et que les différences de proportions des pollinisateurs n'ont pas avantage les mêmes combinaisons génétiques dans chaque espèce.

Weinig et ses collaborateurs (2003) ont montré que la sélection au niveau de QTL pour des caractères liés à la valeur sélective chez *Arabidopsis thaliana*, diffère selon l'environnement dans lequel la plante évolue. Ils ont réalisé une expérience de détection de QTL dans une population de RIL (Recombinant Inbred Line) issue du croisement entre deux écotypes différant par leurs besoins en vernalisation pour initier la floraison, en mesurant des caractères liés à la survie et à la reproduction des plantes dans deux milieux naturels contrastés différant pour la rigueur de l'hiver. En général, les écotypes ayant besoin d'une période de vernalisation sont rencontrés dans des régions aux hivers froids, ce qui peut être interprété comme une adaptation au climat. A un des QTL, l'allèle qui augmente la viabilité dans un environnement donné, réduit la fécondité dans un autre. Le fond génétique est également important : ainsi à un autre QTL, l'allèle de fitness change de direction selon le génotype à un autre locus. Ces résultats illustrent deux mécanismes de maintien de variation génétique pour la fitness : la sélection environnementale hétérogène et la sélection épistatique.

L'expérience de Lexer et ses collègues (2003) a permis de mieux comprendre comment une nouvelle espèce hybride homoploïde pouvait s'établir. Ils ont croisé des hybrides synthétiques obtenus par croisement de *Helianthus annuus* et *H. petiolaris*, deux espèces de tournesol très répandues en Amérique du Nord et reconnues pour être les espèces parentales de l'espèce hybride homoploïde *H. paradoxus* inféodée à des marais salins. Une

étude de QTL sur la population d'hybrides synthétiques transplantée dans le milieu naturel de *H. paradoxus* a permis de cartographier trois QTL de survie en milieu salin, d'effets opposés et de coefficients de sélection importants. Ainsi l'espèce hybride cumule des allèles des deux espèces parentales qui lui permettent d'être mieux adaptée à un nouveau milieu.

III. Quels outils pour mettre en évidence de l'adaptation ?

La spéciation écologique (divergence, hybridation...) est probablement plus facile à étudier que la spéciation par dérive. En effet, elle laisse des signatures, du moins pour une certaine durée, au niveau des caractères adaptatifs qui ont été sélectionnés. La méthode directe pour mettre en évidence les gènes impliqués dans l'adaptation est, comme nous l'avons vu précédemment, de tester la fitness de différents variants alléliques dans la nature. Les méthodes indirectes consistent à détecter de la sélection au niveau moléculaire par analyse à l'échelle du gène quand on le connaît, ou par analyse d'association. Il s'agit alors de trouver des associations, statistiquement significatives, entre le polymorphisme de marqueurs et la variance des caractères quantitatifs (voir revue de Ehrenreich et Purugganan 2006). De même, on peut analyser les relations entre variations d'expression de gènes candidats et la variance des caractères quantitatifs.

1. Détecter de la sélection à partir de données moléculaires

Le génome peut être vu comme une mosaïque génétique de l'adaptation et toutes les régions, de même que tous les marqueurs moléculaires, ne racontent pas la même histoire. Si la démographie affecte tous les locus de la même manière, ce n'est pas le cas de la sélection qui est locus spécifique et qui peut être déduite du patron de polymorphisme de locus liés (Cavalli-Sforza 1966). En effet, Maynard-Smith et Haigh (1974) ont démontré l'effet de l'autostop génétique. Lorsqu'une mutation apparaît dans une population, elle se trouve le plus souvent en une copie unique chez un individu, dans un contexte haplotypique particulier, c'est-à-dire en déséquilibre de liaison avec les locus de la région concernée. S'il s'agit d'une mutation avantageuse, elle va augmenter en fréquence et entraîner l'augmentation de la fréquence de tous les allèles liés génétiquement et présents sur l'haplotype initial. Le polymorphisme de la région est alors balayé (on parle de balayage sélectif) par la sélection. L'effet est d'autant plus important que la sélection est forte et que le taux de recombinaison est faible (Maynard-Smith et Haigh 1974 ; Kaplan *et al.* 1989, Stephan *et al.* 1992). Ainsi, en général l'effet génétique de l'auto-stop est une réduction du niveau de diversité et une distribution de fréquences alléliques biaisée vers les fréquences extrêmes aux locus neutres liés (Tajima 1989 ; Fu et Li 1993 ; Fay et Wu 2000 ; Kim et Stephan 2000, 2002 ; Kreitman 2000 ; Andolfatto 2001)

De nombreux tests statistiques ont été développés pour mettre en évidence les signatures de la sélection (Nielsen 2001, Storz 2005). La significativité de ces signatures peut-être testée par rapport à une hypothèse nulle : les variations génétiques observées sont-elles conformes à ce qu'on observerait sous l'hypothèse de neutralité (Kimura, 1983), ou observe-t-on une déviation significative ? Elles peuvent aussi être testées par comparaison entre populations pour mettre en évidence des signatures population spécifiques.

1.1. Analyse du polymorphisme d'un gène candidat, potentiellement sous sélection entre espèces

Plusieurs tests de neutralité sélective ont été développés pour détecter de la sélection sur du polymorphisme de séquences entre espèces proches. Le test de Hudson-Kreitman-Aguade (HKA, Hudson *et al.* 1987) est un test communément utilisé. Il mesure le polymorphisme intra- et inter-spécifique à plusieurs locus et teste l'écart entre le polymorphisme observé et celui attendu sous l'hypothèse de neutralité. Le test de McDonald-Kreitman (MK, McDonald et Kreitman, 1991), consiste à comparer le taux de substitutions synonymes d_s (ou K_s) et non synonymes d_n (ou K_a) entre espèces au niveau de polymorphisme synonyme et non synonyme au sein de ces espèces. D'autres méthodes étudie si le ratio $d_n : d_s$ (ou $K_a : K_s$) s'écarte de 1. Un gène sous l'hypothèse nulle aura un ratio égal à 1 et une déviation de cette valeur est une indication d'une sélection possible (Nielsen, 2001; Yang, 1997; Yang et Nielsen 2002).

1.2. Analyse du polymorphisme le long du génome : tests multi-locus

L'étude de polymorphisme le long du génome entre populations, ou à l'intérieur d'une même population a connu un essor avec le développement des marqueurs moléculaires, en particulier les microsatellites, et les SNP (revue de Storz 2005).

Lewontin et Krakauer (1973) ont introduit le premier test de différenciation entre populations adaptées localement. Ils utilisent les valeurs de F_{st} (Wright 1943, 1951) qui sont des mesures de changements de diversité génétique entre populations. Ainsi, leur test consiste à regarder si la variance des valeurs de F_{st} estimées à différents locus est plus importante que ce que l'on attend par hasard. Un résultat significatif indique une divergence entre populations probablement due à de la sélection divergente. Leur modèle a ensuite été complété. Bowcock *et al.* (1991) ont incorporé des informations phylogénétiques sur l'histoire de divergence entre populations et construit une distribution nulle de F_{st} par simulation. Beaumont et Nichols (1996) ont, eux, utilisé une approche par coalescence pour générer les distributions nulles de

F_{st} . Vitalis *et al.* 2001 ont développé une approche basée sur l'estimation de la longueur des branches entre paires de populations.

Une étude théorique de Slatkin (1995) a permis d'étendre le modèle d'auto-stop (Maynard-Smith et Haigh 1974 ; Kaplan *et al.* 1989) et de considérer que des marqueurs microsatellites liés à un locus sous sélection font partie intégrante de la zone sélectivement balayée. Wiehe (1998) a généralisé ce modèle et ajouté que la force de l'auto-stop dépend du taux de mutation, en plus du coefficient de recombinaison et de sélection. Ces études théoriques ont permis à Schlötterer *et al.* (1997) de construire un modèle reposant sur la variance des tailles d'allèles de marqueurs microsatellites pour détecter les locus déviant de l'équilibre mutation-dérive. Les balayages sélectifs spécifiques à une population doivent produire une déviation par rapport à ce que l'on attend à l'équilibre pour la variance de la taille des allèles microsatellites liés aux locus sélectionnés. Schlötterer (2002) propose une approche empirique basée sur une comparaison par paires de populations sur une centaine de locus. Sous le modèle de mutations par sauts (Stepwise Mutation Model) et à l'équilibre mutation-dérive, la variance des tailles d'allèles et l'hétérozygotie attendue peuvent être utilisées pour estimer le paramètre $\theta = 4N_e\mu$, où $N_e\mu$ est le produit du taux de mutation et de l'effectif efficace de la population (Ohta et Kimura 1973 ; Moran 1975). Ainsi pour deux populations, il est possible de calculer le ratio attendu de la variance des tailles d'allèles RV et le ratio attendu de l'hétérozygotie RH . La transformation logarithmique de ces ratios suit une distribution normale pour un éventail de conditions démographiques (Schlötterer 2002 ; Schlötterer et Dieringer 2005). Ainsi les ratios $\ln RV$ et $\ln RH$ peuvent être calculés à chaque locus et utilisés pour construire un histogramme de distribution de ces statistiques. Les locus se trouvant aux extrémités de la distribution, permettent d'identifier les régions ayant potentiellement subi des balayages sélectifs.

2. Cartographie de gènes d'adaptation

De multiples approches basées sur des méthodes d'association entre caractères et marqueurs existent pour cartographier les gènes d'adaptation (Mackay 2001, Remington *et al.* 2001, Phillips 2005). Le principe de toutes cartographies repose sur le fait que la covariance entre le caractère étudié et les marqueurs génotypés est une fonction du déséquilibre de liaison, et que le déséquilibre de liaison est lui-même une fonction du taux de recombinaison entre le QTL et les marqueurs.

2.1. Cartographie de QTL.

Au départ des individus qui diffèrent pour le caractère qu'on étudie sont croisés et leur progéniture est recroisée soit entre elle, soit avec les individus de départ pour générer une population recombinante pour les allèles des deux parents à chaque locus. L'identification de QTL a joué un rôle important pour mettre en évidence des régions génomiques qui contrôlent les variations phénotypiques. Elle a été largement utilisée pour localiser des régions génomiques impliquées dans la variation phénotypique intra- ou inter-spécifique. La limite de cette approche est la possibilité d'obtenir des individus recombinants pour des faibles distances génétiques. Ainsi les régions caractérisées sont larges (> 500kb) et contiennent de nombreux gènes. La résolution peut être améliorée en augmentant le nombre de marqueurs et le nombre d'individus dans la population.

Nous avons vu que pour l'étude des différences entre espèces les méthodes de détection de QTL ont largement été développées. Cependant, les populations utilisées sont le plus souvent des populations de type F2 ou BC, très polymorphes et faiblement recombinantes, ce qui réduit l'efficacité de la méthode.

2.2. Cartographie par déséquilibre de liaison

En dehors des méthodes classiques de cartographie génétique basées sur le taux de recombinaison, le déséquilibre de liaison (DL) peut être utilisé pour cartographier des locus. Le DL est l'association non due au hasard de locus polymorphes à l'intérieur d'une population (Bennett 1954). Les locus en DL peuvent être associés à des différences phénotypiques et donc exploités pour cartographier les régions génomiques qui sont à la base de l'adaptation. Dans la pratique la cartographie par déséquilibre de liaison requiert un échantillon d'individus pour lesquels on possède des données génotypiques et phénotypiques. Les corrélations, entre les variants génétiques observés et la variation des caractères, peuvent être mesurées et conduire à l'identification de polymorphismes ou d'haplotypes spécifiques qui expliquent la variation des caractères adaptatifs. La résolution de cartographie est directement liée au pourcentage de décroissance du DL avec la distance génétique et varie de façon importante selon l'espèce considérée (généralement entre 20 et 50kb). L'un des problèmes majeur de ce type de cartographie provient du fait que les populations sont souvent structurées et qu'une partie du DL détecté est lié à cette structure (Lewontin 1964, 1974 ; Hudson 2001). Un autre problème est la nécessité d'avoir un minimum de diversité aux locus étudiés pour que le DL soit informatif. A noter, que le DL peut être vu comme fractal et donc

utilisé à une autre échelle. En effet, du DL a été trouvé au-delà de plusieurs centiMorgans autour de QTL ou gènes majeurs importants (Farnir *et al.* 2000, Tenesa *et al.* 2004).

Si le DL a été utilisé pour cartographier les régions génomiques importantes dans l'adaptation de certaines populations à leur environnement, à notre connaissance il n'a jamais été utilisé dans le cadre d'études inter-spécifiques.

3. Analyse des variations d'expression

Une approche alternative de la cartographie de gènes d'adaptation consiste à étudier l'expression génique. L'analyse transcriptionnelle du génome à l'aide de microarrays d'ADNc peut être utilisée pour identifier les locus qui interviennent dans les différences phénotypiques entre individus. Plus récemment, la liaison entre les données d'expression génique et des marqueurs moléculaires a été analysée, afin d'identifier des régions d'eQTL (expression Quantitative Trait Loci) (Damerval *et al.* 1994 ; Gibson et Weir, 2005). Associées à la cartographie génique, les études d'expression sont des outils intéressants pour connecter les variations géniques à des différences phénotypiques pour des caractères adaptatifs et déterminer les bases fonctionnelles des changements adaptatifs. Dans le cadre de la spéciation, Bernatchez et son équipe utilisent cette technique pour repérer les différences d'expression génique qui existent entre populations divergentes de poissons *Coregonus* spp. Ils ont ainsi détecté 3 842 eQTL, impliqués dans la croissance, le métabolisme et le comportement, dont 60 ont une expression différentielle entre deux populations (Bernatchez, communication personnelle).

IV. La spéciation au sein du groupe *Helianthus* – le cas de *Helianthus paradoxus*

1. *Helianthus paradoxus* dans le genre *Helianthus*

Le genre *Helianthus* comprend une cinquantaine d'espèces, annuelles ou pérennes toutes originaires d'Amérique du Nord. Certaines espèces sont diploïdes, d'autres tétraploïdes ou encore hexaploïdes et toutes ont un nombre de paires de chromosomes multiple de 17. Schilling et Heiser (1991) ont divisé le genre *Helianthus* en quatre sections (*Agregstes*, *Helianthus*, *Ciliares*, *Divaricati*) selon les caractères morphologiques et la fertilité inter-spécifique. La section *Helianthus* (Figure 6) comprend 12 espèces diploïdes géographiquement présentes dans une partie du Mexique, aux Etats-Unis et dans une partie du Canada. La plupart des espèces de cette section sont annuelles, à l'exception de *H. debilis* et *H. niveus* qui sont pérennes. Les espèces sont toutes allogames. Trois espèces de cette section *H. anomalus*, *H. deserticola* et *H. paradoxus*, sont des espèces hybrides diploïdes résultant de croisement inter-spécifiques entre *H. annuus* et *H. petiolaris* appartenant également à la section *Helianthus* (Rieseberg, 1991). Elles sont toutes les trois inféodées à des milieux différents de celui des parents, des milieux très particuliers.

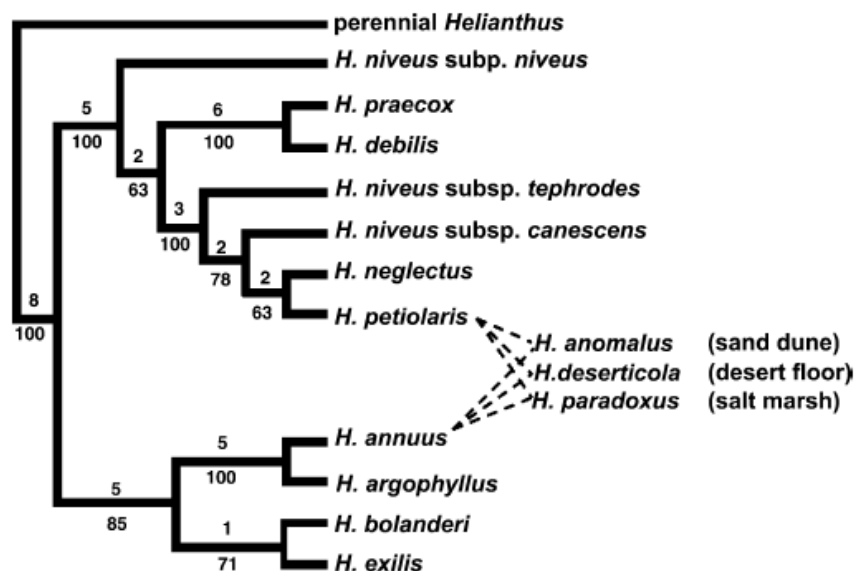


Figure 6 : Arbre phylogénétique de la section *Helianthus*, construit à partir de données d'ADN chloroplastique et ribosomal (Rieseberg 1991). Le nombre de mutations est indiqué au dessus des branches et le pourcentage de bootstrap en dessous des branches. Les lignes pointillées indiquent le lien de parenté avec les espèces hybrides homoploïdes. (Figure tirée de Gross et Rieseberg, 2005).

Fiche d'identité du tournesol

Surnom : Tournesol ou Soleil en français, Sunflower en anglais, Girasol en espagnol et Girasole en italien.

Le tournesol est soumis à l'héliotropisme avant floraison, d'où son surnom.

Famille : Asteraceae, comprise dans les Composées

Genre : *Helianthus*

Il existe 51 espèces de tournesols toutes originaires d'Amérique du Nord

14 de ces 51 espèces sont des espèces annuelles et 37 sont pérennes

Espèce cultivée : *Helianthus annuus* L.

H. annuus reste l'espèce de tournesol sauvage la plus répandue en Amérique du Nord.

Domestication :

Elle aurait eu lieu il y a 4 000 ans, relativement à l'Est en Amérique du Nord.

La plante cultivée n'est pas ramifiée, elle ne possède généralement qu'un seul capitule et produit des akènes plus gros, qui restent accrochés au capitule à maturité.

Le tournesol a été importé en Europe par les espagnols au 16^{ème} siècle.

Le tournesol et l'agriculture

Les plus gros producteurs sont la Russie et l'Ukraine mais le tournesol est cultivé partout dans le monde. Il constitue une des sources principales d'huile alimentaire avec le colza et l'olivier.

Il a un avenir prometteur en tant que biocarburant. Il est également utilisé comme plante fourragère et ses graines sont largement appréciées dans les volières.

Différents cultivars de tournesol sont aussi cultivés en tant que plantes ornementales.

Le tournesol et les mathématiques

Les graines sur le capitule sont disposées en deux spirales inversées : une dans le sens des aiguilles d'une montre, la deuxième dans le sens inverse. Le nombre de spirales appartient à la suite de Fibonacci : généralement 34 et 55, parfois 89 et 144. La suite de Fibonacci est composée de nombres qui sont égaux à la somme des deux précédents et les deux premiers éléments de la suite sont 0 et 1.

Le tournesol et la peinture

Vincent Van Gogh était tombé amoureux de cette fleur, il a peint une série de natures mortes « Tournesols » entre 1887 et 1888. Le seul portrait que Paul Gauguin fait de son ami Van Gogh, le représente en train de peindre des tournesols.

Claude Monet a également réalisé un tableau intitulé « Soleils » en 1881

Ou encore...

Le tournesol c'est aussi le Professeur tournesol avec l'épisode de « l'affaire Tournesol » d'Hergé.

2. Les trois hybrides homoploïdes : *H. anomalus*, *H. deserticola* et *H. paradoxus*

L'hybridation dont est issue *H. anomalus*, *H. deserticola* et *H. paradoxus* est de type homoploïde, c'est-à-dire que l'hybride est diploïde et possède le même nombre de paires de chromosomes ($n=17$) que chacun de leurs parents. L'origine d'une nouvelle espèce via hybridation homoploïde est, comme nous l'avons vu, théoriquement assez difficile car elle requiert la mise en place d'une barrière reproductive en sympatrie. L'hypothèse souvent avancée pour qu'une hybridation homoploïde soit efficace et que la nouvelle espèce hybride perdure, est qu'il est nécessaire d'avoir une ségrégation transgressive pour certains caractères (Tableau 1) : l'hybride doit cumuler les allèles de même effet pour les caractères transgressifs, allèles dont les effets opposés s'annulent chez les espèces parentales (Lewontin et Birch 1966, Rieseberg *et al.* 1999).

QTL	Valeurs phénotypiques			
	Parent AA	Parent BB	F2 transgressif	F2 transgressif
1	+ 1	- 1	+ 1 (AA)	- 1 (BB)
2	+ 1	- 1	+ 1 (AA)	- 1 (BB)
3	+ 1	- 1	+ 1 (AA)	- 1 (BB)
4	- 1	+ 1	+ 1 (BB)	- 1 (AA)
5	- 1	+ 1	+ 1 (BB)	- 1 (AA)
Total	+ 1	- 1	+ 5	- 5

Tableau 1 : Exemple hypothétique de ségrégation transgressive, due à l'action complémentaire de gènes à effets additifs (tableau inspiré de Rieseberg *et al.* 1999).

Ainsi les hybrides homoploïdes pourraient coloniser des milieux extrêmes par rapport à celui de leurs parents, et la sélection pour les génotypes transgressifs peut aisément conduire à la spéciation. Cela semble être le cas des trois espèces *H. anomalus*, *H. deserticola* et *H. paradoxus* qui sont inféodées respectivement à des dunes de sable en Arizona, au Nevada et dans l'Utah, à des déserts en Arizona et dans l'Utah, et à des marais salins au Texas et au Nouveau Mexique (Figure 7). A noter que même si les espèces parentales vivent partiellement en sympatrie, *H. annuus* se développe sur des sols plutôt argileux et *H. petiolaris* sur des sols plus sableux.

L'isolement reproducteur, en termes de fécondité, entre les trois espèces hybrides et les deux parents est élevé, comme l'indique les études de fertilités (Figure 8, Lai *et al.* 2005). Il est intéressant de noter que les espèces hybrides sont plus fertiles entre elles qu'avec leur propre parent. Dans la nature, *H. anomalus*, *H. deserticola* et *H. paradoxus* n'ont aucune chance de se rencontrer et aucune barrière reproductive spécifique n'a eu besoin d'être mise

en place. Inversement les deux espèces parentales *H. annuus* et *H. petiolaris* sont très peu fertiles entre elles, mais leur répartition géographique fait qu'elles se sont hybridées plusieurs fois et ont données naissance à trois espèces. Des origines multiples ont été mises en évidence par des études chloroplastiques chez *H. anomalus*, qui aurait au moins trois origines distinctes (Schwarzbach et Rieseberg 2002), et *H. deserticola* qui pourrait également avoir trois origines distinctes, même si l'hypothèse d'introgression postérieure à une hybridation unique n'est pas à rejeter (Gross *et al.* 2003).

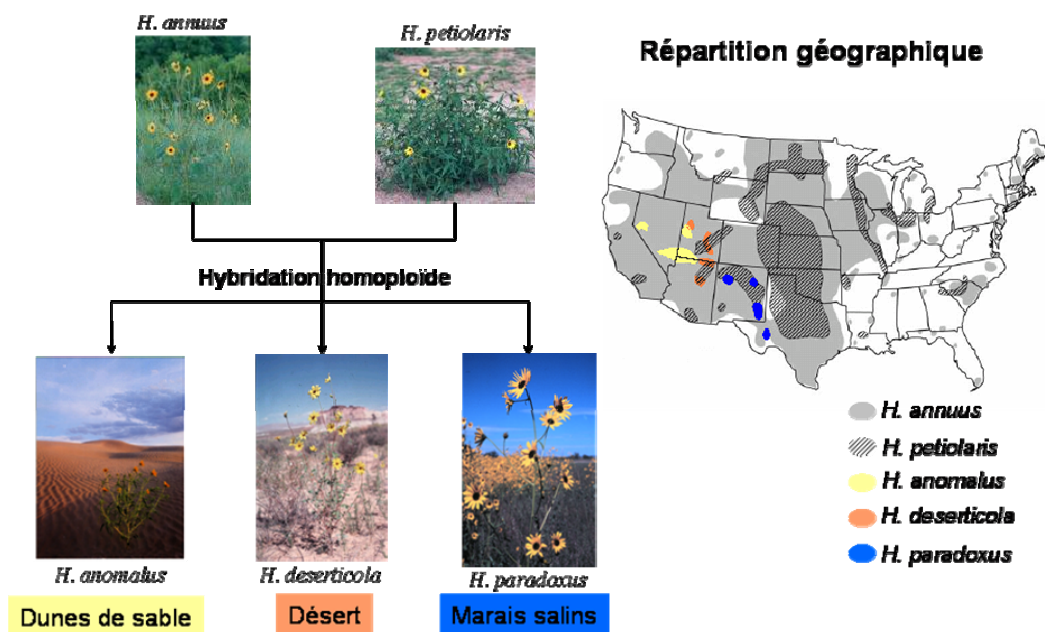


Figure 7 : Les deux espèces parentales *H. annuus* et *H. petiolaris* aux aires de répartition très larges aux Etats-Unis et leurs descendants hybrides *H. anomalus*, *H. deserticola* et *H. paradoxus*, inféodés à trois types de milieux extrêmes.

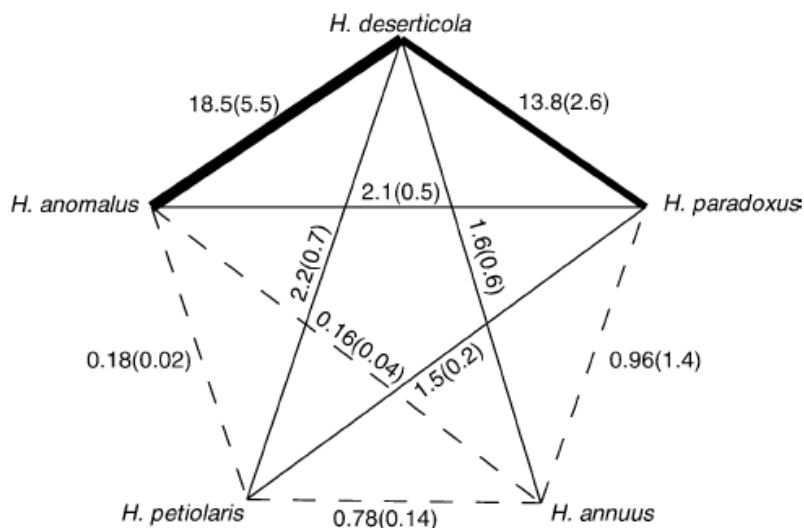


Figure 8 : Pourcentage (moyenne et erreur type) de graines fertiles de tous les croisements possibles entre les espèces parentales *H. annuus* et *H. petiolaris* et leur descendants hybrides *H. anomalus*, *H. deserticola* et *H. paradoxus*. L'épaisseur des lignes est proportionnelle à la capacité de se croiser. Les lignes pointillées indiquent que certains individus issus de ce croisement étaient complètement stériles. (Figure tirée de Lai *et al.* 2005).

3. La biologie et l'histoire de l'espèce *H. paradoxus*

Helianthus paradoxus est donc une espèce sauvage de tournesol annuelle et allogame, actuellement menacée et protégée. Elle a été décrite la première fois par Heiser en 1958. *H. paradoxus* n'apparaît pas dans les collections du 19^{ème} siècle issues de prospections dans les zones où l'espèce vit actuellement. Son statut d'espèce a longtemps été contesté, mais *H. paradoxus* est maintenant reconnue comme une espèce hybride issue d'un croisement entre *H. annuus* et *H. petiolaris*, les deux espèces sauvages de tournesols les plus répandues en Amérique du Nord. En effet *H. paradoxus* est morphologiquement uniforme et possède beaucoup de caractères intermédiaires entre ceux de ses parents, contrairement aux individus issus de croisements sporadiques entre *H. annuus* et *H. petiolaris* dont la gamme des phénotypes est très variée (Heiser 1947, 1958). Une étude effectuée à l'aide d'allozymes et de sites de restrictions appuie cette thèse : les allèles de *H. paradoxus* sont une combinaison de ceux de *H. annuus* et *H. petiolaris* (Rieseberg *et al.* 1990).

H. paradoxus est donc inféodée à des marais salins au Nouveau Mexique et au Texas. Les concentrations en sodium dans ce milieu vont de 2 000 à 20 000 ppm (partie par million), alors que les deux espèces parentales sont sensibles au sel et évoluent dans des sols où la concentration en sodium est comprise entre 20 à 150 ppm (Welch et Rieseberg 2002a). Les populations de *H. paradoxus*, sur chacun des 10 sites où on les retrouve, varie entre 5 000 et 10 000 individus (Rieseberg, communication personnelle). Ces sites sont principalement répartis le long de deux fleuves, le Pecos River et le Rio Grande.

L'âge de l'espèce *Helianthus paradoxus* est très controversé : Turner (1981) explique la répartition géographique restreinte de l'espèce (qui n'en est pas une selon lui) par l'apparition tardive de *H. paradoxus* : elle serait due à un système d'irrigation mis en place à Fort Stockton au Texas, après la seconde guerre mondiale. Heiser (1949) a émis l'hypothèse selon laquelle la large répartition de *H. annuus* en Amérique du Nord n'est que la conséquence d'une dispersion d'origine humaine. Schemske (2000) s'appuie sur cette hypothèse pour dire, que l'homme n'est pas étranger à l'apparition de *H. paradoxus*. Cependant une étude plus récente de Welch et Rieseberg (2002b) basée sur une analyse de microsatellites chez les trois espèces, suggère que l'origine de *H. paradoxus* remonte à entre 75 000 et 208 000 ans. Le temps a été estimé en nombre de générations, ce qui revient à un nombre d'années chez une espèce annuelle, à l'aide de l'équation : $T_D = \frac{D_1}{2w} - \frac{V_0}{w}$ (Zhivotovsky 2001), où D_1 est la moyenne sur tous les locus étudiés du carré de la différence

en terme de nombre de répétitions, w le taux de mutation estimé ici à 2.10^{-4} par locus et par génération et V_0 la moyenne sur tous les locus de la variance du nombre de répétition de la population ancestrale. V_0 n'étant pas connue elle est estimée par $V_{\min} = 0$ et V_{\max} qui est la moyenne de la variance du nombre de répétitions intra-population pour deux populations comparées deux à deux. Si le taux de mutation est sous-estimé, l'âge de l'espèce le sera aussi. Le chiffre minimum de 75 000 ans calculé prouve que *H. paradoxus* était déjà présente avant toute présence humaine en Amérique du Nord

Les différentes populations de *H. paradoxus* semblent être le fruit d'un unique événement d'hybridation (Rieseberg *et al.* 1990 et Welch et Rieseberg 2002b). Mais l'ensemble des individus des populations de *H. paradoxus* génotypés est fixé pour l'haplotype chloroplastique le plus fréquent de *H. annuus*. Si ce résultat suggère que *H. annuus* est probablement le parent maternel de *H. paradoxus*, cela n'exclut pas l'hypothèse d'hybridations multiples, d'autant plus que l'étude sur le nombre d'origines chez *H. anomalous* (Schwarzbach et Rieseberg 2002) a souligné le fait qu'il n'était pas clairement possible de mettre en évidence des signatures d'origines multiples aux niveaux des microsatellites, quand bien même on en trouvait au niveau chloroplastique.

Enfin, si *H. paradoxus* est le fruit d'un unique événement d'hybridation, le niveau de diversité génétique plus élevé dans les populations du sud du Nouveau Mexique et du Texas, n'est pas en adéquation avec une dispersion passive des akènes le long des fleuves de la région. Welch et Rieseberg (2002b) avancent l'hypothèse que la dispersion se serait faite par de larges mammifères comme des bisons.

4. Les outils dont on dispose et les expériences réalisées autour du complexe *H. annuus* et *H. petiolaris*

Le modèle biologique tel qu'il a été présenté précédemment possède plusieurs avantages :

1- *H. annuus* est l'espèce de tournesols qui a été domestiquée. Du fait de l'importance de cette culture, de nombreux outils ont été développés : cartes génétiques (la plus complète étant Tang *et al.* 2002), librairie d'ADNc, bases EST (<http://cgpdb.ucdavis.edu>).

2- Les espèces parentales du système existent toujours, autorisant des croisements manuels entre elles pour obtenir des populations d'hybrides synthétiques.

3- Pour *H. paradoxus* le milieu naturel de l'espèce est aisément reproductible pour faire des expériences en serre par exemple.

L'hybridation homoploïde a été très étudiée au sein du complexe des cinq espèces *H. annuus* et *H. petiolaris*, *H. anomalous*, *H. deserticola* et *H. paradoxus*. En plus de populations

naturelles de ces cinq espèces, des populations BC₂ d'hybrides synthétiques ont été construites (Lexer *et al.* 2003 ; Rieseberg *et al.* 2003, Rosenthal *et al.* 2005) (Figure 9).

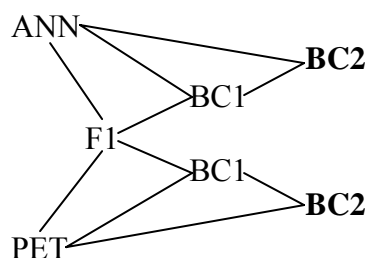


Figure 9 : Construction de populations d'hybrides synthétiques BC₂ à partir d'un individu *H. annuus* (ANN) et *H. petiolaris* (PET) issus de populations naturelles.

La plupart des caractéristiques phénotypiques des espèces hybrides naturelles peuvent être retrouvées chez les hybrides synthétique (BC₂PET : Rieseberg *et al.* 2003, BC₂ANN : Rosenthal *et al.* 2005). Ce qui est d'autant plus intéressant, c'est que la même combinaison de segments chromosomiques parentaux présente chez les hybrides naturels est nécessaire pour générer des phénotypes extrêmes chez les hybrides synthétiques (Rieseberg *et al.* 2003).

Ainsi, en seulement trois générations, il est possible d'obtenir des individus adaptés à des milieux semblables à ceux dans lesquels vivent les espèces hybrides. Dans ce complexe d'espèces, l'hybridation favorise l'adaptation à de nouveaux environnements et les hybrides synthétiques apparaissent comme étant de bons prédictors de la structure des hybrides naturels.

5. Quelles méthodes pour détecter et comprendre les bases génétiques de l'adaptation au milieu salin de *H. paradoxus* ?

Ma thèse a consisté à caractériser les bases génétiques de l'adaptation de *H. paradoxus* au milieu salin par différents types d'approches :

- Je me suis intéressée aux patrons de diversité génétique dans des populations naturelles de *H. paradoxus* et des deux espèces parentales dans les régions voisines de QTL détectés dans une population d'hybride synthétique BC₂, j'ai mis en évidence une baisse de diversité autour de ces QTL chez l'espèce hybride et non chez ses parents (premier chapitre).

- Une intensification de marquage microsatellites dans ces régions montre que le patron de diversité est en mosaïque et non uniformément faible. Cependant l'ordre des marqueurs utilisés n'étant pas connu avec certitude chez l'espèce hybride, j'ai développé une méthode basée sur le déséquilibre de liaison entre des marqueurs génotypés pour reconstruire une carte de déséquilibre de liaison (deuxième chapitre). Ceci a permis de détecter un

ensemble de marqueurs en fort déséquilibre de liaison dans les régions des QTL des populations de *H. paradoxus* étudiées.

- Parallèlement j'ai participé à une étude physiologique visant à comprendre le mécanisme d'adaptation à la salinité de *H. paradoxus*. L'espèce hybride se porte mieux en milieu salin, ses feuilles sont plus succulentes, avec une concentration en sodium et sulfate très élevée dans ses tissus. Le sodium semble remplacer d'autres cations comme le potassium, le calcium et le magnésium, ce qui permettrait à *H. paradoxus* de garder une pression osmotique constante (article de Karrenberg *et al.* 2006 en annexe). Cette étude s'est poursuivie par une étude d'expression génique, afin de comparer l'expression de 10 gènes candidats entre traitements salins (quatre traitements différents ont été appliqués) chez les trois espèces de tournesols. Des différences d'expression ont été trouvées pour certains gènes selon le traitement ; le patron d'expression est variable entre espèces et semble confirmer le rôle de certains de ces candidats dans la tolérance à la salinité de *H. paradoxus* (troisième chapitre).

CHAPITRE 1

Signature génétique de la sélection autour de 3 QTL
de survie dans le génome de *Helianthus paradoxus*

Signature génétique de la sélection autour de 3 QTLs de survie dans le génome de *Helianthus paradoxus*

La sélection laisse différents types de signature sur le génome, notamment une réduction de la diversité (Maynard Smith et Haigh, 1974) et une augmentation de la différenciation entre populations (Lewontin et Krakauer, 1973) aux locus neutres génétiquement liés à des gènes sous sélection directionnelle. Nous avons testé si nous pouvions utiliser ces propriétés pour détecter une signature de la sélection autour de 3 QTLs potentiellement impliquée dans la spéciation.

Dans notre système biologique, l'hybridation de deux espèces parentale, *Helianthus annuus* et *H. petiolaris* a donné naissance à l'espèce *H. paradoxus* adaptée aux marais salins.

Une étude de recherche de QTL de survie en milieu salin et d'absorption en ions a été menée (Lexer et al. 2003) sur une population de BC2 issue d'un croisement entre *H. annuus* et *H. petiolaris* recroisé pour deux générations à un individu *H. petiolaris*, afin d'augmenter la fertilité et donc le nombre de graines, puis de plantes de la population. La population d'hybrides synthétiques ainsi construite a été transplantée dans le milieu naturel de *H. paradoxus* et 14 QTL d'ions minéraux et trois QTL de survie ont été identifiés (Lexer et al. 2003). Ces trois QTL de survie représentent des régions potentielles soumises à sélection dans les populations naturelles d'hybrides.

Ceci constituait un modèle de choix pour tester s'il était possible de détecter une signature de la sélection dans le génome autour de régions impliquées dans l'adaptation, c'est-à-dire dans une meilleure performance des individus dans un milieu donné. Nous avons donc comparé le niveau de diversité autour des trois QTL de survie de l'espèce *H. paradoxus* à celui de régions potentiellement neutres. Pour cela nous avons génotypé deux microsatellites liés à chacun des trois QTL (microsatellites PSR pour Putative Selectionned Regions, en rouge sur la carte génétique, Figure 1.1) et six autres localisés dans des régions du génome où aucun QTL de survie, ni d'ions minéraux n'avait été détecté (microsatellites PUR pour Putative Unselectionned Regions, en bleu sur la Figure 1.1). La diversité génique a été comparée entre les deux catégories de microsatellites, mais également entre l'espèce hybride et chacun de ses parents. Nous avons travaillé sur trois populations naturelles pour chacune des trois espèces, en choisissant des populations parentales relativement proches géographiquement des populations hybrides (Figure 1.2).

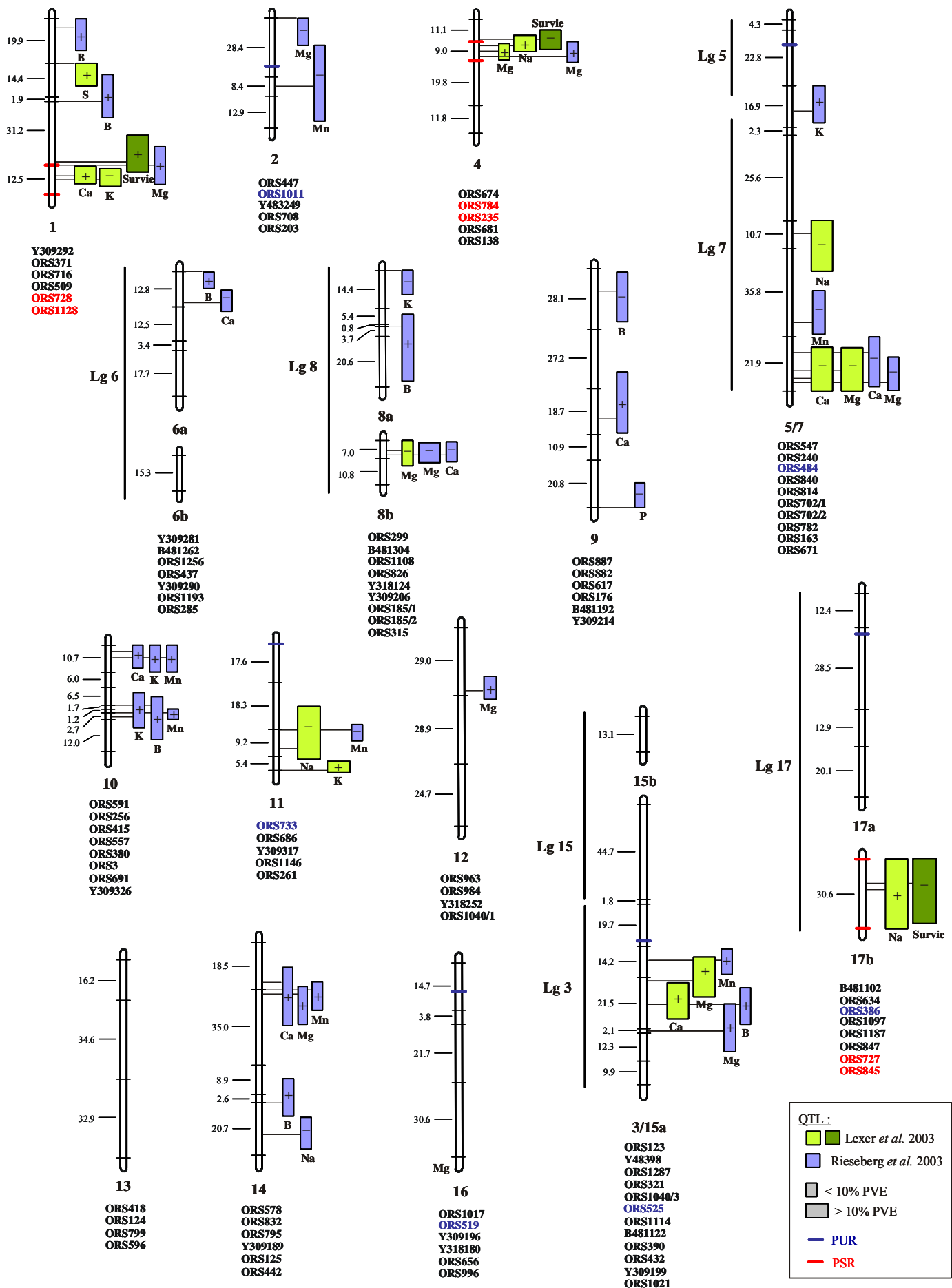


Figure 1.1 : Carte génétique d'une population issue d'un croisement interspécifique de deuxième génération entre *H. annuus* et *H. petiolaris*. Les QTL en vert ont été détectés dans une population de BC₂ (172 individus) transplantés dans le milieu naturel de *H. paradoxus* (Lexer et al. 2003), alors que ceux en bleu ont été mis en évidence dans une population BC₂ (384 individus) issus du même croisement et cultivés en serre (Rieseberg et al. 2003). La position des marqueurs est indiquée sur chaque groupe de liaison par des traits horizontaux avec les distances génétiques (cM) les séparant sur la gauche. Les noms des marqueurs commençant par ORS font référence à des microsatellites isolés chez *H. annuus* (Tang et al. 2002), les autres sont des marqueurs AFLP (données non publiées). Les marqueurs en rouge correspondent aux microsatellites situés dans les régions PSR (Putative Selected Regions) et les marqueurs en bleu à ceux des régions PUR (Putative Unselected Regions). La position des QTL est représentée par des lignes horizontales, leur magnitude et l'intervalle de LOD sont indiqués par des rectangles de largeur et hauteur variables. Le symbole +/- dans chaque rectangle donne la direction de l'effet de l'allèle de *H. annuus* dans la population de cartographie BC₂.

Répartition géographique des populations

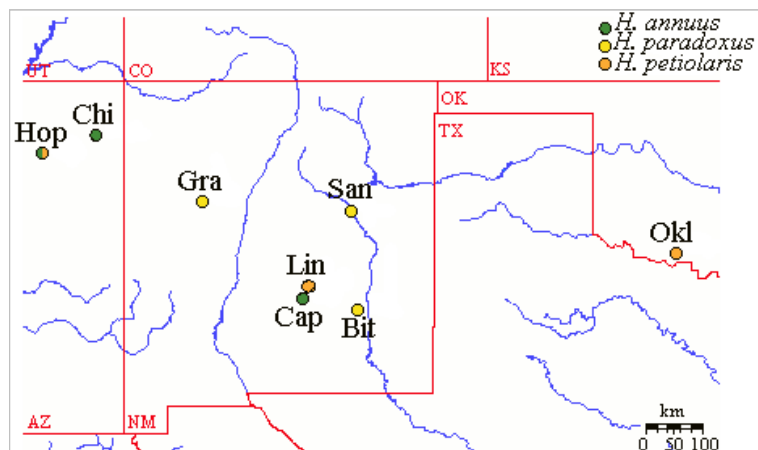


Figure 1.2 : Répartition des populations de notre étude. Les trois populations de l'espèce hybride *H. paradoxus* PARgra, PARsan et PARbit sont situées au Nouveau Mexique (NM). Deux populations de l'espèce parentale *H. annuus* ANNchi et ANNhop proviennent de l'Arizona (AZ), tandis que la troisième population ANNCap vient au Nouveau Mexique. La seconde espèce parentale *H. petiolaris* a une population originaire de l'Arizona PEThop, une autre du Nouveau Mexique PETLin et une troisième de l'Oklahoma PETokl. En bleu sont représentés les principaux fleuves et en rouge les frontières des différents états.

Microsatellite signature of ecological selection for salt tolerance in a wild sunflower hybrid species, *Helianthus paradoxus*

CÉCILE EDELIST,*†CHRISTIAN LEXER,‡CHRISTINE DILLMANN,†DELPHINE SICARD† and LOREN H. RIESEBERG*

*Department of Biology, Indiana University, Bloomington, IN 47405, USA, †UMR de Génétique Végétale, INRA/Université Paris XI/CNRS/INA PG, Ferme du Moulon, 91190 Gif-sur-Yvette, France, ‡Jodrell Laboratory, Royal Botanic Gardens, Kew, Richmond, Surrey, TW9 3AB, UK

Abstract

The hybrid sunflower species *Helianthus paradoxus* inhabits sporadic salt marshes in New Mexico and southwest Texas, USA, whereas its parental species, *Helianthus annuus* and *Helianthus petiolaris*, are salt sensitive. Previous studies identified three genomic regions – survivorship quantitative trait loci (QTLs) – that were under strong selection in experimental hybrids transplanted into the natural habitat of *H. paradoxus*. Here we ask whether these same genomic regions experienced significant selection during the origin and evolution of the natural hybrid, *H. paradoxus*. This was accomplished by comparing the variability of microsatellites linked to the three survivorship QTLs with those from genomic regions that were neutral in the experimental hybrids. As predicted if one or more selective sweeps had occurred in these regions, microsatellites linked to the survivorship QTLs exhibited a significant reduction in diversity in populations of the natural hybrid species. In contrast, no difference in diversity levels was observed between the two microsatellite classes in parental populations.

Keywords: *Helianthus*, hybridization, microsatellites, salt tolerance, signature of natural selection, speciation

Received 19 April 2006; revision accepted 25 July 2006

Introduction

A central interest in evolutionary biology is to understand the genetic basis of population divergence and adaptation. This issue can be addressed directly by testing the fitness effects of allelic variants in the wild (e.g. Bradshaw & Schemske 2003; Lexer *et al.* 2003; Weinig *et al.* 2003) or indirectly through the search for molecular signatures of positive selection from genome scans (Harr *et al.* 2002; Schlötterer 2002). This latter approach not only makes it possible to infer the strength and direction of historical bouts of selection, but it also provides a means for identifying genes of interest to ecology, agriculture, and medicine.

Selection leaves several different kinds of signatures on patterns of molecular variation. Perhaps the most striking of these is the reduction of diversity at neutral sites linked

to loci under directional selection. When an advantageous mutation appears at a very low frequency in a population, it may rapidly increase in prevalence, resulting in the loss of less fit alleles. Due to linkage or ‘genetic hitchhiking’, variability is predicted to be lowest near the target of selection and to rise with increasing genetic distance from the selected site (Maynard Smith & Haigh 1974). The physical size of the hitchhiked fragment will depend on the selection coefficient of the favourable mutation and on the local recombination rate; strong selection and low recombination will increase the size of affected regions. Hence hitchhiking increases correlations in diversity levels among linked loci (Kim & Stephan 2002) and creates a local skew in the frequency spectrum (first tested by Tajima 1989). Directional selection may also increase genetic differentiation between populations at selected loci and adjacent genomic regions (Lewontin & Krakauer 1973).

Numerous statistics have been developed to quantify levels of genetic diversity and distance for tests of selection,

Correspondence: Cécile Edelist, Fax: 33 1 69 33 23 40; E-mail: edelist@moulon.inra.fr

including nucleotide diversity π (Nei & Li 1979), gene diversity H_e (Nei 1978), F_{ST} -based estimators (first introduced by Lewontin & Krakauer 1973), and the $lnRV$ and $lnRH$ statistics (Schlötterer 2002; Schlötterer & Dieringer 2005). The latter were recently developed to screen for reductions in variability at microsatellite loci.

However, confident detection of selection remains challenging because changes in variability or differentiation may have multiple causes. As a consequence, a broad array of methods have been developed to minimize false positives caused by differences in mutation and/or recombination rates among loci or by genome-wide effects on variation due to population size expansions or contractions, sampling design, population structure, and population admixture (reviewed in Storz 2005). These methods include both model-based approaches that account for population structure (e.g. Beaumont & Nichols 1996; Beaumont & Balding 2004) and model-free approaches based on the distribution of summary statistics (e.g. Schlötterer 2002). Evaluation of these methods by computer simulation indicates that loci detected by multiple tests (each with different parameters and assumptions) represent robust candidates for further study (Schlötterer & Dieringer 2005).

As predicted by theory (Barton 1999; McKay & Latta 2002; Le Corre & Kremer 2003), selective sweeps appear to be frequent in genomic regions that contain quantitative trait loci (QTLs) or genes known to be under directional selection. These include sweeps associated with plant domestication (Vigouroux *et al.* 2002, 2005; Clark *et al.* 2004, 2006; Burke *et al.* 2005), animal breeding (Wiener *et al.* 2003), and parasite resistance to antibiotics (Wootton *et al.* 2002; Nair *et al.* 2003). The signature of selection also appears to be widespread in natural populations of organisms as diverse as fruit flies (Harr *et al.* 2002; Kauer *et al.* 2003), mice (Storz & Nachman 2003; Storz & Dubach 2004), salmon (Vasemägi *et al.* 2005) and humans (Kayser *et al.* 2003; Storz *et al.* 2004). Here, we test for the signature of ecological selection around major QTLs involved in adaptation to saline stress during the origin and evolution of a wild, annual sunflower, *Helianthus paradoxus*.

Helianthus paradoxus belongs to a complex of five annual sunflower species. Three of these (*Helianthus anomalus*, *Helianthus deserticola* and *H. paradoxus*) are stabilized diploid hybrid derivatives of the same two parental species, *Helianthus annuus* and *Helianthus petiolaris* (Rieseberg 1991). A remarkable feature of the hybrid neospecies is their adaptation to severe environments: sand dunes, deserts and salt marshes. The two parental species cannot survive in the salt marsh habitat of *H. paradoxus* (Lexer *et al.* 2003) and do not grow well in the sand dune (Ludwig *et al.* 2004) and desert habitats (Gross *et al.* 2004) of *H. anomalus* and *H. deserticola*, respectively. This implies that despite their outcrossing breeding system, gene flow between the hybrid lineages and their parental taxa must have been low relative

to the strength of environmental selection during the establishment of the hybrid species. The necessary reduction in gene flow appears to have been aided by karyotypic restructuring during hybrid speciation (Rieseberg *et al.* 1995; Lai *et al.* 2005). Despite the karyotypic changes, the species are closely related and synteny extends over large genomic regions of all 17 linkage groups (Lai *et al.* 2005). As a consequence, it is possible to recreate many aspects of the genotype and phenotype of the hybrid species by crossing contemporary populations of the parental species and tracing parentage and phenotypic effects in synthetic hybrid lineages (Rieseberg *et al.* 1996; Rieseberg *et al.* 2003; Rosenthal *et al.* 2005). These synthetic hybrids also represent valuable experimental material for locating QTLs or candidate genes responsible for adaptation to extreme habitats currently occupied by the hybrid species (Lexer *et al.* 2003; Lexer *et al.* 2004).

In the present study we focused on three QTLs that had a significant effect on the survivorship of BC_2 synthetic hybrids of *H. annuus* \times *H. petiolaris* that were grown in the brackish salt marsh habitat of *H. paradoxus* in New Mexico (Lexer *et al.* 2003). To determine whether these same regions were also under selection during the origin and/or subsequent evolution of *H. paradoxus*, we compared levels of polymorphism and differentiation for six microsatellite markers close to the three survivorship QTLs with six from elsewhere in the genome.

Materials and methods

Study organisms and populations

The progenitor species (*Helianthus annuus* and *Helianthus petiolaris*) are abundant throughout the central and western USA. In contrast, *Helianthus paradoxus* is restricted to a handful of brackish salt marshes in New Mexico and southwestern Texas and is listed as an endangered species by the US government. We collected achenes (one-seeded fruits) from three populations of *H. paradoxus* from across its range in the fall of 2001 (Table 1). In addition, three populations of each parental species were sampled from the same general area (Table 1); we assumed that geographically local populations would be most similar genetically to the ancestral parental populations that gave rise to *H. paradoxus*. Note that while the three species are broadly sympatric, because of differences in habitat preferences they typically are parapatric or allopatric locally. Indeed, we know of only two locations where habitat variation is sufficiently fine-grained that *H. paradoxus* grows intermixed with one of its parental species (*H. annuus*).

For *H. paradoxus*, 64 achenes from each population were germinated in the Indiana University greenhouses and fresh leaf tissue was collected for DNA extraction. Genomic DNA from the parental species was derived from a previous study (Welch & Rieseberg 2002). Sample sizes ranged from 14 to 25 individuals per population for the parental species (Table 1).

SELECTIVE SWEEPS NEAR THREE SURVIVORSHIP QTLs 4625

Table 1 Populations, sample sizes n , gene diversity H_e , allelic richness R_s and variance of repeat number V of the parental species *Helianthus annuus* and *Helianthus petiolaris* and their hybrid derivative *Helianthus paradoxus*, at Putative Selected Regions (PSRs) and Putative Unselected Regions (PURs). For comparative purposes, we also included gene diversity H_e corrected for sample size over 17 microsatellites from Welch & Rieseberg (2002). Collections were made by Loren H. Rieseberg (LHR) or Mark E. Welch (MEW)

Name	n	H_e		R_s		V		H_e (Welch & Rieseberg 2002)	Location	Reference
		PSR	PUR	PSR	PUR	PSR	PUR			
<i>H. annuus</i>										
ANNcap	18	0.86	0.79	9.00	7.96	66.95	44.30	0.77	Capitan, Lincoln Co., NM	MEW 1-1998
ANNchi	25	0.85	0.83	8.89	8.74	81.71	56.97	0.77	Chinle, Apache Co., AZ	LHR 1297
ANNhop	25	0.78	0.84	6.30	7.44	64.07	52.41	0.73	Hopi Indian Reservation, AZ	LHR 1294
<i>H. paradoxus</i>										
PARgra	64	0.37	0.52	3.06	3.53	2.32	18.85	0.42	Grants, Cibola Co., NM	LHR 1300
PARsan	64	0.35	0.49	2.59	3.00	7.46	23.39		Santa-Rosa, Guadalupe Co., NM	LHR 1302
PARbit	64	0.67	0.89	5.61	9.59	12.40	53.99	0.76	Bitterlake NWR, Chaves Co., NM	LHR 1303
<i>H. petiolaris</i>										
PETHop	24	0.79	0.82	7.45	8.43	30.33	61.82	0.82	Hopi Indian Reservation, AZ	LHR 1283
PETlin	14	0.83	0.83	7.77	8.45	29.22	49.34	0.77	Lincoln Co., NM	MEW 10-1998
PETokl	25	0.73	0.77	6.91	9.27	11.36	115.97	0.82	Cotton Co., OK	LHR 1220

Microsatellite choice and genotyping

Because synteny relationships are mostly conserved across the three species (Lai *et al.* 2005), we were able to employ a detailed map of the cultivated sunflower, *H. annuus* (Tang *et al.* 2002), as well as a second-generation backcross (BC_2) map of *H. annuus* \times *H. petiolaris* (Lexer *et al.* 2003), to identify markers within the 1-LOD confidence intervals of the QTLs of interest. Note that the BC_2 map was also employed for mapping the survivorship and mineral ion uptake QTLs that are the focus of the present study (Lexer *et al.* 2003). We chose six microsatellites (ORS728, ORS1128, ORS235, ORS784, ORS727 and ORS845) that were linked to the three survivorship QTLs on linkage groups (LG) 1, 4 and 17b (hereafter referred to as Putative Selected Regions or PSRs). For comparative purposes, six microsatellite loci [ORS386 (LG17a), ORS484 (LG5), ORS519 (LG16), ORS525 (LG3), ORS733 (LG11) and ORS1011 (LG2)] were chosen from regions that had no effect on survivorship in the *H. paradoxus* habitat (i.e. Putative Unselected Regions or PURs) and that did not contain mineral uptake QTLs whether tested under natural (Lexer *et al.* 2003) or greenhouse conditions (Rieseberg *et al.* 2003). Primer sequences are available at <http://cropandsoil.oregonstate.edu/knapp-lab/>. The 12 microsatellites chosen were codominant, readily amplified in all three species, and present in a single copy.

Total genomic DNA was isolated using the DNeasy Plant Kit (QIAGEN 2001), from 30 to 40 mg of fresh leaf tissue collected in the greenhouse (3×64 *H. paradoxus* individuals). After extraction, DNA concentration was quantified using a TKO-100 fluorometer (Hoefer Scientific Instruments). Polymerase chain reactions (PCRs) were run in a 10- μ L total volume, following protocols described in

Burke *et al.* (2002). To reduce nonspecific amplification, touchdown PCR (Don *et al.* 1991) was used. Reactions were given an initial denaturation step of 3 min at 95 °C. This was followed by 10 cycles of 30 s at 94 °C, 30 s at the annealing temperature plus 10 °C (minus 1 °C at each cycle), and 45 s at 72 °C, and then 30 cycles of 30 s at 94 °C, 30 s at the annealing temperature, and 45 s at 72 °C. A final extension step cycle of 20 min at 72 °C was employed to ensure full-length amplification of all fragments. Annealing temperatures varied between 48 °C and 57 °C depending on the best average amplification. All PCRs were run in 96 Well Flexible Plates (Fisherbrand). Microsatellite fragments were separated by electrophoresis on an ABI PRISM 3700 capillary sequencer (Applied Biosystems 2000). PCR products from different loci were pooled on the basis of size or fluorophore. Between two and six PCR products were combined after 20-fold (forward labelled primer system) or 10-fold (M13 three-primer system for ORS845, ORS1011, and ORS1128) dilutions. Samples were then prepared by mixing 1 μ L of the diluted PCR pool with 9.8 μ L ddH₂O and 0.20 μ L GenSize R500 ROX standard sizes (GenPak), heated for 5 min at 95 °C, chilled on ice, and placed on the ABI PRISM 3700. Chromatographs were generated using GENESCAN 3.7 software (Applied Biosystems 2000) and microsatellite fragment lengths were scored with GENOTYPER 3.6 (PE Biosystems 2000).

Analysis of genetic diversity

Allele frequencies were computed for each locus and a principal components analysis (PCA) was performed on allele size to provide an initial overview of the data.

Levels of genetic diversity (or differentiation) were compared between the two classes of loci in each species using

four estimators: He , Rs , $\ln RV$ and $\ln RH$. Gene diversity or expected heterozygosity He (Nei 1978) corrected for sample size was estimated with the software MICROSATELLITE ANALYSER (MSA) (Dieringer & Schlötterer 2003). Allelic richness Rs (El Mousadik & Petit 1996), which is a measurement of number of alleles per locus corrected by the rarefaction index (Hurlbert 1971), was calculated with the software FSTAT (Goudet 2001). The ratio of variance in repeat number corrected for sample size, $\ln RV$, between two populations at a locus

$$\left(\text{here } \ln RV = \ln \frac{V_{\text{hybrid}}}{V_{\text{parent}}} \right)$$

(Schlötterer 2002) was calculated with the software MSA (Dieringer & Schlötterer 2003). This statistic assumes that mutations at microsatellite loci follow an unbiased stepwise model (Ohta 1973), in which mutations are confined to single repeat unit changes. However, this model is violated for two microsatellites, ORS1128 and ORS784 (some alleles had deletions in flanking regions), so these were excluded from analyses of $\ln RV$. The ratio of the expected heterozygosity, $\ln RH$

$$\left(\text{approximated by } \ln RH = \ln \frac{\left(\frac{1}{1 - H_{\text{hybrid}}} \right)^2 - 1}{\left(\frac{1}{1 - H_{\text{parent}}} \right)^2 - 1} \right),$$

was also calculated. Note that when a locus was monomorphic within a population, a single allele that differed by a single repeat unit change from the common allele was added to the data set so that both $\ln RV$ and $\ln RH$ could be calculated.

Mann–Whitney tests (JMP SAS Institute 1997) were used to test for differences in He and Rs among species. To determine within species whether the diversity indices He , Rs , $\ln RV$ and $\ln RH$ differed among populations or between PSRs and PURs, we employed a mixed linear model for each species: $y_{ijk} = \mu + \text{Class}_i + \text{Pop}_j + (\text{Class}_i \times \text{Pop}_j) + \varepsilon_{ijk}$, where $i = 1$ to 2 is the microsatellite class (PSRs vs. PURs), $j = 1$ to 3 refers to the population of a given species and, $k = 1$ to 6 represents the microsatellite loci from each class, with loci treated as independent replicates. The $\ln RV$ and $\ln RH$ statistics were calculated separately for comparisons of *H. paradoxus* with *H. annuus* and *H. paradoxus* with *H. petiolaris*, and ratios derived from comparisons with each of the three populations of a given parental species were considered to be independent replicates. Hence, $k = 1$ to 18 (three populations \times six loci) for comparisons among populations of *H. paradoxus*. To test for differences in diversity between classes of loci, we employed nonparametric tests (despite lower statistical power relative to their parametric counterparts) because assumptions of normality and equality of variance were violated by some of our data. The significance

of population, marker class, and interaction effects was tested by random permutations of the data (RDevelopment Core Team 2004). Population effects were tested by permuting microsatellites among populations, class effects by permuting microsatellites between classes, and interaction effects by permuting microsatellites globally. For marker class and interaction effects, the statistic

$$S = \frac{MST}{MSR}$$

was computed, whereas for the population effects

$$S = \frac{MST}{MSI}$$

was employed because populations were considered as a random effect (MST is the mean square of the effect being tested, MSR is the mean square of the residual and MSI is the mean square of the interaction). P values were calculated as the proportion of permutations with S values that exceed observed values.

Because levels of variation at markers linked to QTLs under selection in *H. paradoxus* are expected to be influenced by hitchhiking, we developed regression models for the diversity estimators He , Rs , $\ln RH$ and $\ln RV$ that considered the strength of selection on QTLs and the recombination distance of a given marker from the most likely QTL position: $y_{ij} = \mu + (1 - r_i)s_i + \varepsilon_{ij}$, where i is the microsatellite locus, j the hybrid species populations considered as replicates, and r the recombination rate. r was approximated here by the Haldane function

$$r = \frac{1}{2}[1 - e^{(-2d)}],$$

with d the distance of a microsatellite to the most likely QTL position ($d_{\text{ORS728}} = 1$ cM, $d_{\text{ORS1128}} = 13.5$ cM, $d_{\text{ORS235}} = 10$ cM, $d_{\text{ORS784}} = 1$ cM, $d_{\text{ORS727}} = 10.7$ cM, $d_{\text{ORS845}} = 19.9$ cM), and s is the selection coefficient estimated at ORS728 ($s = 0.126$), ORS784 ($s = -0.084$), and ORS727 ($s = -0.097$) (Lexer *et al.* 2003). For microsatellites located in PURs, r and s were approximated by $r = 0.5$ and $s = 0$.

Differentiation among populations

Pairwise F_{ST} values (Weir & Cockerham 1984) were estimated with the software MSA (Dieringer & Schlötterer 2003). Mean F_{ST} values for the two classes of microsatellites were compared among populations of the same species or populations of different species using Mann–Whitney tests (JMP SAS Institute 1997).

Blind test for microsatellite loci with reduced diversity

One possible approach for identifying loci that have been subject to selection within a population is to identify outliers

in the distribution of a given summary statistic under the null hypothesis of random genetic drift. In our data set, two classes of microsatellites, PSR and PUR, were defined a priori and tested for a difference in genetic diversity. Under a null hypothesis of pure drift, there should be no more difference between PSR and PUR for genetic diversity than expected by chance between two subsets of six microsatellites chosen at random from the genome. The difference between classes for the level of genetic diversity (He, Rs, lnRV and lnRH) was assessed by the summary statistic S (see above). All

$$n = C_{12}^6 = 924$$

possible permutations were tested and $S_{0.05}$ was computed as the threshold value for rejecting the null hypothesis at the probability level $\alpha = 0.05$. To control for false-positive results, the $n\alpha = 46$ significant permutations were stored and for each permutation, the class with the significantly lower diversity level was defined as the blind-PSR class. Thereafter, the number of times a given microsatellite appeared in the blind-PSR class was tabulated. If our hypothesis of selection near survivorship QTLs is correct, then we expect the blind-PSR class to be enriched with microsatellites belonging to the PSR class defined in terms of genomic location.

Admixture proportion

The software LEADMIX (Wang 2003) performs maximum-likelihood estimation of admixture proportions in a model where an ancestral species P_0 is split into two parental populations P_1 (here *H. annuus*) and P_2 (here *H. petiolaris*), which evolve independently before they contribute in genetic proportion p_1 and $(1 - p_1)$ to form a hybrid population P_h (here *H. paradoxus*). In our analysis, the admixture proportion p_1 was defined as the contribution of the species *H. annuus* to the hybrid species *H. paradoxus*. After the admixture event, P_1 , P_2 and P_h evolve independently (no gene flow) before a sample is taken from each species. From allelic frequencies in each species, LEADMIX estimates different parameters: the time before the admixture event, the time from the admixture event to present, and the admixture proportion. As time is measured in units of generations per effective population size ($2Ne$), there are two different estimates for the time before the admixture event, one for *H. annuus* (t_1) and one for *H. petiolaris* (t_2). Similarly, there are three estimates for the time after the admixture event: T_1 for *H. annuus*, T_2 for *H. petiolaris* and T_h for *H. paradoxus*. Differences between t_1 and t_2 or between T_1 , T_2 and T_h represent differences in effective population sizes, with higher estimates of T corresponding to lower relative population sizes. The estimates can be used to provide a rough approximation of the number of generations τ_A since the admixture event, compared to the total number of

generations $\tau_A + \tau_B$ since the divergence of the parental species *H. annuus* and *H. petiolaris*:

$$\frac{\tau_A}{\tau_A + \tau_B} = \frac{1}{1 + \frac{t}{T} \frac{N_B}{N_A}},$$

where N_B is the effective population size before the admixture event and N_A is the effective population size after the admixture event.

$$\frac{T_{PARENT}}{T_h} = \frac{N_h}{N_{PARENT}}$$

is the ratio of the effective population size of the hybrid species relative to that of the parental species.

Results

Microsatellite polymorphism over all individuals

The 12 microsatellite loci genotyped in this study have either di- or trinucleotide repeats, with motif lengths varying from 10 to 71 nucleotides and allele sizes from 111 to 492 nucleotides. The number of alleles detected across the 12 loci ranged from 147 alleles in *Helianthus paradoxus*, to 206 alleles in *Helianthus annuus*, to 218 alleles in *Helianthus petiolaris*, despite the fact that close to three times as many individuals were genotyped in *H. paradoxus* than its parental species. The polymorphism level for the three species is comparable to that reported by Welch & Rieseberg (2002). Also, the paucity of allelic variation in the hybrid species is concordant with observations from this earlier study (Table 1).

The two microsatellite classes, from PSRs and PURs, do not have intrinsic properties that might lead to reduced diversity at PSR loci. For example, allele numbers are often correlated with repeat numbers at a given locus because of increased slippage and reading frame shifts (Wierdl *et al.* 1997). In the present study, the average motif length was lower in the PSR (25 nucleotides) than in the PUR (42 nucleotides). However, the difference between the two microsatellite classes is not significant ($t_5 = 1.651$, $P = 0.080$). Also, interruptions within a motif and irregular repeats are known to decrease microsatellite variability (Petes *et al.* 1997; Brinkmann *et al.* 1998). In our study, microsatellites from the PURs were interrupted more frequently (ORS484, ORS519, ORS525 and ORS733 have interrupted motifs) than those linked to the survivorship QTLs (only ORS728 is interrupted). Thus, in the absence of selection, microsatellites from PURs would be predicted to have less diversity.

The distributions of allele frequencies are significantly different between species ($P < 0.01$ for all chi-squared tests) and among populations within each species ($P < 0.01$ for all tests). The most obvious difference among species is that microsatellite length variance is reduced in *H. paradoxus*

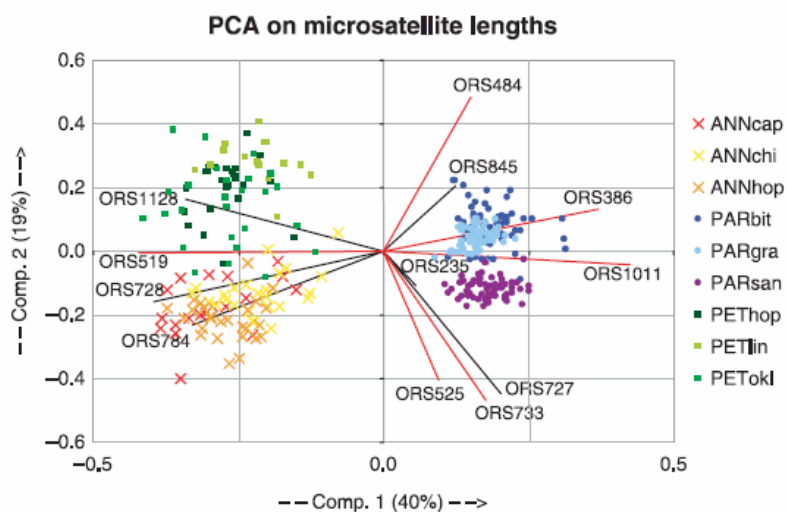


Fig. 1 Principal components analysis of allele size variation across the 12 microsatellite loci in *Helianthus annuus* (crosses), *Helianthus paradoxus* (closed circles), and *Helianthus petiolaris* (squares). The contribution of each microsatellite to each component is computed to maximize the difference between individuals. Vectors represented by black lines indicate the contribution of PSR microsatellites, whereas red line vectors show the contribution of PUR markers. The first two components explain 59% of the total variation for allele sizes.

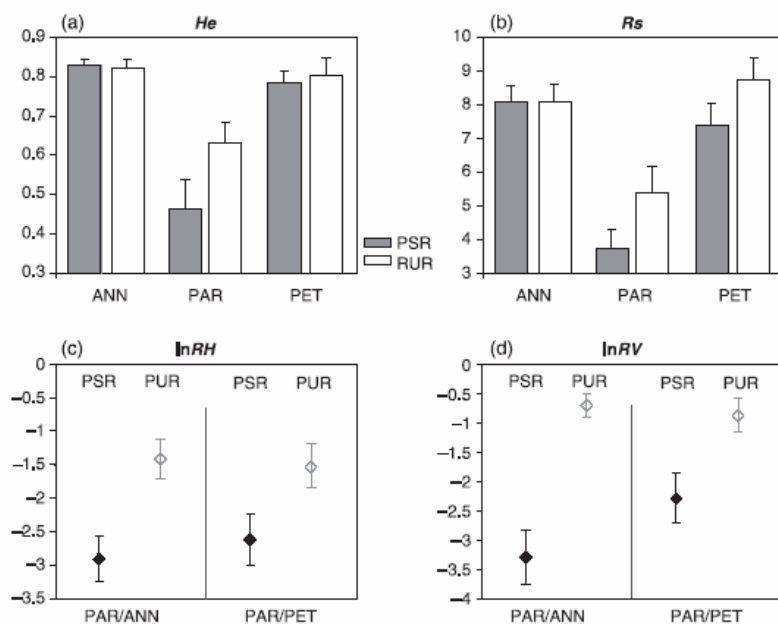


Fig. 2 Comparison of variability levels (means and standard errors) between microsatellites in PSRs (grey) and in PURs (white). (a) gene diversity, H_e ; (b) allelic richness, R_s ; (c) $\ln RH$ of *Helianthus paradoxus* over *Helianthus annuus* and over *Helianthus petiolaris*, respectively; (d) $\ln RV$ of *H. paradoxus* over each one of its parents. Note that standard errors do not take into account population effects and therefore cannot be used to test for differences between microsatellite classes in *H. paradoxus*.

relative to the parental species, as was previously reported by Welch & Rieseberg (2002).

A PCA of allele size variation across the 12 microsatellite loci clusters individuals from the same species together (Fig. 1). The PCA graphically illustrates the reduced allelic diversity in *H. paradoxus* populations relative to its parents. The first component separates the parental species from the hybrid taxon, whereas the second component distinguishes the parental species from each other. On this second component, the hybrid species' populations are projected between the parental species, a placement consistent with their hybrid origin.

Genetic variation among populations and between microsatellite classes

Gene diversity, H_e , and allelic richness, R_s , do not differ between parental species for either PSRs or PURs (Fig. 2a, b; $U_{H_e,PSR} = 316$, $P = 0.602$; $U_{H_e,PUR} = 354$, $P = 0.517$; $U_{R_s,PSR} = 312$, $P = 0.517$; $U_{R_s,PUR} = 190$, $P = 0.380$). By contrast, the hybrid species *H. paradoxus* has less diversity as measured by H_e or R_s than either parental species and for both PSR and PUR markers (Fig. 2a, b; $U_{H_e,PSR} = 208$, $P < 0.001$; $U_{H_e,PUR} = 260$, $P = 0.022$; $U_{R_s,PSR} = 200$, $P < 0.001$; $U_{R_s,PUR} = 242$, $P = 0.012$ for the comparisons between *H. paradoxus*

SELECTIVE SWEEPS NEAR THREE SURVIVORSHIP QTLs 4629

Table 2 Results from blind classifications of microsatellite loci performed over the 924 permutations for each of the four diversity estimators. Proportions in cells indicate the frequency at which the corresponding microsatellites were detected by the blind-test. Cells with crosses denote microsatellites that were discarded because they violated the stepwise mutation model in *Helianthus paradoxus*

		Status before blind-test	Status after blind-test	<i>He</i>	<i>Rs</i>	<i>lnRH</i>		<i>lnRV</i>	
				PAR	PAR	ANN	PET	ANN	PET
ORS728	LG1	PSR	blind-PSR	1	1	1	1	1	1
ORS1128	LG1	PSR	blind-PSR	0.79	0.90	0.75	0.88	X	X
ORS235	LG4	PSR	blind-PSR	1	1	1	0.96	1	1
ORS727	LG17b	PSR	blind-PSR	0.68	0.79	0.82		0.71	
ORS525	LG3	PUR	excluded	0.67		0.66	0.65		
ORS484	LG5	PUR	excluded	0.78		0.67	0.66		
ORS386	LG17a	PUR	blind-PUR						0.68
ORS733	LG11	PUR	blind-PUR						
ORS1011	LG2	PUR	blind-PUR						
ORS519	LG16	PUR	blind-PUR				0.65		0.54
ORS845	LG17b	PSR	excluded						
ORS784	LG4	PSR	excluded		0.56			X	X

Table 3 R^2 and P associated with linear regression models of four diversity estimators in *Helianthus paradoxus*. The models take into account the strength of selection on QTLs and distance of markers from the putative selected sites to better account for differences in variability among microsatellite loci, either with the initial microsatellite classes (upper rows) or with the blind-test markers only (lower rows)

		<i>He</i>	<i>Rs</i>	<i>ln RH</i> ANN	<i>ln RH</i> PET	<i>ln RV</i> ANN	<i>ln RV</i> PET
All markers	R^2	0.112	0.096	0.142	0.099	0.399	0.13
	P	0.046	0.066	<0.001	<0.001	<0.001	<0.001
blind-test markers	R^2	0.315	0.155	0.383	0.487	0.519	0.635
	P	0.004	0.024	<0.001	<0.001	<0.001	<0.001

and *H. annuus* and $U_{He,PSR} = 433$, $P = 0.002$; $U_{He,PUR} = 411$, $P = 0.014$; $U_{Rs,PSR} = 438.5$, $P = 0.001$; $U_{Rs,PUR} = 242$, $P = 0.012$ for the comparisons between *H. paradoxus* and *H. petiolaris*).

Within the two parental species, none of the factors tested (population, marker class or interaction) had a significant effect on either *He*, or *Rs*. By contrast, in the hybrid species, *H. paradoxus*, population and class significantly affected variability levels. The population effect was significant when tested on *He* ($S = 73.643$, $P = 0.014$), *Rs* ($S = 28.532$, $P < 0.001$) and both *lnRH* ratios ($S = 49.007$, $P = 0.015$ for *H. paradoxus* vs. *H. annuus* and $S = 49.007$, $P = 0.022$ for *H. paradoxus* vs. *H. petiolaris*). This effect is a consequence of the greater diversity of *H. paradoxus* from Bitter Lake relative to the two other populations. Permutations on microsatellite classes revealed significant differences between PSR and PUR microsatellites. However, despite the trend toward lower diversity at PSRs than PURs for all four diversity estimators (Fig. 2), significant differences between PSR and PUR microsatellites are found only for *Rs* ($S = 8.038$, $P = 0.050$) and *lnRV* of *H. paradoxus* vs. *H. annuus* ($S = 41.801$, $P = 0.017$). No interaction effects were found.

The blind search for loci that have been under selection deviated somewhat from the predictions based on the

locations of survivorship QTLs. Consistent with expectations, four out of the six microsatellites linked to survivorship QTLs were repeatedly found in the blind-PSR class (Table 2): ORS728 and ORS1128 (LG1), ORS235 (LG4) and ORS727 (LG17b). However, two microsatellites, ORS525 and ORS484, from the PUR class appear in the blind-PSR class for *He* (Fig. 3) and *lnRH*, but in the blind-PUR for *Rs* or *lnRV*. These two microsatellites are interrupted in their repeat motif and may have a lower mutation rate. Also, six microsatellites occur most commonly in the blind-PUR class (ORS386, ORS733, ORS845, ORS1011, ORS519, and ORS784), two of which (ORS845 and ORS784) were initially assigned to the PSR class. In summary, the blind search confirmed the initial assignments to PSR and PUR classes for eight of the 12 microsatellites. The four microsatellites that failed the blind-test were excluded from some of the analyses described below.

Within *H. paradoxus*, the linear regression models showed a significant linear relationship between *He*, *lnRH*, *lnRV* and $(1-r)s$ and marginal significance for comparisons with *Rs* (Table 3). This indicates that levels of variation are correlated with the strength of selection on QTLs in the wild and with the distance to the selected site. Values of R^2

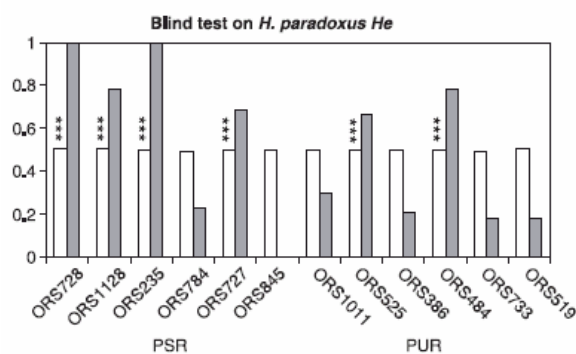


Fig. 3 Results of a blind-test performed on the genetic diversity estimator (H_e) in *Helianthus paradoxus*. For each microsatellite, the frequency of the marker in the class with the lower diversity is in white. The difference between classes for the level of genetic diversity was assessed by the summary statistic S , and $S_{0.05}$ was computed as the threshold value for rejecting the null hypothesis of no diversity differences between classes at the probability level $\alpha = 0.05$. The frequency at which S exceeds $S_{0.05}$ is in grey, stars indicate when S exceeds $S_{0.001}$.

Table 4 Mean F_{ST} s within (diagonal) or between species performed after the blind-test. Values above the diagonal are for microsatellites located in the blind-PSR, whereas those below the diagonal are derived from markers in the blind-PUR. * = $P < 0.05$

PUR/PSR	ANN	PAR	PET
ANN	0.080/0.038	0.433*	0.158
PAR	0.251*	0.211/0.257	0.478*
PET	0.132	0.273*	0.116/0.117

are higher and relationships between all four diversity statistics and $(1-r)s$ are significant when these models are applied to the subset of eight microsatellites identified by the blind-test (Table 3).

Differentiation among populations

Levels of genetic differentiation, as measured by F_{ST} for the blind-test microsatellites, did not differ between blind-PSRs and blind-PURs for comparisons among populations within species (Table 4). In contrast, for comparisons between the hybrid and parental species, mean F_{ST} was significantly greater for loci located in blind-PSRs than blind-PURs: *H. paradoxus* vs. *H. annuus* ($U = 1009$, $P < 0.001$) and *H. paradoxus* vs. *H. petiolaris* ($U = 969$, $P < 0.001$) (Table 4).

Admixture

Our analyses of demographic history employed the eight blind-test microsatellites only (Table 3). For each parameter of the admixture model applied by LEADMIX, there were six

different estimates: three populations of *H. paradoxus* times two microsatellite classes.

Estimates of time (generations/ N_e) before admixture differed according to microsatellite class for *H. petiolaris* [$t_{2(PUR)} = 0.074$; C.I. 0.045–0.113 and $t_{2(PSR)} = 0.136$; C.I. 0.093–0.195], but not for *H. annuus* (mean $t_1 = 0.115$; C.I. 0.080–0.167). Estimates of time after the admixture event did not significantly differ between microsatellite classes in the parental species and the estimates were $T_1 = 0.00053$ (C.I. < 0.00001–0.010) for *H. annuus*, and $T_2 = 0.00162$ (C.I. < 0.00001–0.011) for *H. petiolaris*. Those different estimates gave a ratio t/T ranging from 46 to 216; therefore, the relative number of generations after the admixture event ranged from 0.005 to 0.021, assuming N_e did not change dramatically after the admixture event. This implies that *H. paradoxus* arose recently relative to the origin of its parental species. Estimates of the time after admixture for *H. paradoxus* varied significantly among populations and between microsatellite classes ($0.157 \leq T_h \leq 0.470$). In all cases, T_h was significantly higher than T_1 or T_2 , the ratio

$$\frac{T_{PARENT}}{T_h}$$

ranged from 0.001 to 0.007, indicating that N_e is much lower in the hybrid species than its parents.

Interestingly, admixture proportions (i.e. the proportion of genes in the hybrid species that were inherited from *H. annuus*) depends on the microsatellite class. For the PUR class, the admixture proportions were 0.757, 0.809 and 0.708 for populations PARgra, PARsan and PARbit, respectively; thus *H. annuus* is the major contributor. For the PSR class, in contrast, the corresponding admixture proportions were 0.201, 0.004, and 0.325, and *H. petiolaris* is the primary donor.

Discussion

Species history

Our results are largely consistent with previous estimates of the phylogenetic and demographic history of *Helianthus paradoxus* (Rieseberg *et al.* 1990; Welch & Rieseberg 2002; Lai *et al.* 2005). The lower overall genetic diversity in *H. paradoxus* compared to its parents suggests that the species experienced a narrow bottleneck during its origin, a demographic event previously implied by large-scale karyotypic re-patterning detected in the hybrid lineage (Lai *et al.* 2005). Likewise, the estimated time since admixture, as computed by LEADMIX, suggests that *H. paradoxus* arose recently, a conclusion in agreement with previous microsatellite (Welch & Rieseberg 2002) and chloroplast DNA studies (Rieseberg *et al.* 1991), which place the origin of *H. paradoxus* within the past 210 000 years. Even if populations of the parental species were 10 times smaller before than after the hybridization event, the proportion of generations after the admixture

event never exceeds 20% of the total number of generations since the divergence of the parental species, which has been estimated at between 750 000 and 1 million years ago (Rieseberg *et al.* 1991).

Paradoxically, F_{ST} values between *H. paradoxus* and its progenitors are higher than between the two parental species themselves (Table 4). This result is most likely due to the large reduction in allelic diversity in *H. paradoxus*, which may inflate estimates of F_{ST} (Hedrick 2005). However, it might also be that *H. annuus* and *H. petiolaris* continue to exchange genes with each other, but not with *H. paradoxus*. *H. annuus* and *H. petiolaris* are partially sympatric and average genetic differentiation between them is less than for a pair of more closely related but historically allopatric congeners, *H. annuus* vs. *H. bolanderi* (Gardner 2004; Yatabe & Rieseberg, in review). In contrast, *H. paradoxus* comes into contact with its parental species only sporadically. Thus, despite a stronger sterility barrier between parental species than between the parental species and *H. paradoxus* (Lai *et al.* 2005), interspecific gene flow may be more important in the former.

The three hybrid populations differ significantly in diversity, with the most southerly population (Bitter Lake) exhibiting the greatest variability. This pattern, which is opposite that predicted by passive colonization along rivers, was also reported by Welch & Rieseberg (2002). The higher genetic diversity of the Bitter Lake population could be due to its large population size. Surprisingly, *H. paradoxus* from Bitter Lake is more similar to the Grants population than to Santa Rosa (Fig. 1), despite the fact that Bitter Lake and Grants are on different drainages and are separated by the tail end of the Rocky Mountain chain. The lack of congruence between geographical and genetic distance could be explained by multiple origins of *H. paradoxus* or by a single origin followed by long distance dispersal by animals. Welch & Rieseberg (2002) favoured the latter hypothesis since *H. paradoxus* had a single chloroplast DNA haplotype and was monophyletic for nuclear loci. Our results also favour this hypothesis since estimates of time since admixture are highly consistent across the three populations of *H. paradoxus*.

Testing for differences in diversity levels among markers

To test for the molecular signature of the adaptation to an extreme environment, we used four different diversity estimators H_e , R_s , $\ln RV$ and $\ln RH$ with the hope of minimizing biases associated with each individual statistic. For example, gene diversity H_e is sensitive to the mutation rate (or frequency of rare alleles), whereas allelic richness R_s is sensitive to population size changes caused by bottlenecks and admixture (Petit *et al.* 1998; Comps *et al.* 2001). The $\ln RV$ and $\ln RH$ statistics, which were developed for the analysis of microsatellite data, are robust to genome-wide

effects such as population expansions, bottlenecks, and admixture (Schlötterer & Dieringer 2005) and control for variation in mutation rates (Schlötterer 2002).

Thus, it is perhaps not surprising that the four statistics provide slightly different estimates of diversity differences between marker classes. While all four statistics show the same trend, with PSR markers exhibiting lower diversity than PUR markers (Fig. 2), these differences were significant only for R_s and $\ln RV$ for *H. paradoxus* over *H. annuus*. These two estimators were also most discriminating when performing the blind-test (Table 2). They both take into account the number of alleles rather than their frequencies and may therefore be more sensitive to bottlenecks (whether caused by selection or drift) than the other two estimators, H_e and $\ln RH$. A similar result was reported in European Beech, where a greater decrease in R_s than H_e was observed following postglacial colonization (Comps *et al.* 2001). These authors concluded that R_s might be more suited for identifying population historical processes.

The blind-test provided an alternative means for identifying loci with low relative diversity and explaining differences in diversity levels. For example, we were able to confirm the predicted reduction in diversity for four of six candidate PSR loci and show that reduced variation in two candidate PUR loci was likely due to a low mutation rate. More generally, the blind-test represents a promising method for identifying subsets of loci with reduced diversity, even in the absence of mapping data or other predictive information.

Signature of selection around three survivorship QTLs in *H. paradoxus*

The loss of diversity at markers associated with the three survivorship QTLs in *H. paradoxus* is consistent with the hypothesis that these regions have been subjected to one or more selected sweeps, presumably due to ecological adaptation to salt marsh conditions. This conclusion is supported by observations that the hybrid species is more differentiated from its parental species for PSR than PUR microsatellites. The fact that intraspecific F_{ST} s are not significantly different between PSR and PUR microsatellites suggests that the selective sweeps occurred prior to geographical expansion of the species or that selection was homogeneous in the three *H. paradoxus* populations.

An alternative hypothesis is that the observed distribution of variability across the *H. paradoxus* genome is a byproduct of one or more population bottlenecks, which are known to increase the variance in diversity among loci. In this scenario, the low variability of loci from selected regions simply represents the extreme values in such a distribution. However, this hypothesis seems unlikely because of the correlation between variability levels and selected regions (above) and because the selective sweep hypothesis is supported by multiple lines of evidence (below).

BC₂ synthetic hybrid populations descended from contemporary populations of *H. annuus* and *H. petiolaris* have already helped to recreate many aspects of the genotype and phenotype of the ancient hybrid (Rieseberg *et al.* 2003; Rosenthal *et al.* 2005). In our study, the three survivorship QTLs detected in BC₂s were good predictors of regions under selection in the natural hybrid species. Interestingly, *H. petiolaris* alleles for two of the three survivorship QTLs were favoured in the BC₂s (Lexer *et al.* 2003), which is in agreement with the high proportion of *H. petiolaris* alleles from PSRs in the present study. Paradoxically, *H. petiolaris* is even more salt sensitive than *H. annuus* (Lexer *et al.* 2003). However, previous work indicates that the origin of *H. paradoxus* and other homoploid hybrid sunflower species was facilitated by transgressive segregation, whereby complementary genes from both parental species were combined to colonize extreme environmental conditions (Rieseberg *et al.* 2003). That is, despite being salt sensitive, *H. petiolaris* does contain alleles that contribute to salt tolerance and these alleles appear to have facilitated ecological adaptation of *H. paradoxus*.

We find significant differences between microsatellite classes for the level of genetic diversity within *H. paradoxus*. It is noteworthy that these differences are much better explained by regression models that account for selection coefficients and recombination distances to selected sites. This result strongly supports our interpretation that the lower diversity of PSR markers is a consequence of selection and genetic hitchhiking (Maynard Smith & Haigh 1974).

So when did these sweeps occur? One possibility is that we have detected the footprint of very strong sweeps that occurred during the origin of *H. paradoxus*. This hypothesis is supported by the observation that diversity is reduced at PSR loci for all three *H. paradoxus* populations and that very large chromosomal segments, such as those created by admixture (Ungerer *et al.* 1998), seemed to have been affected by the sweeps. This would also imply that the admixture event was recent, since the molecular signature of selection will be eroded quickly over time (Slatkin 1995; Wiehe 1998).

An alternative hypothesis is that the survivorship QTLs are rich in genes that contribute to salt adaptation and have been subjected to repeated selective sweeps since the origin of *H. paradoxus*. This explanation would not require a recent origin of *H. paradoxus*, but it would imply that selection was homogenous across the three *H. paradoxus* populations. Genetic modelling as well as fine-scale analyses of genetic diversity levels within these QTL regions are currently underway in an attempt to distinguish between these possibilities and to account for the seemingly large width of the sweeps observed.

Acknowledgements

We thank Frederic Austerlitz for helpful discussion on the data analyses, Jennifer Durphy and Mark Welch for plant material

development and Jared Strasburg for comments on the manuscript. This research was funded by grants from the National Science Foundation (DEB-0314654 and DBI0421630) and the National Institutes of Health (GM059065) to L.H.R. and by a thesis grant from the French Ministère de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche to C.E.

References

- Applied Biosystems (2000) *ABI PRISM GENESCAN Analysis Software 3.7*. Applied Biosystems, Foster City, California.
- Barton NH (1999) Clines in polygenic traits. *Genetical Research*, **74**, 223–236.
- Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology*, **13**, 969–980.
- Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **263**, 1619–1626.
- Bradshaw HD, Schemske DW (2003) Allele substitution at a flower colour locus produces a pollinator shift in monkeyflowers. *Nature*, **426**, 176–178.
- Brinkmann B, Klintschar M, Neuhuber F, Hühne J, Rolf B (1998) Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. *American Journal of Human Genetics*, **62**, 1408–1415.
- Burke JM, Tang S, Knapp SJ, Rieseberg LH (2002) Genetic analysis of sunflower domestication. *Genetics*, **161**, 1257–1267.
- Burke JM, Knapp SJ, Rieseberg LH (2005) Genetic consequences of selection during the evolution of cultivated sunflower. *Genetics*, **171**, 1933–1940.
- Clark RM, Linton E, Messing J, Doebley JF (2004) Pattern of diversity in the genomic region near the maize domestication gene *tb1*. *Proceedings of the National Academy of Sciences, USA*, **101**, 700–707.
- Clark RM, Wagler TN, Quijada P, Doebley J (2006) A distant upstream enhancer at the maize domestication gene *tb1* has pleiotropic effects on plant and inflorescent architecture. *Nature Genetics*, **38**, 594–597.
- Comps B, Gomory D, Letouzey J, Thiebaut B, Petit RJ (2001) Diverging trends between heterozygosity and allelic richness during postglacial colonization in the European beech. *Genetics*, **157**, 389–397.
- Dieringer D, Schlötterer C (2003) MICROSATELLITE ANALYSER (MSA): a platform independent analysis tool for large microsatellite data sets. *Molecular Ecology Notes*, **3**, 167–169.
- Don RH, Cox PT, Wainwright BJ, Baker K, Mattick JS (1991) 'Touchdown' PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Research*, **19**, 4008.
- El Mousadik A, Petit RJ (1996) Chloroplast DNA phylogeography of the argan tree of Morocco. *Molecular Ecology*, **5**, 547–555.
- Gardner K (2004) *Genetic Investigations of Sunflower Hybrid Zones*. PhD Thesis, Indiana University, Bloomington, Indiana.
- Goudet J (2001) *FSTAT: a program to estimate and test gene diversities and fixation indices* (version 2.9.3). University of Lausanne, Lausanne, Switzerland. Available at <http://www.unil.ch/lizea/software/fstat.html/>
- Gross BL, Kane NC, Lexer C *et al.* (2004) Reconstructing the origin of *Helianthus deserticola*: survival and selection on the desert floor. *American Naturalist*, **164**, 145–156.

- Harr B, Kauer M, Schlötterer C (2002) Hitchhiking mapping: a population-based fine-mapping strategy for adaptive mutations in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences, USA*, **99**, 12949–12954.
- Hedrick PW (2005) A standardized genetic differentiation measure. *Evolution*, **59**, 1633–1638.
- Hurlbert SH (1971) The nonconcept of species diversity: a critique and alternative parameters. *Ecology*, **52**, 577–586.
- JMP SAS Institute (1997) JMP Statistics and Graphic Guide. SAS Institute, Cary, North Carolina.
- Kauer MO, Dieringer D, Schlötterer C (2003) A microsatellite variability screen for positive selection associated with the 'out of Africa' habitat expansion of *Drosophila melanogaster*. *Genetics*, **165**, 1137–1148.
- Kayser M, Brauer S, Stoneking M (2003) A genome scan to detect candidate regions influenced by local natural selection in human populations. *Molecular Biology and Evolution*, **20**, 893–900.
- Kim Y, Stephan W (2002) Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics*, **160**, 765–777.
- Lai Z, Nakazato T, Salmaso M *et al.* (2005) Extensive chromosomal repatterning and the evolution of sterility barriers in hybrid sunflower species. *Genetics*, **171**, 291–303.
- Le Corre V, Kremer A (2003) Genetic variability at neutral markers, quantitative trait loci and trait in a subdivided population under selection. *Genetics*, **164**, 1205–1219.
- Lewontin RC, Krakauer J (1973) Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics*, **74**, 175–195.
- Lexer C, Welch ME, Durphy JL, Rieseberg LH (2003) Natural selection for salt tolerance quantitative trait loci (QTLs) in wild sunflower hybrids: implications for the origin of *Helianthus paradoxus*, a diploid hybrid species. *Molecular Ecology*, **12**, 1225–1235.
- Lexer C, Lai Z, Rieseberg LH (2004) Candidate gene polymorphisms associated with salt tolerance in wild sunflower hybrids: implications for the origin of *Helianthus paradoxus*, a diploid hybrid species. *New Phytologist*, **161**, 225–233.
- Ludwig F, Rosenthal DM, Johnston JA *et al.* (2004) Selection on leaf ecophysiological traits in a desert hybrid *Helianthus* species and early-generation hybrids. *Evolution*, **58**, 2682–2692.
- Maynard Smith J, Haigh J (1974) The hitchhiking effect of a favourable gene. *Genetical Research*, **23**, 23–35.
- McKay JK, Latta RG (2002) Adaptive population divergence: markers, QTL and traits. *Trends in Ecology & Evolution*, **17**, 285–291.
- Nair S, Williams JT, Brockman A *et al.* (2003) A selective sweep driven by pyrimethamine treatment in Southeast Asian malaria parasites. *Molecular Biology and Evolution*, **20**, 1526–1536.
- Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, **89**, 583–590.
- Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences, USA*, **76**, 5269–5273.
- Ohta T (1973) Slightly deleterious mutant substitutions in evolution. *Nature*, **246**, 96–98.
- PE Biosystems (2000) *ABI PRISM GENOTYPING Software 3.6*. PE Biosystems, Foster City, California.
- Petes TD, Greenwell PW, Dominska M (1997) Stabilization of microsatellite sequences by variant repeats in the yeast *Saccharomyces cerevisiae*. *Genetics*, **146**, 491–498.
- Petit RJ, El Mousadik A, Pons O (1998) Identifying populations for conservation on the basis of genetic markers. *Conservation Biology*, **12**, 844–855.
- QIAGEN (2001) *DNeasy 96 Plant Protocol for Isolation of DNA from Fresh Plant Leaves Using the Mixer Mill MM 300*. QIAGEN, Valencia, California.
- R Development Core Team (2004) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rieseberg LH (1991) Homoploid reticulate evolution in *Helianthus* (Asteraceae): evidence from ribosomal genes. *American Journal of Botany*, **78**, 1218–1237.
- Rieseberg LH, Carter R, Zona S (1990) Molecular tests of the hypothesized hybrid origin of two diploid *Helianthus* species (Asteraceae). *Evolution*, **44**, 1498–1511.
- Rieseberg LH, Beckstrom-Sternberg S, Liston A, Arias D (1991) Phylogenetic and systematic inferences from chloroplast DNA and isozyme variation in *Helianthus* sect. *Helianthus*. *Systematic Botany*, **16**, 50–76.
- Rieseberg LH, Van Fossen C, Desrochers AM (1995) Hybrid speciation accompanied by genomic reorganization in wild sunflowers. *Nature*, **375**, 313–316.
- Rieseberg LH, Sinervo B, Linder CR, Ungerer MC, Arias DM (1996) Role of gene interactions in hybrid speciation: evidence from ancient and experimental hybrids. *Science*, **272**, 741–745.
- Rieseberg LH, Raymond O, Rosenthal DM *et al.* (2003) Major ecological transitions in wild sunflowers facilitated by hybridization. *Science*, **301**, 1211–1216.
- Rosenthal DM, Rieseberg LH, Donovan LA (2005) Re-creating ancient hybrid species' complex phenotypes from early-generation synthetic hybrids: three examples using wild sunflowers. *American Naturalist*, **166**, 26–41.
- Schlötterer C (2002) A microsatellite-based multilocus screen for the identification of local selective sweeps. *Genetics*, **160**, 753–763.
- Schlötterer C, Dieringer D (2005) A novel test statistic for the identification of local selective sweeps based on microsatellite gene diversity. In: *Selective Sweep* (ed. Nurminsky D), pp. 55–64. Landes Bioscience, New York.
- Slatkin M (1995) Hitchhiking and associative overdominance at a microsatellite locus. *Molecular Biology and Evolution*, **12**, 473–480.
- Storz JF (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology*, **14**, 671–688.
- Storz JF, Dubach JM (2004) Natural selection drives altitudinal divergence at the albumin locus in deer mice, *Peromyscus maniculatus*. *Evolution*, **58**, 1342–1352.
- Storz JF, Nachman MW (2003) Natural selection on protein polymorphism in the rodent genus *Peromyscus*: evidence from interlocus contrasts. *Evolution*, **57**, 2628–2635.
- Storz JF, Payseur BA, Nachman MW (2004) Genome scans of DNA variability in humans reveal evidence for selective sweeps outside of Africa. *Molecular Biology and Evolution*, **21**, 1800–1811.
- Tajima F (1989) The effect of change in population size on DNA polymorphism. *Genetics*, **123**, 597–601.
- Tang S, Yu JK, Slabaugh B, Shintani K, Knapp J (2002) Simple sequence repeat map of the sunflower genome. *Theoretical and Applied Genetics*, **105**, 1124–1136.
- Ungerer MC, Baird SJ, Pan J, Rieseberg LH (1998) Rapid hybrid speciation in wild sunflowers. *Proceedings of the National Academy of Sciences, USA*, **95**, 11757–11762.

- Vasemägi A, Nilsson J, Primmer CR (2005) Expressed sequence tag-linked microsatellites as a source of gene-associated polymorphisms for detecting signatures of divergent selection in Atlantic salmon (*Salmo salar* L.). *Molecular Biology and Evolution*, 22, 1067–1076.
- Vigouroux Y, McMullen M, Hittinger CT *et al.* (2002) Identifying genes of agronomic importance in maize by screening microsatellites for evidence of selection during domestication. *Proceedings of the National Academy of Sciences, USA*, 99, 9650–9655.
- Vigouroux Y, Mitchell S, Matsuoka Y *et al.* (2005) An analysis of genetic diversity across the maize genome using microsatellites. *Genetics*, 169, 1617–1630.
- Wang J (2003) Maximum-likelihood estimation of admixture proportions from genetic data. *Genetics*, 164, 747–765.
- Weinig C, Dorn LA, Kane NC *et al.* (2003) Heterogeneous selection at specific loci in natural environments in *Arabidopsis thaliana*. *Genetics*, 165, 321–329.
- Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution*, 38, 1358–1370.
- Welch ME, Rieseberg LH (2002) Patterns of genetic variation suggest a single, ancient origin for the diploid hybrid species *Helianthus paradoxus*. *Evolution*, 56, 2126–2137.
- Wiehe T (1998) The effect of selective sweeps on the variance of the allele distribution of a linked multi-allele locus: hitchhiking of microsatellites. *Theoretical Population Biology*, 53, 272–283.
- Wiener P, Burton D, Ajmone-Marsan P *et al.* (2003) Signatures of selection? Patterns of microsatellite diversity on a chromosome containing a selected locus. *Heredity*, 90, 350–358.
- Wierdl M, Dominska M, Petes TD (1997) Microsatellite instability in yeast: dependence on the length of the microsatellite. *Genetics*, 146, 769–779.
- Wootton JC, Feng X, Ferdig MT *et al.* (2002) Genetic diversity and chloroquine selective sweeps in *Plasmodium falciparum*. *Nature*, 418, 320–323.
- Yatabe Y, Rieseberg LH (in review) Persistent gene flow and adaptive differentiation between the annual sunflowers, *Helianthus annuus* and *H. petiolaris*. *Genetics*.
-
- Loren Rieseberg and Christian Lexer study the genetics of hybridization and speciation in plants. Christine Dillmann and Delphine Sicard are working on the genetics and evolution of quantitative traits. This study is part of Cecile Edelist's PhD project carried out in collaboration between the laboratories of Rieseberg and Dillmann. Cecile studies the genetic basis of adaptation to saline stress to explore the origin of the homoploid hybrid species *Helianthus paradoxus*.
-

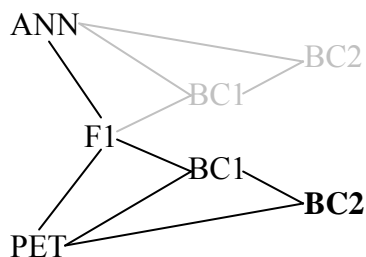
CHAPITRE 2

Cartographie génétique sur des populations naturelles et signature de la sélection à l'échelle d'un groupe de liaison

I. Intensification du marquage sur les groupes de liaisons contenant les QTL de survie

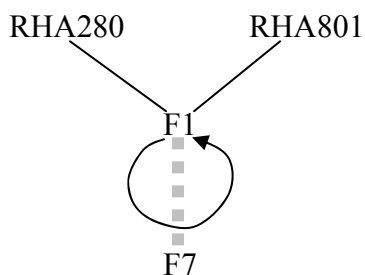
Afin de délimiter plus précisément les zones où le polymorphisme est réduit chez *H. paradoxus*, nous avons intensifié le génotypage de marqueurs microsatellites dans les trois régions contenant les QTL de survie : 11 marqueurs microsatellites sur le groupe de liaison LG1, 20 sur LG4 et 11 sur LG17 (voir cartes comparatives, Figure 2.1). Ces 42 nouveaux microsatellites ont été génotypés chez les mêmes individus des trois populations de *H. paradoxus*. 22 de ces nouveaux microsatellites, dont le nom commence par HT, sont contenus dans des EST (Expressed Sequence Tag). Les marqueurs génotypés proviennent de trois cartes génétiques différentes construites i) sur une population de backcross issue du croisement entre les deux espèces parentales *H. annuus* et *H. petiolaris*, ii) sur une population de lignées recombinantes (RIL) issue d'un croisement entre deux génotypes cultivés de *H. annuus*, iii) sur une population issue du croisement entre un hybride de deux individus de populations naturelles de *H. paradoxus* et une lignée cultivée de *H. annuus*. Les caractéristiques de ces populations sont présentées ci-dessous :

i) Population BC2 issue du croisement entre les deux espèces parentales *H. annuus* et *H. petiolaris* :



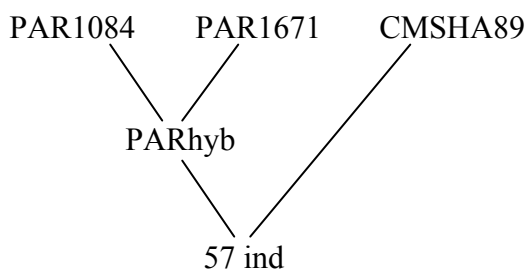
C'est sur cette population de 384 individus BC2, issue d'un double backcross avec *H. petiolaris*, que sont cartographiés les QTL de survie et ions minéraux. Elle comprend 76 microsatellites, dont ceux proches des QTL de survie, et 20 AFLP et sa taille est de 1252 cM.

ii) Population de lignées recombinantes issue d'un croisement entre deux génotypes cultivés de *H. annuus* :



Cette carte a été construite sur 94 individus RIL de septièmes générations issus du croisement entre deux lignées de tournesols *H. annuus* RHA280 et RHA801. 408 microsatellites sont cartographiés et la carte fait 1368.3 cM (Tang *et al.* 2002). Le développement de microsatellites dans des EST a permis d'intensifier cette carte (1887.3 cM) (Knapp, non publié).

iii) Population issue du croisement entre un hybride de deux individus de populations naturelles de *H. paradoxus* et une lignée cultivée de *H. annuus* :



PAR1084 et PAR1671 sont deux populations de *H. paradoxus* situées au Texas et au Nouveau Mexique respectivement. Un individu hybride issu d'un croisement entre un individu PAR1084 et PAR1671 a été croisé avec une mère de la lignée CMSHA89, une lignée de *H. annuus* isogénique. Cette carte a été construite sur 57 individus à l'aide de 120 microsatellites et 552 AFLP et s'étend sur 1420.5 cM (Lai *et al.* 2005).

Ces cartes présentent peu de marqueurs en commun (Figure 2.1). Bien qu'il semble exister une syntenie importante entre les différentes espèces de tournesols étudiées (Lai *et al.* 2005), la provenance des marqueurs de différentes cartes ne permet pas de connaître avec précision leur ordre et leur distance les uns par rapport aux autres chez *H. paradoxus*. Nous avons développé deux stratégies pour ordonner les marqueurs : l'une de cartographie classique et la seconde d'ordonnement des marqueurs par déséquilibre de liaison.

Intensification de la carte de *H. paradoxus* :

Nous avons essayé de génotyper les microsatellites choisis à partir de la carte de *H. annuus* (Figure 2.1), sur l'ensemble des 57 individus de la population de cartographie de *H. paradoxus*. Cependant la qualité des ADN de ces individus ne nous a pas permis de les génotyper de manière satisfaisante. Il est important de noter que même avec des données de qualité, il aurait été très difficile d'ordonner beaucoup de marqueurs dans une région

relativement petite, la taille de la population n'étant pas très importante et la distance entre les nouveaux marqueurs sur leur carte génétique respective faible.

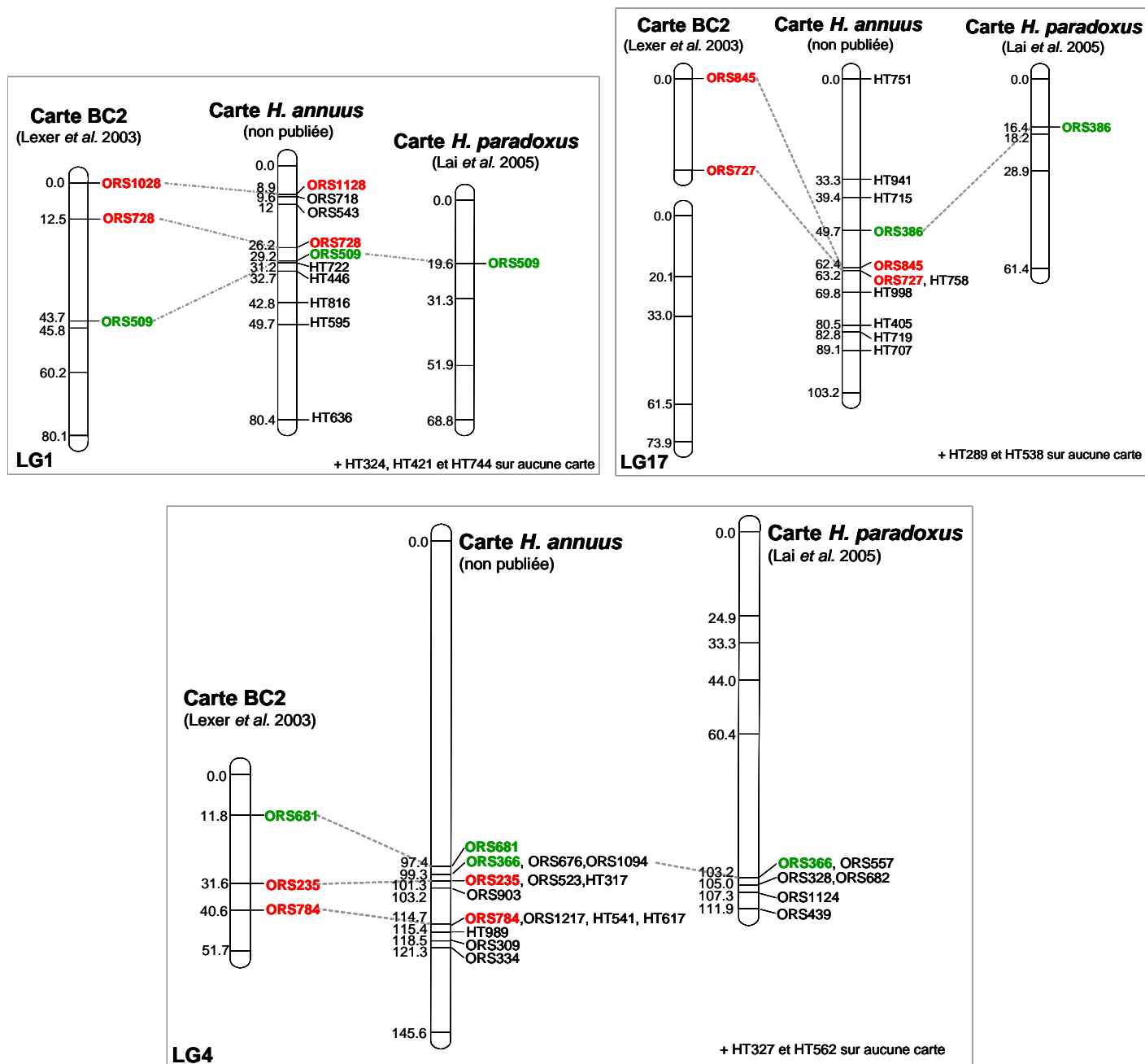


Figure 2.1 : Cartes comparées de trois groupes de liaison LG1, LG4 et LG17, d'une population BC2 (Lexter et al. 2003 – 1252cM), de *H. annuus* (Knapp non publiée – 1887.3cM, mais qui est une version complétée de Tang et al. 2002) et de *H. paradoxus* (Lai et al. 2005 – 1420.5cM), sur lesquelles sont localisées l'ensemble des microsatellites génotypés. Les microsatellites de couleur signifient : en rouge que les microsatellites sont proches de QTL de survie, en vert que les microsatellites sont localisés sur plusieurs cartes. Les microsatellites dont le nom débute par HT sont dans des EST (Expressed Sequence Tag). Les microsatellites 'sur aucune carte' appartiennent au groupe de liaison concerné (Knapp non publié), mais n'ont pas été localisé plus précisément.

Utilisation du déséquilibre de liaison entre marqueurs microsatellites

Nous avons développé une méthode basée sur le déséquilibre de liaison pour ordonner les marqueurs situés sur un même groupe de liaison. Cette méthode a été testée sur des populations de RIL de maïs dont la carte génétique est connue, sur des populations naturelles simulées et enfin sur des populations naturelles de *H. annuus*, sachant que nous disposons d'une carte génétique de cultivar de cette espèce. La méthode et les résultats sont décrits dans les pages qui suivent.

II. Using linkage disequilibrium patterns for ordering genetic markers in natural populations

(To be submitted, Journal of Heridity)

Cécile Edelist¹, Delphine Sicard¹, Adrienne Ressayre¹, Yoko Yatabe², Loren H. Rieseberg^{3,4} and Christine Dillmann¹

1 UMR de Génétique Végétale, INRA / Univ. Paris XI / CNRS / INA PG, Ferme du Moulon, 91190 Gif-sur-Yvette, France

2 Tsukuba Botanical Garden, 4-4-1 Amakubo, Tsukuba, Ibaragi, 305-0005 Japan

3 Department of Botany, University of British Columbia, Vancouver, B.C. V6T 1Z4, Canada

4 Department of Biology, Indiana University, Bloomington, IN 47405, USA

Address correspondence to Cécile Edelist, UMR de Génétique Végétale, INRA / Univ. Paris XI / CNRS / INA-PG, Ferme du Moulon, 91190 Gif-sur-Yvette, France; Fax: 33 1 69 33 23 40, or e-mail: edelist@moulon.inra.fr

Acknowledgements

We thank Frederic Austerlitz, Matthieu Falque and Olivier Martin for helpful discussion and comments on the manuscript. This research was funded by grants from the French Ministère de l'Éducation Nationale et de la Recherche to CE.

Keywords

LD mapping, ordering markers, natural population

Running title: Linkage disequilibrium mapping in natural populations

Abstract

Information on genetic polymorphism in natural populations of non-model organisms is now widespread thanks to the increasing ease with which molecular markers can be developed. However, it is less common to have a genetic map for non-model species because many species are difficult to culture in the lab or greenhouse or have long generation times. Thus, the genomic location of genotyped markers is rarely known in these systems. Here we develop and test an algorithm for ordering genetic markers based on the associations or linkage disequilibrium between them. The best marker order was assumed to be the one for which linkage disequilibrium decreased most consistently with genetic distance. The best order was found by replacing unknown genetic distances with the number of intervals between the markers and running an iterative algorithm. The method was tested across three maize RIL populations and simulated populations before being applied to natural populations of the common sunflower, *Helianthus annuus*. High Spearman rank correlations were found between the best order and the correct map order in the maize RILs and in simulated populations, thereby validating the ordering algorithm. However, no significant congruence was found between the best order generated by our algorithm in natural *H. annuus* populations and the genetic map for domesticated sunflower. These results imply that factors other than genetic distance have a strong influence on patterns of LD in natural populations.

Introduction

The genome can be viewed as a mosaic of loci that have been subjected to different evolutionary pressures, with each genome segment reflecting a unique, but partially overlapping part of the evolutionary history of the population (Storz 2005). Evolutionary pressures like migration, mutation, selection, and genetic drift not only determine the polymorphism of a single locus, but they also create linkage disequilibrium (LD) between loci, i.e., deviations from random associations of alleles at different loci (Bennett 1954). Hence, changes in the frequencies of gametes within a population are likely to affect both the frequencies of alleles and the strength of associations or LD between them. Once created, LD can be broken down by recombination. Therefore, LD decreases with genetic distance and the pattern of LD along chromosomes can be employed to map the location of genes or mutations underlying quantitative trait variation (Lewontin 1964; Lewontin 1974; Hudson 2001; Stephan *et al.* 2006).

LD mapping has first been particularly effective in human populations where the decay of LD with physical distance is fast and barely detectable beyond 5kb (Dunning *et al.* 2000). This approach was employed, for example, to identify the causal mutation in *CHEK2*, which is associated with increased risk for breast cancer (Einarsdottir *et al.* 2006). In populations that exhibit LD over greater genetic distances due to selfing, population bottlenecks, or admixture, mapping precision is reduced and false positive rates are higher (Thornsberry *et al.* 2001; Rieseberg & Buerkle 2002; Olsen *et al.* 2004). Nonetheless, with appropriate sampling and data analyses strategies, it may be possible to map quantitative trait loci (QTLs) or quantitative trait nucleotides (QTNs) in these more structured populations. Well known examples include the identification of QTLs for pollen viability in wild sunflower hybrid zones (Rieseberg *et al.* 1999) and QTNs underlying pathogen resistance and flowering time in the selfing plant, *Arabidopsis thaliana* (Aranzana *et al.* 2005). Numerous other examples come from studies of domesticated plants and animals (Sandor *et al.* 2006; Breseghello & Sorrells 2006).

While extensive LD is an impediment to precision mapping of QTNs (Huttley *et al.* 1999; Farnir *et al.* 2000; Tenesa *et al.* 2004), it may be useful for estimating locus orders at broader genetic distances. Genome mapping is usually accomplished using artificial F₂ or backcross populations, but this may not be practical in many wild organisms that are difficult to propagate in the lab or greenhouse or that have long generation spans. The possibility of genetic mapping in natural populations has been explored in hybrid zones (e.g., Rieseberg & Buerkle 2002), but not, as far as we are aware, for populations within species.

In the present paper, we ask whether the strength of LD between molecular markers could be used to order them in natural populations of organisms that lack a genetic map. As mentioned above, LD is expected to decline with genetic distance, so there is theoretical justification for such an approach. However, because LD is affected by both genome-wide (drift) and locus-specific (e.g., mutation and selection) evolutionary forces, it might be that the variance in LD will be too high to make such an approach practical. What is needed is an empirical test of “long distance” LD mapping.

Here, we provide such a test. Our method is based on the estimated pairwise LD between all pairs of markers that are to be ordered. We assume that the best order is the one for which the LD pattern of decrease with genetic distance is the most consistent. This best order is found by replacing the unknown genetic distances by the number of intervals between the markers and running an iterative algorithm. The efficiency of the method is tested over three mapping maize RIL populations and simulated populations before being applied to natural populations of the common sunflower, *Helianthus annuus*.

Material and Methods

Long distance LD mapping: LD estimator and ordering criterion

A set of n markers belonging to the same linkage group was considered. The $\frac{n(n-1)}{2}$ pairwise LD estimation R_{ij}^2 , was estimated following (Maruyama 1982; Hedrick 1987):

$$R_{ij}^2 = \frac{1}{(l_i - 1)(l_j - 1)} \sum_k \sum_l \frac{(f_{ikjl} - f_{ik}f_{jl})^2}{f_{ik}f_{jl}} \quad (5)$$

where f_{ikjl} is the frequency of the gamete carrying allele k at marker i and allele l at marker j , f_{ik} and f_{jl} are the allele frequencies at markers i and j respectively, and l_i and l_j are the number of alleles at markers i and j respectively. All LD estimators R_{ij}^2 , were associated with a P value P_{ij} . Because all markers from a linkage group may not be in significant LD, LD Groups (LDG) were formed on the basis of the R_{ij}^2 and associated P_{ij} values between pairs of markers. Each LDG was comprised of all markers that were in significant LD with at least one other marker of the group. This procedure is very similar to the one employed by Mapmaker (Lander *et al.* 1987) to form linkage groups. The threshold P value can be adjusted depending on the level of LD in the population under consideration.

The true order of markers was set at $\{1\ 2\ \dots\ n\}$. If markers are fairly evenly spaced along the chromosome, then the genetic distance between any pair of markers i and j will be proportional to the number of intervals separating them: $\Delta_{ij} = (j - i)$. Given that LD between marker pairs decreases with genetic distance, then LD between markers will decrease as Δ_{ij} increases. The ‘‘LD distance’’ between two markers can be defined as:

$$D_{ij} = -\log R_{ij}^2 \quad (6)$$

R_{ij}^2 decreases exponentially with D_{ij} . Here, we supposed that D_{ij} is a linear function of the number of marker intervals Δ_{ij} :

$$D_{ij} = -\log R_{ij}^2 = C + \lambda\Delta_{ij} + \varepsilon_{ij} \quad (7)$$

where C is a constant, λ the slope of the curve and ε_{ij} the random noise of variance σ^2 . Note that the higher the value of λ , the faster the increase of LD distance relative to marker intervals, and the steeper the decrease of linkage disequilibrium with genetic distance. Hence, the ordering criterion was to maximize λ and the best order was chosen as the one for which λ is maximal. For each order tested, we performed a linear regression of D_{ij} on Δ_{ij} to estimate

C and λ . We also computed the *SSE* of the regression as $SSE = \sum_i \sum_{j>i} (D_{ij} - \hat{C} - \hat{\lambda}\Delta_{ij})^2$.

Algorithm for ordering markers

We developed a steepest descent algorithm to find the order that maximized the ordering criterion, called best order $B = \{b_1 b_2 \dots b_n\}$, with exploring only a subset of the $n!$ possible orders. Starting from markers in arbitrary order $\{j_1 j_2 \dots j_n\}$, we explored n different initial trial sets, by permuting j_1 with each of the $\{j_1 j_2 \dots j_n\}$, for which we computed the ordering criterion. In each trial set, the second marker was determined by choosing among the $n-1$ markers the one that has the highest R^2 with the first marker. The third marker was chosen as the one that has the highest R^2 with the second, etc... until all the markers were ordered. Then all the permutations between two markers of the trial set were tested, and the permutation having the best value of the search criterion was defined as the new trial set. At the end, there were n values corresponding to the n best orders starting with each of the markers available. The order with the highest value for the search criterion was chosen as the best order. All the analyses were performed using R software (R Development Core Team 2004).

Testing the ordering method

The algorithm for marker ordering was tested by simulating 10 000 distance matrices between $n = 10$ markers from equation (7). For each replicate, a random value was chosen for C , λ and σ , and the $\frac{n(n-1)}{2}$ values of ε_{ij} were drawn in a normal distribution $N(0, \sigma)$. The intervals of variation were [0-3] for C , [0-1] for λ , and [0-10] for σ . The LD distances were computed according to (7) for the map order $\{1\ 2\ \dots\ n\}$. Negative D_{ij} values were set to 0, corresponding to R_{ij}^2 values of 1. After applying the ordering algorithm to a simulated distance matrix, we computed the rank correlation r_S between the map order and the best order found. We also estimated \hat{C}_{map} , $\hat{\lambda}_{map}$ and SSE_{map} for the map order by linear regression. We considered that the algorithm was successful when the rank correlation r_S between the map order and the best order found was higher than 0.80. We counted the percentage of success for different values of $\hat{\lambda}_{map}$ and SSE_{map} . This procedure provides an estimate of the power of our ordering method.

Population and genotyping data

Maize RIL populations

For maize, we used RFLP data from three populations (A, B, and C) of recombinant inbred lines, all from generation six (Causse *et al.* 1996). The F1 individuals resulted from self-pollination of A) a cross between a male French early flint line ($F2$) and a female from the Iodent group (Io); B) a male $F2$ and a female American early dent line ($F252$), and C) a male $F252$ and a female Io . A total of 200 F2 plants from each population were self-pollinated, and 7 to 10 F3 seeds per F2 plant were grown. The next generation was produced by randomly choosing one F3 plant for each F2 plant. The process was repeated three more times to obtain the three F6 RIL populations. Final population sizes at generation six were 145 individuals for population A, 129 for population B and 152 for population C (Causse *et al.* 1996). A total of 126, 112 and 116 RFLP markers for the populations A, B and C respectively, were mapped on the 10 maize chromosomes (see Causse *et al.* (1996) for more details) with the software Mapmaker 3.0 (Lander *et al.* 1987). The number of markers per chromosome ranged between nine and 22, eight and 17, and eight and 18 for populations A, B and C respectively.

Simulated populations

We simulated diploid populations of an effective size Ne evolving for T generations. Each individual was represented by a pair of chromosomes of 100 centiMorgans (cM). 25 markers were randomly distributed along the 100 cM. At each marker, the initial number of repetitions for each of the $2Ne$ microsatellite alleles was drawn in a Gamma distribution. The parameters of the Gamma distribution were chosen to fit empirical observations in natural populations of the genus *Helianthus* (personal observation): the average length of the repetition motif was chosen equal to 20 nucleotides and the variance of the repetition motif was fixed to 1.3, so that the shape and scale parameters of the distribution were $\Gamma\left(\frac{20^2}{1.3}, \frac{1.3}{20}\right)$.

The mutation model was the same as described in Kruglyak *et al.* (1998). At each generation, an allele can mutate by strand slippage mimicking polymerase error, or point mutation events, or transition due to specific base substitutions. An allele of size x_i can gain or lose one repeat by slippage at a rate $\mu_a(x_i-1)$. A point mutation can decrease the number of repeats to any value between 1 and (x_i-1) at a rate μ_b . Finally, base substitutions occurring at a small rate μ_c can cause a transition from $x_i = 1$ to $x_i = 2$, and therefore avoid having an allele absorbed by stage one. The values of the initial mutation rates per repeat unit per generation were $\mu_a = 10^{-6}$ for slippage, $\mu_b = 10^{-8}$ for point mutations and $\mu_c = 10^{-8}$ for substitutions (Kruglyak *et al.* 1998). This model led to an effective mutation rate per marker and per generation of approximately $3.8 \cdot 10^{-4}$ at the start of the simulation. This mutation rate changes through time with the allele frequency distribution. At the end of the simulation the mutation rate ranged from $3.2 \cdot 10^{-4}$ and $4.7 \cdot 10^{-4}$ and varied among markers. At each generation, crossovers between microsatellite markers were generated according to a Poisson distribution assuming no interference (Hospital & Chevalet 1993). The parameter of the Poisson distribution was the chromosome length expressed in Morgans. For the simulated populations, the marker order, as well as the genetic map, were known precisely. All the simulations were performed using R software (R Development Core Team 2004).

Helianthus annuus wild populations

H. annuus is the most common wild sunflower species across North America and is the progenitor of the domesticated sunflower. In this study we used microsatellite data from three wild populations of *H. annuus*. Achenes were collected on 60 plants over three relatively large wild populations (> 1000 plants): one in North Dakota (NDC), a second one in South Dakota (SDD) and a third one in Nebraska (NEG) (Yatabe & Rieseberg In review).

Achenes were germinated at Indiana University and after DNA extraction 108 microsatellites were genotyped on 39 NDC, 52 SDD and 54 NEG individuals. The microsatellites were chosen from a genetic map developed from a cross between two domesticated inbred lines (Tang *et al.* 2002). The map spans the entire genome of sunflower, which is composed of 17 chromosome pairs (see Yatabe & Rieseberg In review, for more details). For this study, we focused on the three linkage groups (LG) for which at least nine microsatellite markers had been genotyped: LG10 (nine markers), LG13 (13 markers) and LG17 (16 markers).

Computation of pairwise R_{ij}^2 and associated P values

For the sunflower experimental data, the gametic phase was unknown and R_{ij}^2 was estimated as the Burrow's composite measure (Weir 1979). The estimated linkage disequilibrium was calculated with Linkades (Garnier-Gere & Dillmann 1992). For all experimental data, LD estimators were associated with a P value P_{ij} obtained by an exact test with Genepop (Raymond & Rousset 1995). For the simulated data, the less conservative P_{ij} values were computed by assuming a Chi2 distribution $\chi_{(l_i-1)(l_j-1)}^2$ for the nR_{ij}^2 statistics, where n is the sample size.

Efficiency of the algorithm on simulated and real populations

The efficiency of the method was assessed by comparing the best marker order found by the algorithm to the known map order. We computed the Spearman rank correlation r_S between the map order and the best order found. A standardized LD distance between adjacent ordered markers was also computed as

$$D'_{i,i+1} = \frac{-\log R_{i,i+1}^2}{\sum_{j=1}^{n-1} (-\log R_{j,j+1}^2)} \quad \text{and} \quad D'_{i,j} = \sum_{k=i}^{j-1} D'_{k,k+1}$$

for $j > i+1$. This allowed us to calculate Pearson's correlation coefficient r_P between pairwise genetic distances $d_{i,j}$ expressed in cM, and standardized LD distances D'_{ij} for the best order.

Results

Power of the marker ordering method

The efficiency of the algorithm for marker ordering was assessed by Monte-Carlo simulations of LD distance matrices according to equation (7). The search criterion was to maximize the slope λ of the regression between pairwise LD distances D_{ij} and the number of intervals between the markers Δ_{ij} . For each replication, we first compared the slope $\hat{\lambda}_{map}$ that was estimated given the map order, to the slope $\hat{\lambda}_{best}$ that was estimated given the best order found. The slope $\hat{\lambda}_{best}$ was always higher or equal to $\hat{\lambda}_{map}$ (results not shown) indicating that our algorithm is efficient in finding the order of the markers which maximizes λ . The power to detect the correct marker order was assessed by counting the number of time the algorithm ends up with a best order having a rank correlation r_s higher than 0.8 with the map order. The power of our method depends on both $\hat{\lambda}_{map}$ and SSE_{map} (Figure 1). Unsurprisingly, it increases with $\hat{\lambda}_{map}$, and is higher when SSE_{map} is low. When SSE_{map} was lower than 1 and $\hat{\lambda}_{map} > 0.25$, a high rank correlation was found with map order in all cases. When SSE_{map} was between 1 and 5, however, this was achieved only for $\hat{\lambda}_{map} > 0.60$ (Figure 1).

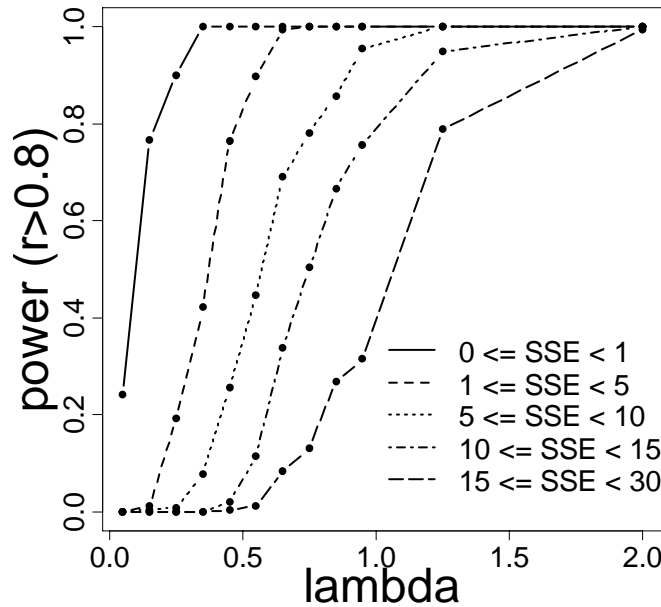


Figure 1: Power for finding correct marker orders when using the following linear relationship: $D_{ij} = -\log R_{ij}^2 = C + \lambda \Delta_{ij} + \varepsilon_{ij}$. A total of 10 000 LD matrices were simulated, for which C , λ and SSE were drawn from different laws. The correct order was supposed to be found whenever the rank correlation r_s between the actual map order and the best order was greater than 0.8. The values of λ and SSE were obtained using the map order.

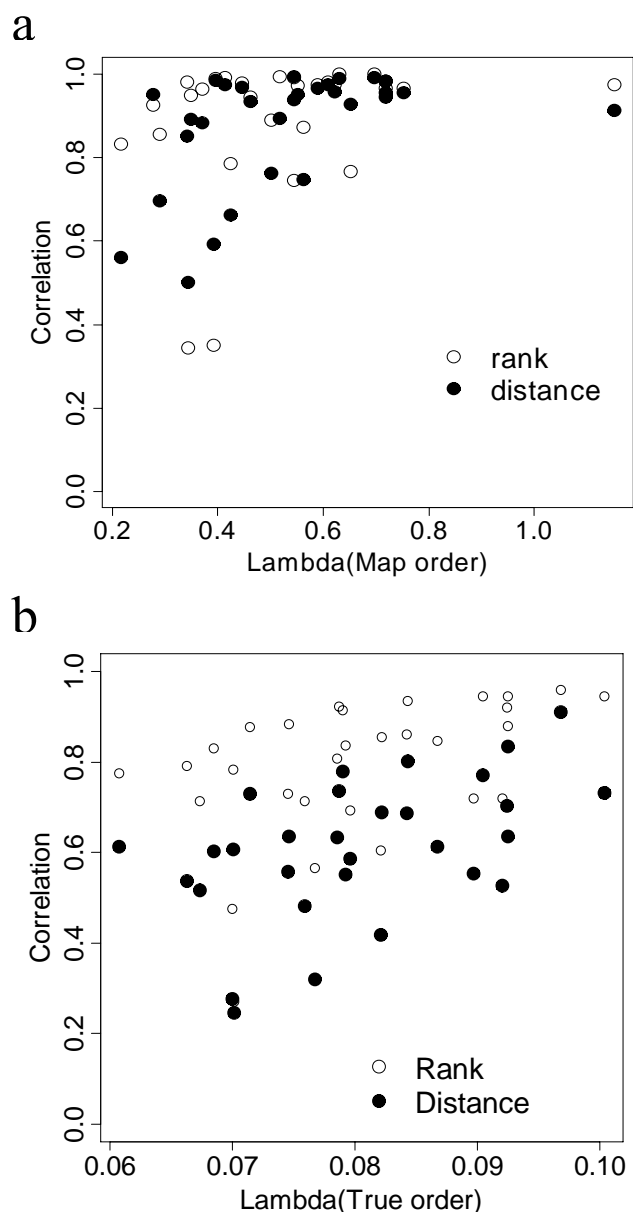


Figure 2: **Efficiency of the ordering algorithm on maize RIL populations (a) or on simulated populations (b).** Rank correlations between the map order and the best order (in white) and Pearson's correlation coefficient between pairwise genetic distances and pairwise LD distances of the best order (in black) are represented as a function of λ estimated with the map order. Pairwise LD distances are computed for the best order found with the algorithm. Each dot of the same color represents a linkage group for one of the three RIL populations (a) or a population of size $N = 1\,000$ that has evolved for 5 000 generations (b).

Maize RIL populations

In the maize RIL populations, the increase of LD distance with the number of intervals between the markers was such that $\hat{\lambda}_{map}$ ranged between 0.22 and 1.15 with an average value of 0.526, and SSE_{map} ranged between 0.47 and 7.20, with an average value of 2.42. On the basis of the $\hat{\lambda}_{map}$ and SSE_{map} averages, we expect the best order found to be close to the map order in about 85% of the cases (Figure 1). Indeed, the rank correlation r_s between the map order and the best order was higher than 0.8 in 25 out of the 30 chromosomes (3 populations x 10 chromosomes per population), and higher than 0.9 in 21 out of the 30 chromosomes tested. The rank correlation tended to increase as $\hat{\lambda}_{map}$ increased (Figure 2a, white circles). It was significant in all cases except for chromosomes B5 and C5, where $r_s = 0.34$ and 0.35 respectively. Pearson correlation coefficients r_p between pairwise map genetic distances and pairwise LD distances for the best order ranged between 0.50 and 0.99 (Figure 2a, black circles). They were all significant, and were higher than 0.8 in 23 out of the 30 chromosomes analyzed. Note that results obtained by ranked or Pearson correlations were consistent except when local inversions occur between the map order and best order.

As expected, an exponential increase of pairwise LD with distance between markers (4) was observed (not shown). To present more details on the output of our ordering method, we show the relationships between map order and best order for three contrasting chromosomes (Figure 3). Chromosome A2 ($r_s = 0.99$, $r_p = 0.99$) is representative of cases where the algorithm was effective in finding the correct map order, with only two markers ordered incorrectly, but even these were placed very closely to their correct map positions (Figure 3a). This generates a linear relationship between pairwise genetic map distances and pairwise standardized LD distances whether computed with the map order (Figure 3b, black dots) or with the best order (Figure 3b, red dots). Chromosome B5 ($r_s = 0.34$, $r_p = 0.50$) represents one of the poorest rank correlations with map order. The set of markers on chromosome B5 can be split into 2 different LDGs. Within each LDG, the ordering of the markers found by our algorithm was the same as the map order, except for one pairwise inversion (Figure 3c). Figure 3d shows that the correlation between the genetic map distance and the standardized LD distance under the map order is high (black dots), whereas the correlation between the genetic map distance and the standardized LD distance under the best order (red dots) is represented by two different clouds of points at right angles due to an inversion between the map order and the best order found. Nevertheless, as marker order is

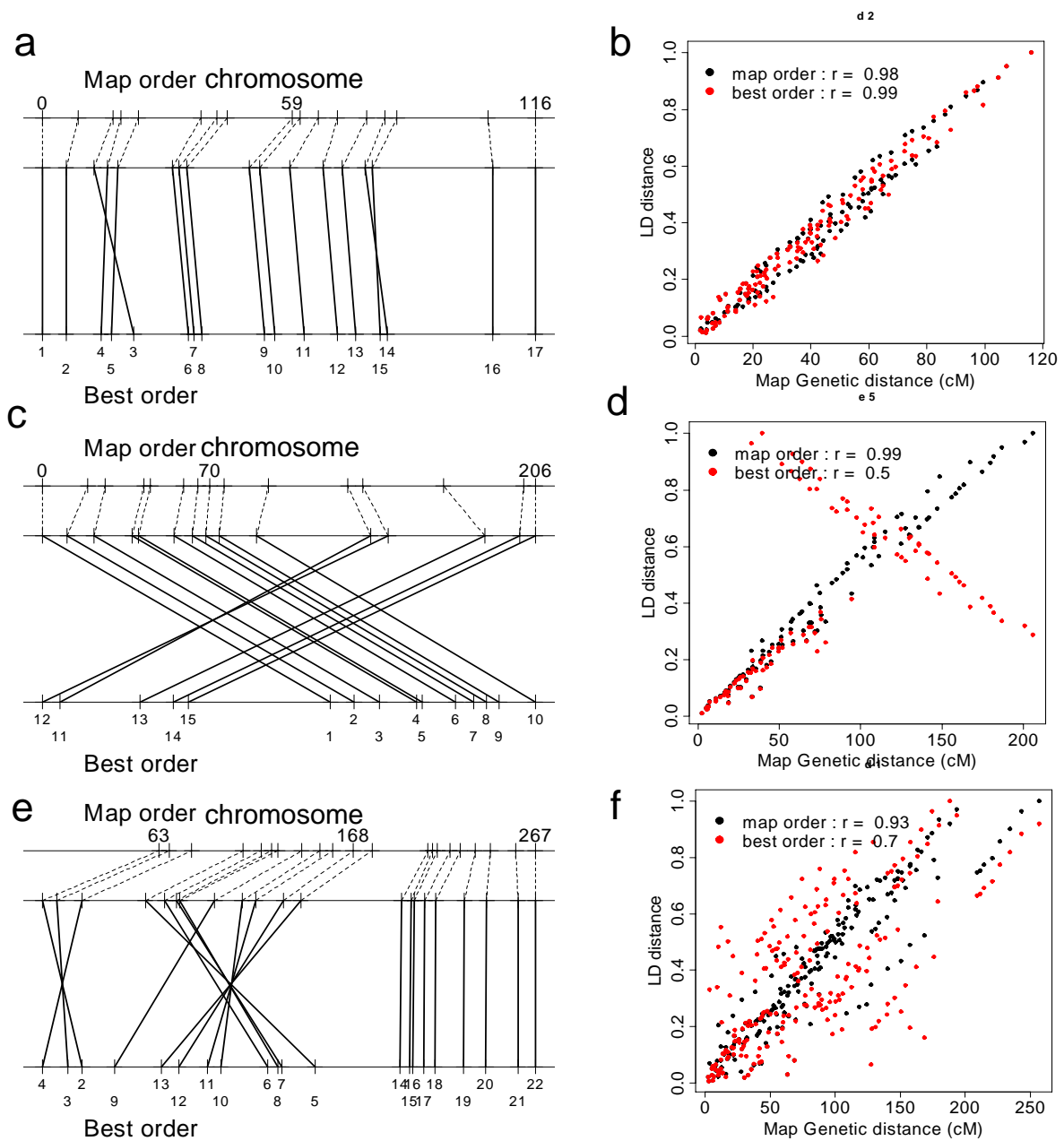


Figure 3: Comparison between the map order and the best order in the maize RIL population. (a, c, e) are the map projections. The first row is the genetic map in cM with the map order, the second row is the LD map with the map order and the third row is the LD map with the best order. (b, d, f) represent the relationship between pairwise genetic map distances and pairwise LD distances for the map order (black) and the best order (red). The three chromosomes (A2 in a, b; B5 in c, d; A1 in e, f) have been chosen as representatives of all possible situations that have been encountered.

well conserved within the inverted LD groups, the red dots are well aligned in the two directions (Figure 3d). Finally, chromosome A1 ($r_s = 0.86$, $r_p = 0.70$) was representative of intermediate cases. The genetic map comprises three clusters of markers separated by marker-free intervals of about 20 cM. The ordering algorithm was efficient in allotting markers to the correct marker subset. Within a marker subset, the best order was consistent with the map order, but the orientation of the markers subsets were inverted (Figure 3e). The resulting correlation between the map genetic distance and the LD distance based on the best order is a mixture of different clouds of points, showing a positive correlation within a marker subset, and a negative correlation between marker subsets (Figure 3f).

Simulated populations

We simulated the neutral changes in patterns of polymorphisms for 25 microsatellites markers in populations of size $N = 1\,000$ individuals during $T = 5\,000$ generations. After 5 000 generations, all the 30 simulated populations had reached mutation/drift equilibrium for single locus genetic diversity, with Nei's diversity index (Nei 1978) approximately $He = 0.50$. At the end of the simulations, only few loci (0 to 3) were fixed for one allele, and the allele distributions differed among the 25 loci: some were fixed, some had a unimodal distribution spanning a wide range of motif repetitions, when some others had split into a bimodal distribution. This is consistent with the high level of allelic variation found for microsatellites in natural populations (Ellegren 2004).

Significant linkage disequilibria were found between markers in all replicate populations. An increase in pairwise LD distances with both genetic map distance and the number of intervals between the markers was observed for all simulated populations (Figure 4). High LD values can be observed at genetic distances up to 40 cM (Figure 4a). Note that the same pattern was observed for all simulated populations, which was one of the assumptions of the model. The increase of LD distance with the number of intervals between the markers was such that $\hat{\lambda}_{map}$ ranged between 0.06 and 0.10 with an average value of 0.08, and SSE_{map} ranged between 0.67 and 2.01, with an average value of 0.91. Based on previous power estimates, we expected the best order found by our ordering method to be close to the map order in about 60% of the cases (Figure 1). The rank correlations r_s between the map order and the best order were all significant except in one replicate. They were indeed higher than 0.8 in 19 out of the 30 replicate populations, and higher than 0.9 in 8 out of the 30 replicate populations. Similar results were found with Pearson's correlation coefficients,

which were all significant. This shows that on average two markers that are close together on the genetic

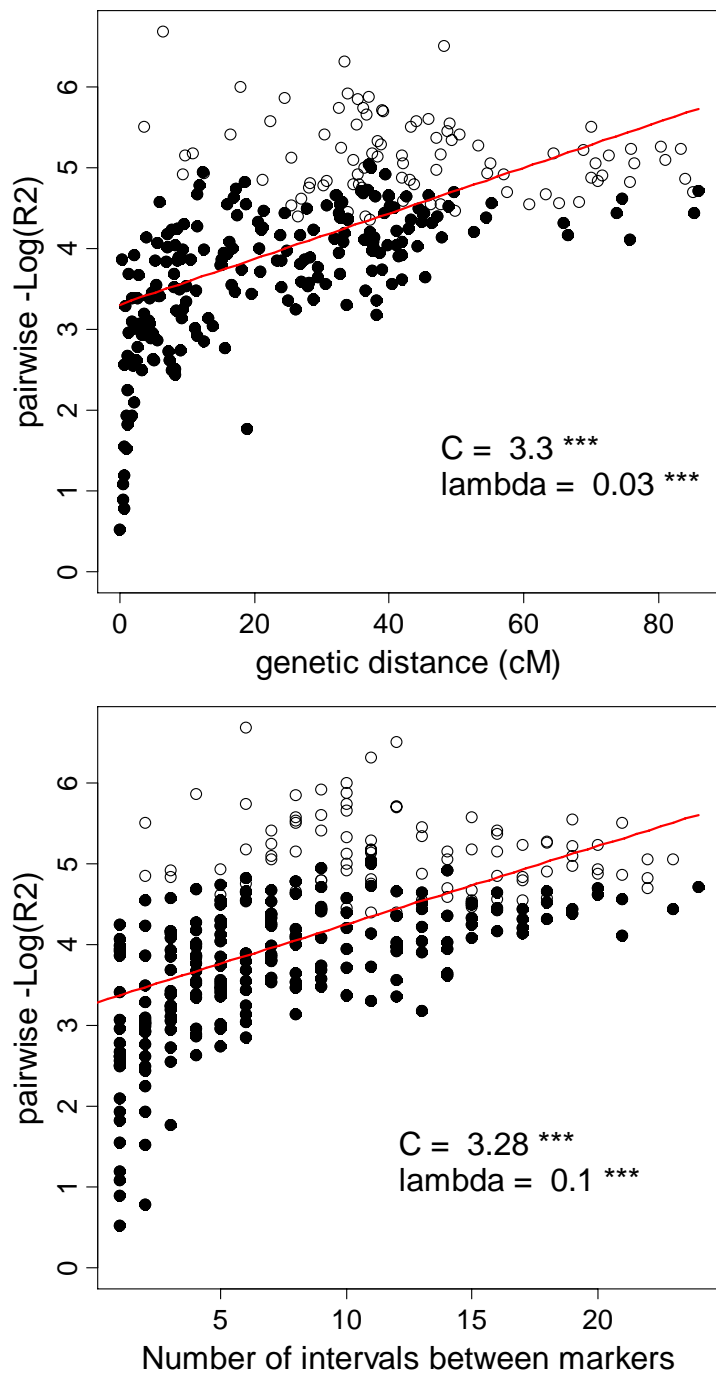


Figure 4: **Increase of pairwise LD with distances between markers in simulated populations.** The LD distance is represented as a function of either the genetic distance between markers (a) or the number of intervals between markers (b). Genetic data from 25 microsatellite markers were obtained from the simulation of a population of size $N = 1000$ during 5000 generations. Black dots are the pairs of markers with a significant LD value at the 5% level, while white dots are the pairs of markers for which the LD value is not significant.

map are also found close together in the best order found by our ordering method (Figure 2b). Moreover, r_P was increasing with $\hat{\lambda}_{map}$, and was much higher when the increase of LD distance with the number of intervals between the markers was close to $\hat{\lambda}_{map} = 0.10$.

In order to check that the simulated populations were at mutation/drift equilibrium, two simulations were run up to $T = 100\,000$ generations (results not shown). In those simulations, the coefficient of the regression of LD distance on either genetic map distance or the number of intervals between markers were the same order of magnitude as the results presented here.

Comparing the genetic map with the best order LD map among populations showed that i) in most cases, markers which belong to a group of linked markers are found together in the best order LD map; ii) as in the maize RIL population, inversions can be found between groups of linked markers; and iii) we generally found one to three markers which were assigned a wrong position in the LD best order, and which may correspond to spurious associations between markers due to genetic drift.

Helianthus annuus natural populations

The three populations of *H. annuus* exhibited high levels of genetic diversity for all 39 microsatellite markers located on linkage groups LG10, LG13 and LG17 (Table 1). Within each population, only a few markers exhibited significant F_{is} values. Global pairwise F_{st} values were low ($F_{st}(NDC - NEG) = 0.002$, $F_{st}(NDC - SDD) = 0.026$, $F_{st}(NEG - SDD) = 0.021$), with the SDD population being slightly divergent from the two other populations. The higher F_{st} between SDD and the two other populations can be

LG	10			13			17		
	NDC	NEG	SDD	NDC	NEG	SDD	NDC	NEG	SDD
Population									
Number of markers	9			13			16		
<i>He</i>	0.69	0.71	0.72	0.71	0.69	0.74	0.72	0.7	0.77
Number of pairs of markers	36			78			120		
Mean R^2 over all pairs of markers	0.02	0.047	0.046	0.018	0.056	0.051	0.019	0.053	0.041
% of significant R^2	0.08	0.11	0.28	0.05	0.06	0.29	0.025	0.12	0.1
Mea of significant R^2	0.013	0.066	0.039	0.018	0.051	0.049	0.024	0.051	0.045
C (standard error)	4.303 ** (0.067)	3.24 ** (0.079)	3.349 ** (0.054)	4.093 ** (0.043)	2.981 ** (0.049)	3.003 ** (0.052)	4.057 ** (0.039)	3.212 ** (0.042)	3.419 ** (0.041)
λ (standard error)	-0.082 ** (0.017)	-0.025 (0.020)	-0.062 ** (0.014)	-0.007 (0.008)	0 (0.009)	0.018 * (0.009)	-0.001 (0.006)	-0.025 ** (0.006)	-0.023 ** (0.006)
SSE	0.161	0.222	0.106	0.142	0.184	0.210	0.180	0.213	0.199

Table 1: Details for the markers located on LG 10, LG 13 and LG 17 in three wild populations of *H. annuus*: NDC, NEG and SDD. For each population and each linkage group, Gene Diversity *He*, LD in terms of R^2 and C, λ and SSE estimated from the *H. annuus* map order are summarized. ** $P < 0.01$, * $P < 0.05$).

explained by greater genetic diversity within the SDD population. Nei's diversity indexes averaged over all the markers were $H_e(SDD) = 0.75$, $H_e(NDC) = 0.71$, and $H_e(NEG) = 0.70$.

Within each linkage group, significant LD values were found between pairs of markers in each of the three populations (Table 1). The SDD population exhibited the highest percentage of significant LD pairs, whereas the NEG population had the lowest. In the NDC population, the percentage of significant pairs did not exceed the 5% expected by chance for linkage groups LG13 and LG17. The averages within linkage group R2 values were similar among linkage groups for a population: $R^2 = 0.050$ for all linkage groups in SDD and NEG, and $\sim R^2 = 0.019$ for all linkage groups in the NDC population. When performing the regression of pairwise LD distances against the number of intervals between markers according to the map order, the estimated values $\hat{\lambda}_{map}$ were all negative or null, except for linkage group LG13 in population SDD where $\hat{\lambda}_{map} = 0.018$. Hence, we do not expect to find a correlation between map order and the best order in this data set. In fact, when markers were ordered according to the genetic map for domesticated *H. annuus*, high LD values often were found between markers that were far apart on the genetic map (Figure 5).

LD matrices – NEG population – LG13

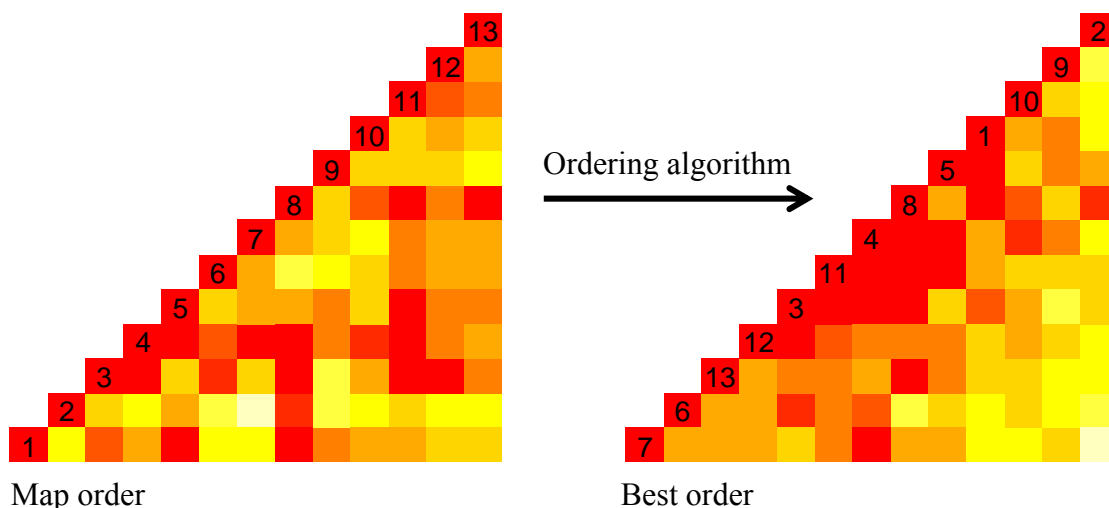


Figure 5: Matrices of LD (R_{ij}^2) in the wild *H. annuus* population from Nebraska (NEG). For the matrix on the left markers were ordered according to the genetic map, whereas markers in the matrix on the right was ordered by our algorithm. The strength of LD is proportional to the strength of red coloration.

After running the algorithm, pairs of markers with the highest LD were grouped together, and the regression of LD distance against the number of intervals between markers for the best order was significant in all populations and for all linkage groups, with $\hat{\lambda}_{best}$ ranging from -1.249 to 2.698. However, the mean squared error SSE_{best} was large and ranged between 0.020 and 42.727. Not surprisingly, none of the rank correlations between map order and best order were significant.

Despite the fact that the best orders found had little to do with the supposed map order, the results were consistent between the three populations. As illustrated on Figure 6 for linkage group LG13, we were able to identify, on each linkage group, two to three subsets of markers that were found close together under the best order of each population.

A subset was defined as a group of markers that were found adjacent in all populations, although inversions within the subset were allowed. Hence, for a subset of n markers, the number of intervals between any two markers of the subset never exceeded $(n-1)$ in all populations. For linkage group LG10, the subsets $\{5,9\}$ and $\{2,3,7,8\}$ were found in all populations. For linkage group LG13, the subsets of markers $\{1,5,10,12\}$, $\{2,9\}$, and $\{4,6,8,11,13\}$ were found in all populations, except in population NEG where marker 12 was found elsewhere, and the set $\{4,6,8,11,13\}$ was split into $\{6,13\}$ and $\{4,8,11\}$. For linkage group LG17, the sets $\{2,5,6,7,15\}$, $\{1,12,13,16\}$, and $\{4,8,11\}$ were found in all populations, except markers 8 and 15, which were placed elsewhere in population NDC. For these subsets of markers, the LD pattern was therefore highly consistent between populations.

To determine whether the LD pattern and the ordering algorithm were sensitive to variation in genetic diversity between markers, we also applied the algorithm to the subset of markers with the highest gene diversity ($He > 0.8$), as well as to the subset of markers with the lowest gene diversity ($He < 0.8$), whenever the number of markers with a low He remaining was not too low (i. e., for LG17 in population SDD). The best orders found with the marker subsets were the same as the best orders found with all the markers.

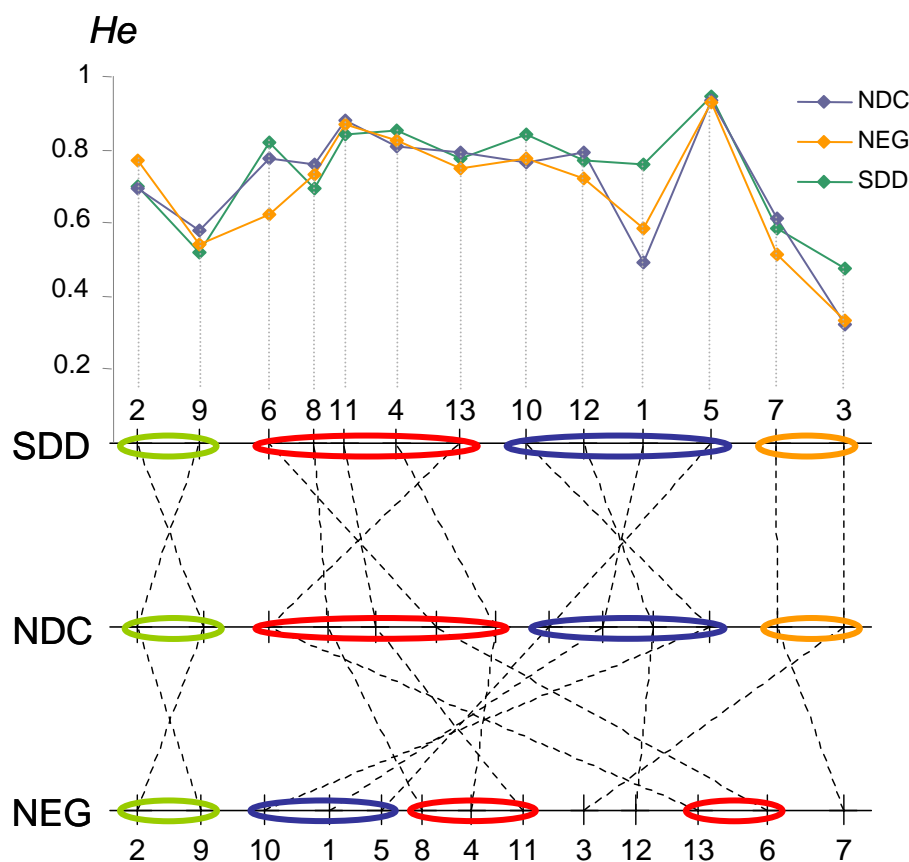


Figure 6: The pattern of gene diversity, He , along linkage group 13 for the South Dakota (SDD) population of *Helianthus annuus* in the upper part of the figure and the “best order” positions of 13 genotyped microsatellites in the three *H. annuus* populations in the lower part of the figure. The congruence of marker positions between populations (SDD, NDC: North Dakota, NEG: Nebraska) are indicated by colored circles. Distances between markers in the three wild populations are LD standardized distances. Microsatellites have been named according to their position on the domesticated *H. annuus* map (marker 1 is located at 0.6 cM, 2, 3 and 4 at 1.5 cM, 5 at 5 cM, 6 at 24.5 cM, 7 at 30.2 cM, 8 at 31.2 cM, 9 at 35 cM, 10 at 40 cM, 11 at 57.4 cM, 12 at 69.8 cM and 13 at 85 cM).

Discussion

In this paper, we proposed a method based on pairwise LD patterns to order a set of markers in a population when the genetic map is unknown. The main hypotheses of our model include, i) there will be a linear increase of the LD distance expressed as $D_{ij} = -\log R_{ij}^2$ with the genetic map distance, and ii) the genetic map distance is proportional to the number of intervals between markers. Instead of constructing a likelihood function for the observed LD matrix, we searched for the best order by maximizing the slope of the regression of the LD distance over the interval distance. A steepest descent algorithm was implemented to explore the parameter space (i.e., all the different orders). All scripts were written in R language and are available on request.

Algorithms that order data points from pairwise distance or similarity matrices already exist. For example multidimensional scaling (MDS) locates data points in a multidimensional space such that the Euclidean distances between these points minimize distortions in the distance matrix. The major difference between MDS and the algorithm developed in this study is that MDS gives the same weight to all data, whereas our algorithm considers an explicit function for the increase of the observed distance with the distance between data points. We compared MDS to LD distances with our method; our method more efficiently retrieved the correct marker order.

We analyzed the efficiency of our algorithm on three different data sets. First, LD matrices were simulated according to equation (7) with a linear increase of LD distances with interval distances and random noise. Second, maize RIL populations were chosen as an experimental case where the increase of LD distance with interval distance is known to be linear (equation 3). Third, Monte-Carlo simulations were performed under a mutation/drift model to test whether the linear model (7) could accurately approximate the increase of LD distance with interval distance in “natural” populations. These three types of data allowed us to explore both the potential utility and limits of our method.

The linear approximation used is not the best model to describe the relationships between LD distance and genetic distance in the simulated populations (Figure 4), but it fitted the data well enough to support our approach. Although the residuals may not be independent from genetic distance in simulated data sets (see for example Figure 4), the power predicted by simulating random LD matrices was highly consistent with our analyses of both the maize RIL and simulated populations. This shows that our approach is robust to departures from the assumptions of the linear regression model. With the data sets that we analyzed, the range of

values for λ and SSE were such that the power was 80% for the maize RILs and 60% for the simulated populations.

The ordering algorithm proved to be efficient when λ was high and SSE low, but there were two caveats. First, by maximizing the slope of the regression curve instead of a likelihood function, high LD values between individual pairs of markers were too heavily weighted. This led to the frequent pairing of genetically distant markers that exhibit spurious associations that presumably result from drift and/or selection. For example, in simulated natural populations, approximately one third of markers were misplaced, apparently due to drift effects alone. Second, the use of interval numbers rather than genetic distances makes the algorithm overly sensitive to departures from an even distribution of markers along the genetic map. While the algorithm accurately retrieved the map order in maize RIL populations, markers separated by large gaps were often inverted by our ordering algorithm.

The choice of linear regression to model the decrease of LD with genetic distance was motivated by the statistical properties of linear regression: estimation of parameters was fast and reliable and no hypotheses about the history of the population, mutation rate, effective population, or mutation mode were required. All of these factors influence our parameter λ . A profitable focus of future work might be to disentangle the effects of these factors λ to better understand the conditions under which LD patterns may be used to order the markers.

For the wild *H. annuus* populations, we were outside of the parameter range required for accurate ordering of markers. However, our method allowed us to find a marker order consistent with the LD pattern. Computer programs such as LDhat (McVean *et al.* 2002) estimate recombination rates between markers from LD patterns in natural populations, given a user-defined order. LDhat also identifies inconsistencies in LD pattern such as incongruence between LD distance and genetic map distance. Possibly, our method could be used as a complement to LDhat to identify the LD order of the markers when such inconsistencies exist in an experimental data set.

No significant correlations were found between the genetic map order for domesticated sunflower and the LD best order obtained from natural populations of *H. annuus*. This could be due to a lack of colinearity between domesticated and natural populations. However, this seems unlikely given the conservation of synteny observed among closely-related species in other plant groups such as the cereals, for which a collinear “crop circle” has been established (Moore *et al.* 1995; Devos 2005). Also, crosses between wild and domesticated sunflower are fully fertile, suggesting that major differences in gene order are unlikely (Heiser 1954; Burke *et al.* 2002). Alternatively, the incongruence between LD

patterns and genetic map distance is due to shared population history and possibly selection. The LD best orders were highly consistent across the three natural populations, all of which are from the North of the USA: Nebraska, North and South Dakota, which suggests a common and recent history for these three populations. Also many of the markers used here were developed from expressed sequence tags (ESTs) and might therefore be affected by selection on the genes they are associated with.

In conclusion, with a simple linear model approximating the increase of LD distance with interval distance we were able to order markers by strength of LD in several types of populations: maize RILs, simulated populations, and wild *H. annuus* populations. The ordering algorithm had proved to be efficient when λ is high and *SSE* low. The best orders found by our algorithm were close to the map order for the RILs and simulated populations. In the wild *H. annuus* populations, however, λ and *SSE* were not in a good range to accurately order markers, and our algorithm failed to find the correct orders. These results indicate that factors other than genetic distance have a strong influence on patterns of LD in natural populations. This problem may prove insurmountable for genetic mapping in natural populations, except perhaps in admixed populations (Burke *et al.* 2002).

Bibliography

- Aranzana MJ, Kim S, Zhao K, *et al.* (2005) Genome-wide association mapping in *Arabidopsis* identifies previously known flowering time and pathogen resistance genes. *PLoS Genetics*, **1**, e60.
- Bennett JH (1954) On the theory of random mating. *Annals of Eugenics*, **18**, 311-317.
- Breseghele F, Sorrells ME (2006) Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. *Genetics*, **172**, 1165-1177.
- Burke JM, Tang S, Knapp SJ, Rieseberg LH (2002) Genetic analysis of sunflower domestication. *Genetics*, **161**, 1257-1267.
- Causse M, Santoni S, Damerval C, *et al.* (1996) A composite map of expressed sequences in maize. *Genome*, **39**, 418-432.
- Devos KM (2005) Updating the 'crop circle'. *Current Opinion in Plant Biology*, **8**, 155-162.
- Dunning AM, Durocher F, Healey CS, *et al.* (2000) The extent of linkage disequilibrium in four populations with distinct demographic histories. *American Journal of Human Genetics*, **67**, 1544-1554.
- Einarsdottir K, Humphreys K, Bonnard C, *et al.* (2006) Linkage disequilibrium mapping of *CHEK2*: common variation and breast cancer risk. *PLoS Medicine*, **3**, e168.
- Ellegren H (2004) Microsatellites: simple sequences with complex evolution. *Nature Reviews. Genetics.*, **5**, 435-445.
- Farnir F, Coppieters W, Arranz JJ, *et al.* (2000) Extensive genome-wide linkage disequilibrium in cattle. *Genome Research*, **10**, 220-227.
- Garnier-Gere P, Dillmann C (1992) A computer program for testing pairwise linkage disequilibria in subdivided populations. *Journal of Heredity*, **83**, 239.
- Hedrick PW (1987) Gametic disequilibrium measures: proceed with caution. *Genetics*, **117**, 331-341.
- Heiser CB (1954) Variation and subspeciation in the common sunflower, *Helianthus annuus*. *American Midland Naturalist*, **51**, 287-305.
- Hospital F, Chevalet C (1993) Effects of population size and linkage on optimal selection intensity. *Theoretical and Applied Genetics*, **96**, 775-780.
- Hudson RR (2001) Two-locus sampling distributions and their application. *Genetics*, **159**, 1805-1817.
- Huttley GA, Smith MW, Carrington M, O'Brien SJ (1999) A scan for linkage disequilibrium across the human genome. *Genetics*, **152**, 1711-1722.
- Kruglyak S, Durrett RT, Schug MD, Aquadro CF (1998) Equilibrium distributions of microsatellite repeat length resulting from a balance between slippage events and point mutations. *Proceedings of the National Academy of Sciences, U S A*, **95**, 10774-10778.
- Lander ES, Green P, Abrahamson J, *et al.* (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics*, **1**, 174-181.
- Lewontin RC (1964) The Interaction of Selection and Linkage II: Optimum Models. *Genetics*, **50**, 757-782.
- Lewontin RC (1974) The problem of genetic diversity. *Harvey Lectures*, **70 Series**, 1-20.
- Maruyama T (1982) Stochastic integrals and their application to population genetics. In: *Molecular Evolution, Protein Polymorphism and the Neutral Theory* (ed. Kimura M), pp. 151-166. Japan Scientific Societies Press, Tokyo.
- McVean G, Awadalla P, Fearnhead P (2002) A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics*, **160**, 1231-1241.
- Moore G, Devos KM, Wang Z, Gale MD (1995) Cereal genome evolution. Grasses, line up and form a circle. *Current Biology*, **5**, 737-739.

- Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, **89**, 583-590.
- Olsen KM, Halldorsdottir SS, Stinchcombe JR, *et al.* (2004) Linkage disequilibrium mapping of *Arabidopsis* *CRY2* flowering time alleles. *Genetics*, **167**, 1361-1369.
- R Development Core Team (2004) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248-249.
- Rieseberg LH, Buerkle CA (2002) Genetic mapping in hybrid zones. *American Naturalist*, **159S**, 159-S150.
- Rieseberg LH, Whitton J, Gardner K (1999) Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics*, **152**, 713-727.
- Sandor C, Farnir F, Hansoul S, *et al.* (2006) Linkage disequilibrium on the bovine X chromosome: characterization and use in quantitative trait locus mapping. *Genetics*, **173**, 1777-1786.
- Stephan W, Song YS, Langley CH (2006) The hitchhiking effect on linkage disequilibrium between linked neutral loci. *Genetics*, **172**, 2647-2663.
- Storz JF (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology*, **14**, 671-688.
- Tang S, Yu JK, Slabaugh B, Shintani K, Knapp J (2002) Simple sequence repeat map of the sunflower genome. *Theoretical and Applied Genetics*, **105**, 1124-1136.
- Tenesa A, Wright AF, Knott SA, *et al.* (2004) Extent of linkage disequilibrium in a Sardinian sub-isolate: sampling and methodological considerations. *Human Molecular Genetics*, **13**, 25-33.
- Thornsberry JM, Goodman MM, Doebley J, *et al.* (2001) *Dwarf8* polymorphisms associate with variation in flowering time. *Nature Genetics*, **28**, 286-289.
- Weir BS (1979) Inferences about linkage disequilibrium. *Biometrics*, **35**, 235-254.
- Yatabe Y, Rieseberg LH (In review) Persistent gene flow and adaptive differentiation between the annual sunflowers, *Helianthus annuus* and *H. petiolaris*. In: *Genetics*.

III. Signature de sélection à l'échelle d'un groupe de liaison chez *H. paradoxus*

Pour analyser la signature de la sélection à une échelle génétique plus réduite que le génome (chapitre 1), nous avons étudié le polymorphisme de marqueurs présents sur les trois groupes de liaison contenant les QTLs de survie. L'absence de carte génétique détaillée chez l'espèce hybride *H. paradoxus* et d'un nombre important de marqueurs communs entre carte génétique nous limitait dans cette étude. Pour cette raison, nous avons ordonné les marqueurs en utilisant la méthode développée précédemment basée sur le déséquilibre de liaison. Puis, nous avons analysé les variations de polymorphisme le long de chaque groupe de liaison dans les trois populations de *H. paradoxus*.

1. Résultats

1.1. Polymorphisme des marqueurs microsatellites dans les populations naturelles de *H. paradoxus*

Au total, 42 nouveaux marqueurs microsatellites ont été génotypés sur 192 individus collectés dans trois populations naturelles de *H. paradoxus* (voir chapitre 1 pour le détail sur les populations). Chaque microsatellite est localisé sur au moins une carte génétique de *Helianthus* (Figure 2.1 en début de chapitre). Sept microsatellites appartiennent à des groupes de liaisons déterminés, sans avoir été néanmoins précisément cartographiés. 11 marqueurs sont localisés sur le LG1, 20 sur le LG4 et 11 sur le LG17. Parmi ces 42 marqueurs, 22 (les marqueurs nommés HT) ont été développés à partir d'EST.

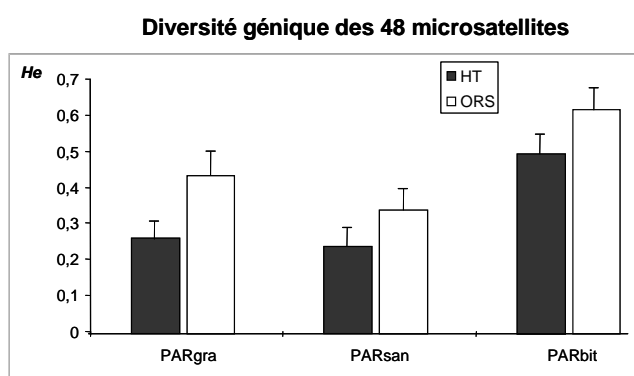


Figure 2.2 : Diversité génique, ou hétérozygotie attendue He moyenne avec les erreurs types, dans les trois populations de *H. paradoxus* de Grants (PARgra), Santa Rosa (PARsan) et Bitter Lake (PARbit), sur les 48 microsatellites localisés sur les groupes de liaison 1, 4 et 17.

En moyenne, les 42 nouveaux microsatellites choisis sont plus polymorphes dans la population de *H. paradoxus* de Bitter Lake (PARbit, Figure 2.1) : $He = 0.553$ pour PARbit

contre $He = 0.346$ pour PARgra et $He = 0.288$ pour PARsan. Les microsatellites situés dans des EST sont moins polymorphes que les autres pour l'ensemble des trois populations de *H. paradoxus* (Figure 2.2).

Pour ordonner les microsatellites nous avons appliqué l'algorithme d'ordonnement décrit précédemment sur les données de génotypage de PARbit car c'est la population la plus polymorphe. Quand il n'y a pas de polymorphisme, le déséquilibre de liaison est inexistant, ou du moins non décelable. Néanmoins certains microsatellites étaient monomorphes sur les 64 individus génotypés dans la population PARbit et ont été retirés de l'étude (HT816 et ORS543 sur LG1, HT327 et ORS309 sur LG4 et HT719 sur LG17). Finalement 43 marqueurs ont été utilisés pour faire l'analyse de cartographie.

1.2. Cartographie des 43 marqueurs par analyse de DL.

Lors de la formation de groupes de microsatellites en déséquilibre de liaison à l'aide des 'P values' associées aux valeurs R_{ij}^2 , certains microsatellites n'étaient pas en déséquilibre de liaison : 10 marqueurs microsatellites étaient en déséquilibre de liaison dans chacun des groupes de liaison LG1, LG4 et LG17 et ont pu être ordonnés (Figure 2.3), les autres n'ont pas pu être cartographiés. A noter que ORS235, un microsatellite proche du QTL de survie sur le LG4, n'était pas en déséquilibre de liaison avec les autres marqueurs.

La coïncidence entre l'ordre des marqueurs sur la carte de *H. paradoxus* trouvée par analyse de DL ou celui des cartes préalablement développées (Carte BC2, carte de *H. annuus*, et carte de *H. paradoxus*) dépend du groupe de liaison (Figure 2.3 et 2.4). Pour LG1 et LG17, les ordres des marqueurs obtenus par analyse de DL ne coïncident pas avec ceux des cartes connues (Figure 2.1 et 2.3). Ceci confirme les résultats chez *H. annuus* qui montrent que nous ne sommes pas dans de bonnes conditions de décroissance de DL avec la distance pour construire une carte génétique sur la base de celui-ci. Par contre, pour LG4, l'ordre des marqueurs obtenu à partir des données de populations naturelles est bien le même que celui observé sur les cartes génétiques basées sur les taux de recombinaisons, à l'exception du marqueur HT317, mais il est situé à un 1cM du bloc auquel il appartient et le marqueur ORS439 qui est en bout de groupe de liaison sur la carte de *H. paradoxus* (Figure 2.4). A noter que les marqueurs en déséquilibre de liaison sur LG4 sont compris dans une distance génétique plus faible (20cM) que ceux sur LG1 (70cM) et LG17 (60cM).

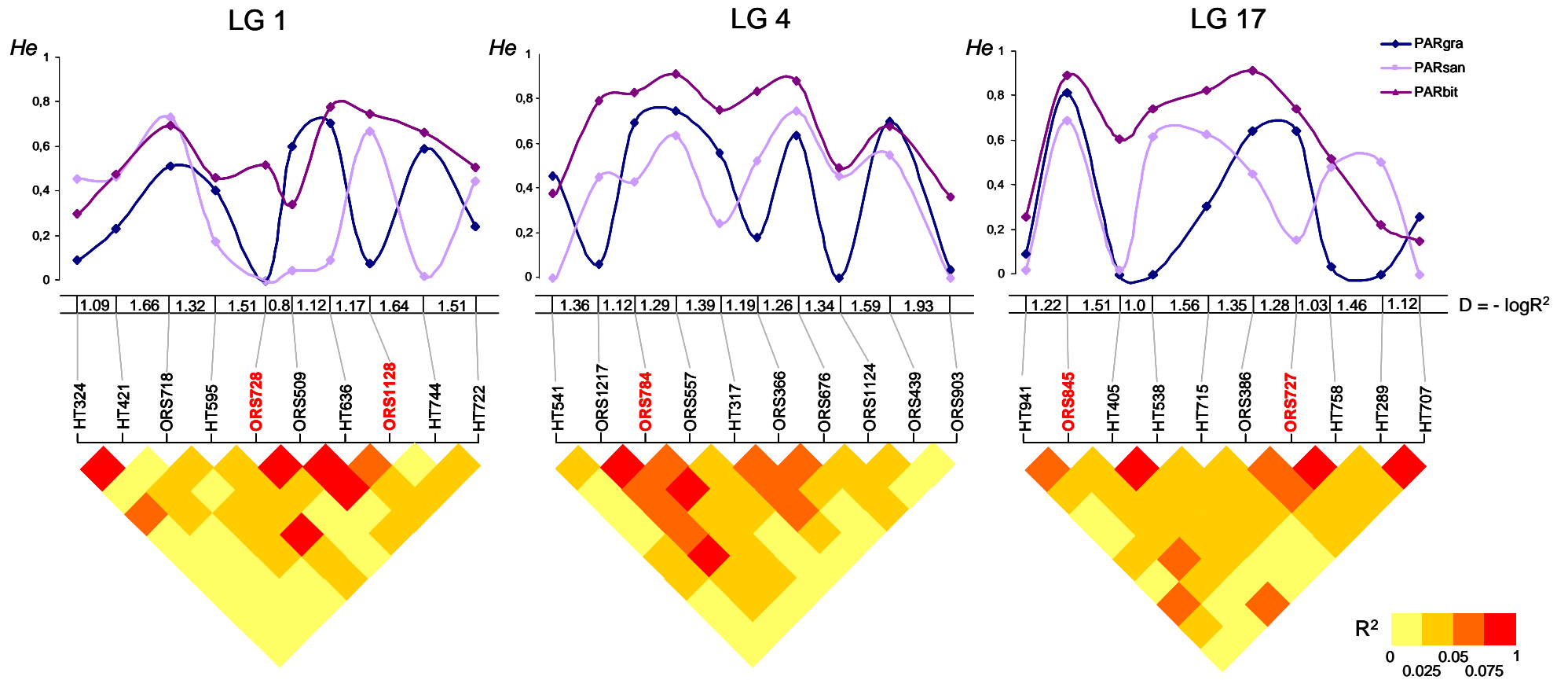


Figure 2.3 : Matrice de R_{ij}^2 de la population PARbit une fois les marqueurs ordonnés pour chacun des groupes de liaison LG1, LG4 et LG17. Les marqueurs en rouges sont ceux localisés à proximité de QTL de survie. La distance $D_{ij} = -\log R_{ij}^2$ entre marqueurs ordonnés est schématisée et le patron de diversité génétique est indiqué pour chacune des populations de *H. paradoxus* PARgra, PARsan et PARbit.

Notons que les microsatellites liés aux QTL de survie se comportent comme des « attracteurs » de DL, en effet c'est avec eux, ou aux microsatellites adjacents une fois les marqueurs ordonnés, que les valeurs de DL sont les plus importantes.

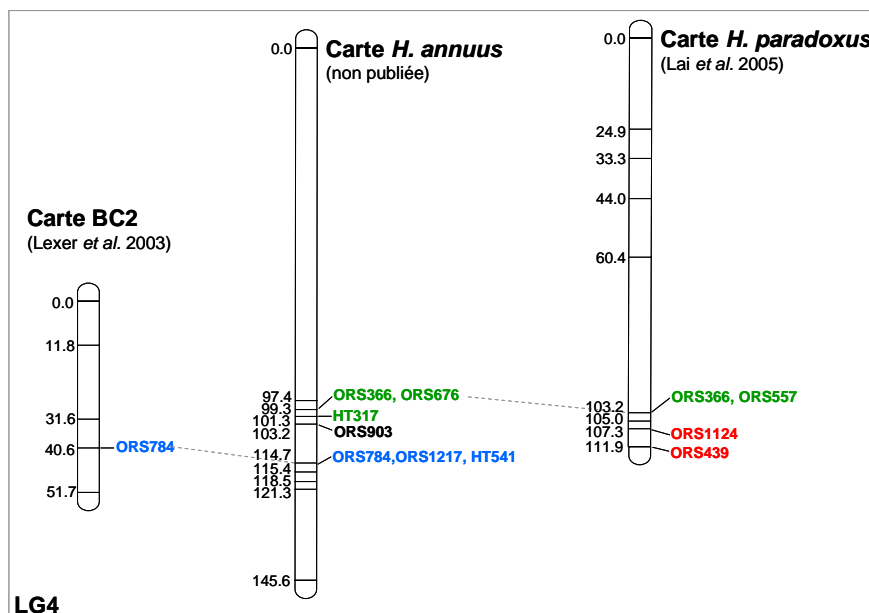


Figure 2.4. Localisation sur les cartes génétiques disponibles des groupes de déséquilibre de liaison obtenues par la méthode de cartographie développée pour le groupe de liaison 4. L'ordre des marqueurs obtenu par la nouvelle méthode de cartographie est présenté Figure 2.2. Trois groupes de marqueurs: bleu (HT541, ORS1217, ORS784), vert (ORS 557, HT317, ORS366 et ORS676), rouge (ORS1124 et ORS439) se retrouvent correctement placés sur les cartes génétiques. Le dixième marqueur est indiqué en noir. La discordance entre la localisation des marqueurs sur les cartes génétiques et la carte de déséquilibre de liaison pour les autres groupes de liaison ne permet pas de présenter le même schéma.

1.3. Variation de la diversité génétique le long de chaque groupe de liaison.

Les variations de H_e le long des trois groupes de liaison sont présentées Figure 2.3. La diversité génique sur les marqueurs ordonnés fluctue le long du groupe de liaison : nous n'observons pas un balayage sélectif unique, mais plutôt une mosaïque de diversité. Les variations de diversité ne sont pas similaires pour les trois populations naturelles. Il s'agit donc de profils complexes qu'il est difficile d'expliquer.

2. Discussion

Nous avons réussi à ordonner des marqueurs sur chacun des groupes en utilisant un algorithme d'ordonnement basé sur le déséquilibre de liaison. Comme nous l'avons montré dans la section précédente de ce chapitre, cet algorithme ne garantit pas de retrouver l'ordre de la carte génétique. Par contre il est efficace pour regrouper les marqueurs en DL. Les marqueurs non polymorphes n'ont pas pu être pris en compte et certains microsatellites,

comme ORS235, un marqueur proche du QTL de survie sur le groupe de liaison 4, étaient probablement trop peu polymorphes pour être en DL significatif avec un autre marqueur du groupe de liaison.

Un résultat intéressant est que cinq des six microsatellites liés aux QTL de survie sont en déséquilibre de liaison avec d'autres marqueurs et même semble être des attracteurs de déséquilibre de liaison. Le fort DL dans ces régions expliquerait pourquoi nous avons été capables de mettre en évidence des balayages sélectifs de plusieurs centiMorgans autour de QTL.

La présence de deux types de marqueurs génotypés, localisés dans des EST ou non, et dont le polymorphisme moyen diffère complique l'interprétation des variations de diversité le long des groupes de déséquilibre de liaison. Pour connaître les marqueurs réellement sous sélection, il serait judicieux de génotyper l'ensemble de ces microsatellites chez des individus des deux espèces parentales et de travailler par exemple sur le ratio d'hétérozygotie attendue $lnRH$. Cela permettrait de s'affranchir des variations de diversité génétiques entre locus dues à des variations du taux de mutation.

On peut toutefois noter que la diversité génétique des microsatellites de type ORS génotypés dans cette étude est comparable à celle des six microsatellites liés aux QTL de survie, ce qui est un autre argument en faveur d'une diminution globale de la diversité génétiques dans ces régions.

CHAPITRE 3

Expression différentielle de gènes candidats potentiellement impliqués dans la réponse à la salinité chez trois espèces de tournesol : l'halophyte *Helianthus paradoxus* et ses parents glycophytes *H. annuus* et *H. petiolaris*

Expression différentielle de gènes candidats potentiellement impliqués dans la réponse à la salinité chez trois espèces de tournesol : l'halophyte *Helianthus paradoxus* et ses parents glycophytes *H. annuus* et *H. petiolaris*

L'expérience en serre a été réalisée par Sophie Karrenberg, Loren Rieseberg et moi-même à Bloomington, Indiana University. C'est ensuite à l'UMR de génétique végétale, à la Ferme du Moulon (Gif sur Yvette), que se sont poursuivies les analyses d'expression génique par RT-PCR semi-quantitative avec Matthieu Falque et Xavier Raffoux.

Introduction

La présence d'eau est une condition incontournable pour que toute plante puisse se développer et assurer ses fonctions physiologiques vitales. Cette ressource n'est pas toujours facile d'accès dans le sol, en fonction du milieu naturel. Les plantes présentes sur des surfaces sèches ou salées vont se retrouver exposées à un stress hydrique important, contre lequel elles devront lutter pour survivre. Dans le cas d'un stress salin, une double problématique se pose à l'organisme végétal : d'un côté, la présence de sel, en abaissant le potentiel hydrique du sol, menace l'approvisionnement en eau de la plante, de l'autre, l'absorption de sel dans les tissus menace le bon fonctionnement physiologique des cellules. Face à ce danger, toutes les plantes ne sont pas égales. Certaines, les glycophytes (du grec glycos = sucré et phytos = plante), ne sont pas capables de supporter la présence de sel. Les halophytes (du grec halos = sel), au contraire, ont développé des réponses physiologiques pour assurer leur approvisionnement en eau tout en préservant leur métabolisme. Une compréhension des mécanismes de réponse au stress salin ouvrirait une voie à une meilleure maîtrise des pratiques agronomiques en milieu saumâtre.

Ainsi, la première difficulté d'une plante en milieu salin est d'assurer son apport en eau. Pour cela, il faut que la plante puisse ajuster la pression osmotique de ses tissus par rapport à la pression osmotique du sol. Elle doit aussi faire face aux changements de concentration ionique. Les plantes halophytes présentent un pouvoir d'ajustement osmotique et ionique élevé, renforcé par l'absorption de sel, et la conduite en sel vers les feuilles, contrairement aux glycophytes, dont un excès de sel dans le cytoplasme provoque des perturbations dans la balance ionique et une prise en potassium, nitrates et calcium réduite. Les mécanismes de réponses au stress salin sont donc de deux ordres : une réponse au stress ionique pour exclure le sel et maintenir les balances ioniques intracellulaires et une réponse au

stress osmotique pour maintenir l'hydratation, la turgescence des tissus et / ou protéger contre la réduction d'hydratation cellulaire (Chinnusamy *et al.* 2005). Il existe également une réponse au stress oxydatif à la suite de l'accumulation de ROS (Reactive Oxygen Species) que l'on ne détaillera pas plus ici (Figure 3.1).

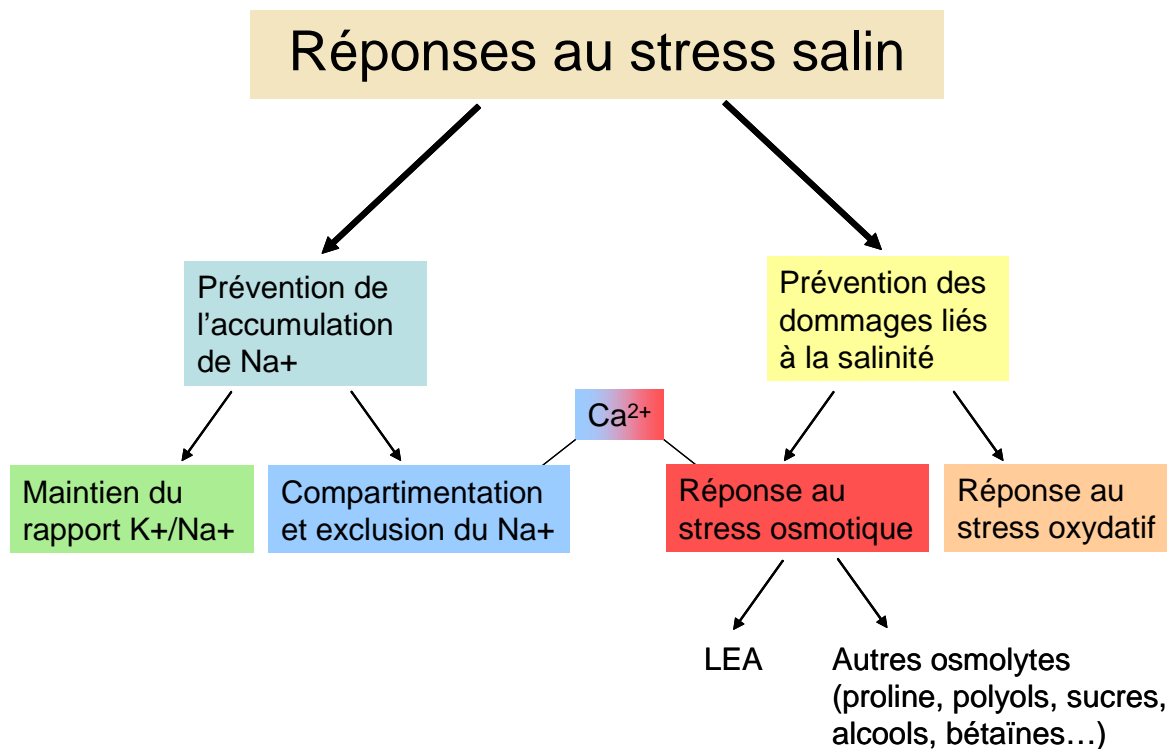


Figure 3.1 : Les différents mécanismes pouvant être mis en place chez les plantes pour répondre au stress salin. LEA = Late embryogenesis Abundant

Le maintien de l'homéostasie ionique est primordial pour le bon fonctionnement de la cellule chez les plantes, notamment le rapport de K⁺/Na⁺ dans le cytosol doit être élevé. Or le Na⁺ est en compétition avec le K⁺ pour entrer dans la cellule à travers des transporteurs Na⁺ - K⁺ et peut bloquer des transporteurs spécifiques de K⁺ lors de stress salin. Accroître l'entrée de K⁺ à travers des transporteurs de K⁺ peut aider à maintenir le ratio K⁺/Na⁺ en conditions de stress salin (Figure 3.2). Restreindre l'entrée de Na⁺ est un autre moyen de maintenir le ratio K⁺/Na⁺ dans le cytosol, cela peut se faire par une affinité moindre des transporteurs de Na⁺. Enfin, pour prévenir de toute intoxication par l'accumulation de Na⁺ dans le cytosol il existe des voies d'exclusion du Na⁺ (Figure 3.2). L'entrée de Na⁺ par des canaux non spécifiques dépolarise la membrane et induit la présence de signaux calcium dans le cytosol. Les oscillations de Ca²⁺ activent une cascade de protéines SOS (Salt Overly Sensitive). La protéine SOS3 sur laquelle se fixe le Ca²⁺ interagit avec SOS2, qui est une sérine/thréonine protéine kinase. Les cibles du complexe SOS3/SOS2 sont des transporteurs antiport

membranaire Na^+/H^+ , qui exportent les ions sodiums hors de la cellule (SOS1), ou les stockent dans des vacuoles (NHX1) (Figure 3.2).

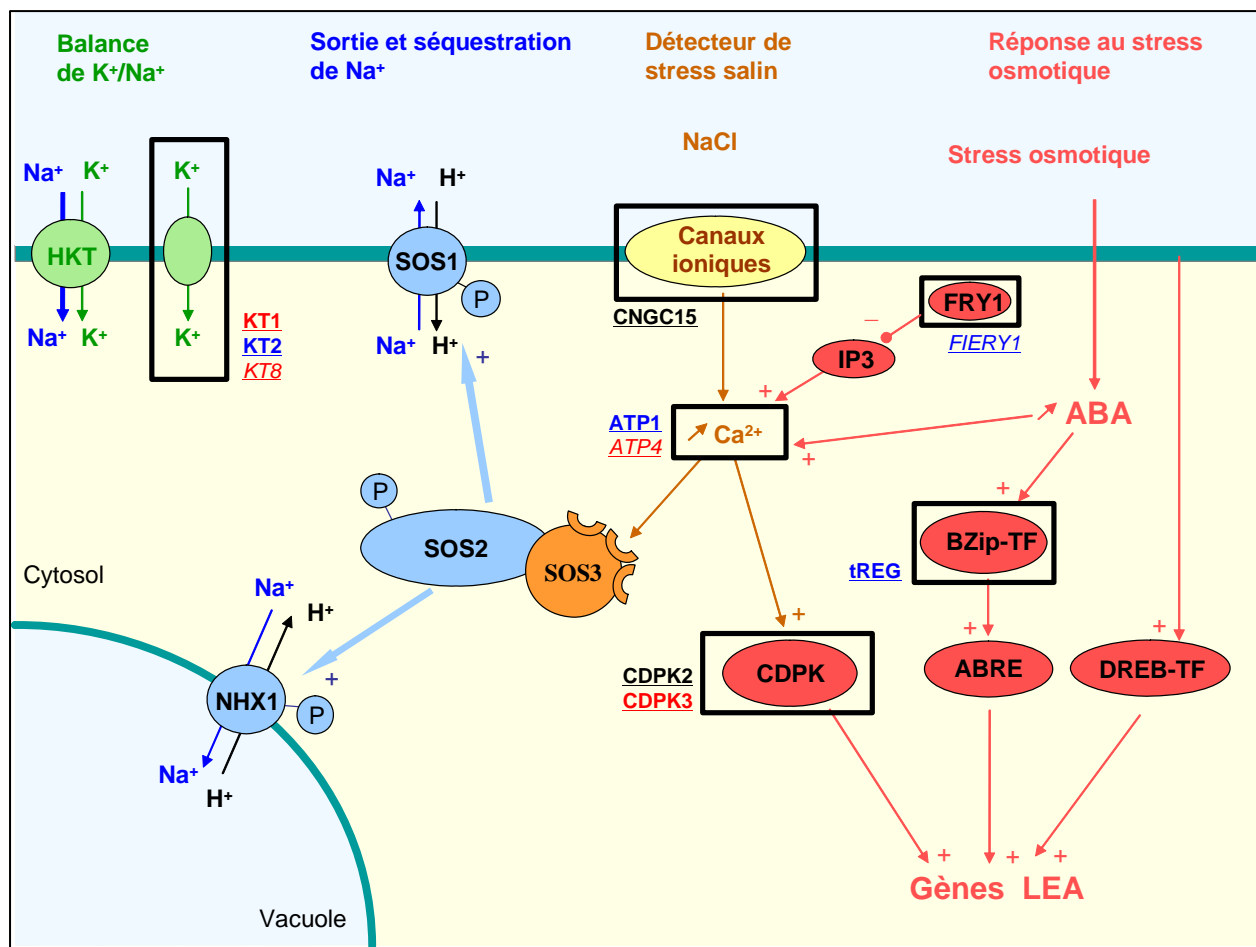


Figure 3.2 : Réponse au stress salin chez les plantes, modifié de Chinnusamy *et al.* 2005.

1- La réponse au **stress ionique** consiste **a-** à maintenir la balance intracellulaire K^+/Na^+ à l'aide de transporteurs membranaires, comme les transporteurs HKT (High affinity K^+ Transporter) ; **b-** **séquestrer** dans des vacuoles ou **exclure le Na^+ intracellulaire** à l'aide du système SOS (Salt Overly Sensitive) et de transporteurs membranaires antiport Na^+/H^+ (NHX1).

2- La réponse au **stress osmotique** passe par **a- l'augmentation de Ca^{2+} intracellulaire** qui est déclenchée par des canaux ioniques membranaires ; la synthèse d'ABA (Abscisic Acid) et l'IP3 (Inositol tris-phosphate) déclenchent également la libération de Ca^{2+} intracellulaire. L'IP3 peut être dégradé par une inositol polyphosphate phosphatase FRY1. Le Ca^{2+} agit comme un messager et active le système SOS, ainsi que des protéines CDPK (Calcium-Dependant Protein Kinase). Les protéines CDPK induisent à leur tour l'expression des protéines LEA (Late Embryogenesis Abundant). Ces protéines ont un rôle protecteur chez les plantes soumises à un stress salin. **b- l'augmentation intracellulaire de l'ABA** Cette augmentation conduit à l'augmentation de Ca^{2+} et plus directement active un facteur transcriptionnel BZip-TF qui induit à son tour la synthèse de protéine LEA via la synthèse d'ABRE (ABA-Responsive elements). Le facteur transcriptionnel DREB-TF est lui directement induit par le stress osmotique et provoque également la synthèse de protéine LEA.

Les 10 gènes candidats de notre étude sont indiqués 'soulignés' au niveau de leur implication et fonction supposée. En rouge les gènes candidats pour lesquels l'expression est plus élevée chez *H. paradoxus*, en bleu ceux pour lesquels l'expression est au contraire moins élevée chez *H. paradoxus*. Les trois gènes en italique indique une différence d'expression significative entre *H. paradoxus* et *H. petiolaris* mais non significative entre *H. paradoxus* et *H. annuus*.

Pour prévenir de la déshydratation du cytosol qui nuirait au fonctionnement du métabolisme, les halophytes et aussi occasionnellement des glycophytes, luttent en produisant des osmolytes ou osmoprotecteurs. Ces composés (acides aminés, ions ammoniums, carbohydrates...), par leur concentration, assurent l'ajustement osmotique entre le cytosol et la vacuole ou le milieu extracellulaire. Leur accumulation dans le cytosol est accompagnée d'une baisse de la concentration de solutés moins compatibles, comme les sels, et d'une augmentation du volume d'eau du cytosol. Parmi les osmoprotecteurs, on notera les protéines LEA (Late Embryogenesis Abundant) (Figure 3.2). Leur accumulation est corrélée à une tolérance au stress salin, suggérant un rôle protecteur lors d'un stress osmotique. Ces protéines LEA sont codées par plusieurs gènes, dont l'expression dépend de voies de transduction complexes induites par le Ca^{2+} et l'ABA (Abscisic Acid). Parmi les régulateurs d'expression, mentionnons les CDPK (Calcium Dependant Proteine Kinase) et des régulateurs contenant des doigts de zinc, tel le facteur de transcription BZip (Figure 3.2).

L'adaptation aux milieux salins a déjà été largement étudiée ces dernières décennies chez l'espèce modèle *Arabidopsis thaliana* et des espèces d'intérêt agronomiques, principalement le riz (revue de Sahi *et al.* 2006) du fait de son système d'irrigation. La profusion des outils génétiques a permis de développer des approches expérimentales à tous les niveaux. Les études physiologiques de réponse au stress salin, dont on vient de décrire les bases cellulaires, ont montré que la réponse au stress salin impliquait de multiples voies et par conséquent un large nombre de gènes. Des études de recherche de QTL liés à la tolérance au stress salin (QTL de prise en ions, de concentrations ioniques, de croissance en milieu salin...) dans des populations de RIL de riz (Flowers *et al.* 2000 ; Koyama *et al.* 2001) ou encore des backcross de tournesols (Lexer *et al.* 2003a ; Rieseberg *et al.* 2003) ont confirmé le nombre important de gènes impliqués dans la réponse à ce stress. Le développement de bases d'EST a permis d'identifier des gènes dont le niveau d'expression change en réponse à la concentration saline (Sahi *et al.* 2003 ; Shiozaki *et al.* 2005), ou entre espèces glycophytes et leurs relatives halophytes (Taji *et al.* 2004). Des approches transgéniques par mutations ciblées ont permis de caractériser un certain nombre de gènes (voir exemples dans la revue de Sahi *et al.* 2006 ; Chinnusamy *et al.* 2005) et d'avoir une idée plus claire des gènes intervenant dans la réponse au stress ionique et osmotique (Figure 3.2). Ainsi, le recoupement des données d'expression d'EST, de séquençage, de gènes ciblés mutés, et l'utilisation d'outils bioinformatiques, fournissent l'opportunité de caractériser plus en profondeur le réseau des voies de réponse au stress salin (Bohnert *et al.* 2001 ; Sahi *et al.* 2006).

Bien que des progrès considérables aient été effectués dans la compréhension de l'adaptation aux stress salins, il est encore difficile de prédire, étant donné la complexité des réponses mises en place, les mécanismes adaptatifs mis en place par les espèces en réponse à ce stress. Notre système d'étude est particulièrement bien adapté pour étudier l'adaptation d'une espèce à un milieu salin : deux espèces parentales *H. annuus* et *H. petiolaris*, glycophytes toutes les deux, ont donné naissance à *H. paradoxus* une espèce halophyte qui vit exclusivement dans des marais salins du Nouveau Mexique et du Texas. Cela permet de comparer facilement les réponses au stress salin entre espèces très voisines. Cette comparaison peut se faire à différents niveaux : au niveau anatomique, comme au niveau physiologique ou génétique.

D'un point de vue anatomique et physiologique, plusieurs études ont déjà été réalisées chez ces trois espèces. Une première étude (Welch et Rieseberg, 2002) en chambre de croissance, où les trois espèces ont été soumises à trois traitements salins, divergents uniquement par leur concentration en NaCl, a montré que *H. paradoxus* est plus tolérante au NaCl que chacun des ses parents. La survie de *H. paradoxus* n'est pas affectée par de fortes concentrations en NaCl. L'efficacité de sa prise en eau est associée à une fraction racinaire plus développée et à une baisse de la conductance à travers les stomates. Comme beaucoup de plantes halophytes ses feuilles sont très succulentes, autorisant une concentration foliaire élevée en Na⁺, probablement stocké dans des vacuoles. Son parent, *H. petiolaris* est apparu comme étant l'espèce la plus sensible au NaCl avec une survie moindre, une concentration foliaire en Na⁺ faible et une fraction racinaire très petite. *H. annuus*, est un peu moins sensible au NaCl appuyant l'hypothèse de sa capacité à exclure le Na⁺ hors de la cellule (Flowers *et al.* 1986 ; Ashraf et O'Leary 1995 ; Francois 1996) ou à le compartimenter dans les racines (Ballesteros *et al.* 1997). Une seconde étude (Lexer *et al.* 2003a et b), en milieu naturel, a consisté à transplanter des populations des trois espèces, ainsi qu'une population synthétique BC₂ issue d'un croisement entre *H. annuus* et *H. petiolaris*, dans le milieu naturel de *H. paradoxus* très riche en NaCl mais aussi en Ca²⁺. Seuls les individus *H. paradoxus* et certains de la population synthétique ont survécu à ces conditions drastiques. Il semble en milieu naturel nécessaire d'avoir un moyen d'exclusion du sodium et une bonne prise de calcium pour tolérer une concentration élevée en sel (Lexer *et al.* 2003b). Cette thèse a été appuyée par la co-localisation de QTL de survie en milieu naturel de *H. paradoxus* et de QTL de prise d'ions Ca²⁺, K⁺, Mg²⁺ et Na⁺ (Lexer *et al.* 2003a). Enfin, dans une troisième étude (Karrenberg *et al.* 2006 voir en annexe), que nous avons menée en serre, des plantes des trois espèces ont été soumises à quatre traitements différents. A l'inverse de Welch et Rieseberg

(2002), les traitements appliqués sont moins sévères pour garantir la survie des plantes, pour les trois espèces, les mélanges salins sont plus complexes et tiennent compte de l'importance des ions K^+ et Ca^{2+} dans la réponse au stress salin. Le premier traitement mime le milieu salin naturel de *H. paradoxus* ('field like'), le second correspond au premier appauvri en Ca^{2+} ('minCa') pour vérifier l'importance d'une concentration élevée en calcium dans la tolérance au sel chez *H. paradoxus*, le troisième est équivalent au premier enrichi en K^+ ('plusK') car les conditions salines causent parfois des déficiences en potassium chez certaines espèces, et le dernier est un traitement réduit en sels considéré comme « non salin » ('no salt'). En conditions salines, *H. paradoxus* se développe mieux. Sa biomasse, sa fraction racinaire et la succulence sont plus importantes et sont associées à des concentrations foliaires accrues en Na^+ et SO_4^{2-} et au contraire diminuées en K^+ , Ca^{2+} et Mg^+ (Karrenberg *et al.* 2006).

En complément de l'étude précédente, analysant la réponse anatomique et physiologique des trois espèces de tournesol au stress salin, nous avons cherché à étudier les bases génétiques de l'adaptation aux marais salins de *H. paradoxus*, bases génétiques potentiellement impliquées dans la spéciation chez ce complexe d'espèces. Pour cela, nous avons comparé l'expression de 10 gènes candidats entre les trois espèces, qui diffèrent pour leur tolérance au sel et leur stratégie de réponse au stress salin. Ces gènes sont co-localisés avec des QTL de survie, d'ions minéraux, ou d'autres caractères potentiellement importants dans l'adaptation à des conditions salines (Figure 3.4). Ils sont tous potentiellement impliqués dans l'une des deux voies de réponse au stress salin (Figure 3.2) : le maintien de l'homéostasie ionique (maintien du rapport K^+/Na^+ , système SOS pour l'exclusion du Na^+) et de l'ajustement osmotique (augmentation de l'expression des gènes codant les LEA). Nous avons analysé la corrélation entre l'expression de ces gènes candidats et les valeurs des caractères phénotypiques et les teneurs foliaires en ions minéraux mesurés par Karrenberg *et al.* (2006), pour étudier si les différences de réponses au stress salin aux niveaux anatomique et physiologique, pouvaient s'expliquer, entre autres, par des niveaux d'expression variables de ces gènes candidats.

Matériel et Méthodes

Populations

Cette étude fait partie d'une expérience de réponse au stress salin et de tolérance aux sels de l'espèce halophyte *Helianthus paradoxus* et de ses deux espèces parentales *H. annuus* et *H. petiolaris* sensibles aux sels (Karrenberg *et al.* 2006). Pour chaque espèce, trois

populations éloignées géographiquement ont été collectées ou obtenues auprès du National Germplasm Resources Laboratory (Beltsville, MD, USA). Les populations de *H. annuus* proviennent du Dakota du Sud, du Colorado et de l'Arizona, celles de *H. paradoxus* proviennent toutes du Nouveau Mexique, d'où l'espèce est originaire, mais de localités éloignées et celles de *H. petiolaris* du Montana, du Colorado et du Texas (voir Karrenberg *et al.* 2006 en annexe pour les références exactes).

Plan d'expérience

L'expérience a été réalisée en deux parties. La première partie portant sur 612 plantes, a permis d'analyser les variations physiologiques des neuf populations végétales soumises à différents traitements salins (Karrenberg *et al.* 2006). La deuxième partie réalisée parallèlement sur 108 plantes, a consisté à étudier les variations d'expression de 10 gènes candidats de réponse au stress salin, dans les neuf populations soumises aux différents traitements salins. Pour ces deux expériences les plantes des neuf populations ont été cultivées au même moment en serre à Bloomington, Indiana University. Elles ont subi les quatre mêmes traitements salins :

1- un traitement mimant le milieu salin dans lequel vit *H. paradoxus*, ou traitement « field like » (fl) riche en sodium, calcium, magnésium et sulfate.

2- un traitement « field like » avec une concentration en calcium et magnésium réduite, ou traitement « minCa ». Une concentration élevée en calcium semble favoriser la tolérance au sel chez *H. paradoxus* (Lexer *et al.* 2003a).

3- un traitement « field like » auquel a été ajouté du potassium, ou traitement « plusK ». Les conditions salines causent parfois des déficiences en potassium chez certaines espèces (Maathuis et Amtmann 1999).

4- un traitement non salin (ns).

Le dispositif expérimental de l'étude physiologique est décrit dans Karrenberg *et al.* 2006. Pour l'expérience sur les variations d'expression, les 108 plantes (3 espèces x 3 populations x 4 traitements x 3 répétitions) ont été regroupées sur 12 rangées de neuf pots, chaque rangée étant constituée d'un individu de chacune des neuf populations (Figure 3.3). Toutes les plantes d'une rangée ont été soumises à un traitement salin identique pour simplifier le protocole expérimental. Les populations intra-rangées ont été randomisées et le type de traitement appliqué à chacune des 12 rangées également, avec un total de trois rangées, équivalent à trois répétitions, pour chacun des quatre traitements. Les 108 plantes ont donc été soumises aux mêmes étapes de germination, et traitements salins que les 612 autres

plantes (voir Karrenberg *et al.* 2006). Le prélèvement final a cependant été pratiqué une semaine plus tard que celui de la première partie de l'expérience et par conséquent, les traitements salins ont été appliqués pendant une semaine supplémentaire. Pour chacune des 107 plantes encore vivantes à la fin de l'expérience des échantillons des feuilles les plus vieilles, mais toujours vertes, ont été prélevés et immédiatement immergés dans de l'azote liquide avant d'être conservés à -80°C.

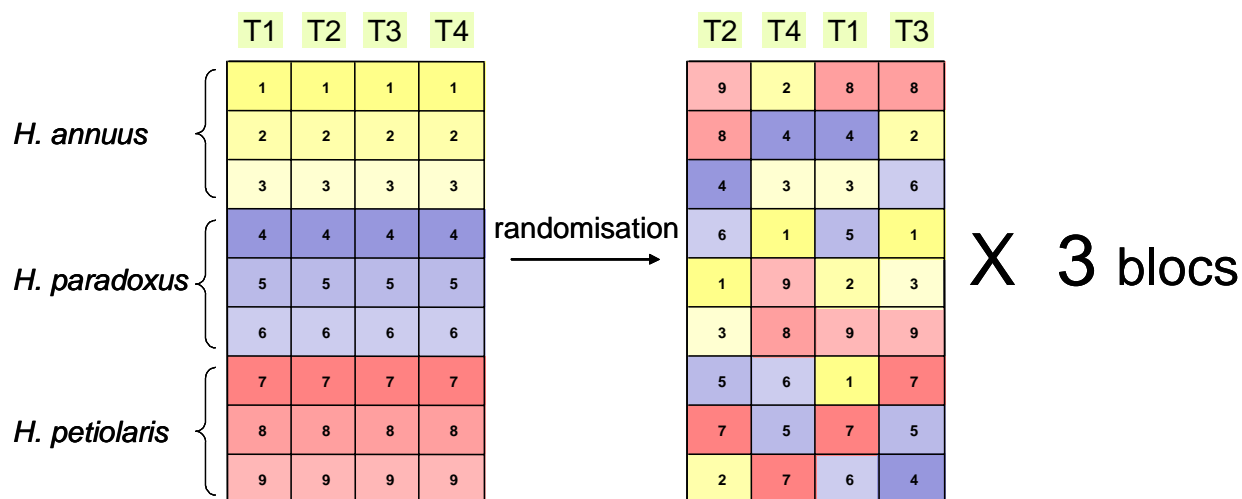


Figure 3.3 : Plan d'expérience comprenant 3 blocs de 4 traitements effectués sur 3 populations de 3 espèces.

Choix des gènes candidats

Plusieurs gènes potentiellement impliqués dans la réponse aux stress salins ont été cartographiés aux cours de deux études antérieures et certains sont co-localisés avec des QTL de survie ou d'ions minéraux (Figure 3.4).

1- Dans une première étude (Lexer *et al.* 2004) les gènes candidats impliqués dans le flux de calcium à travers les membranes, les voies de transduction activées par le Ca^{2+} ou plus généralement les stress ont été recherchés dans la base de données d'EST du Compositae Genome Project (<http://cgpdb.ucdavis.edu>). Sur 36 gènes candidats identifiés, 11 ont été cartographiés sur une population d'hybrides synthétiques BC_2 issus d'un croisement entre *H. annuus* et *H. petiolaris* et transplantés dans le milieu naturel de l'espèce hybride *H. paradoxus*. Parmi ces 11 gènes, des études de génétique d'association ont révélé que le polymorphisme de trois gènes (*CDPK3*, *ATP1* et *tREG*) était corrélé à la survie de la population hybride synthétique dans le milieu naturel de *H. paradoxus* (Lexer *et al.* 2004). Ces trois gènes peuvent donc être considérés comme des QTL de survie. Par ailleurs, le gène

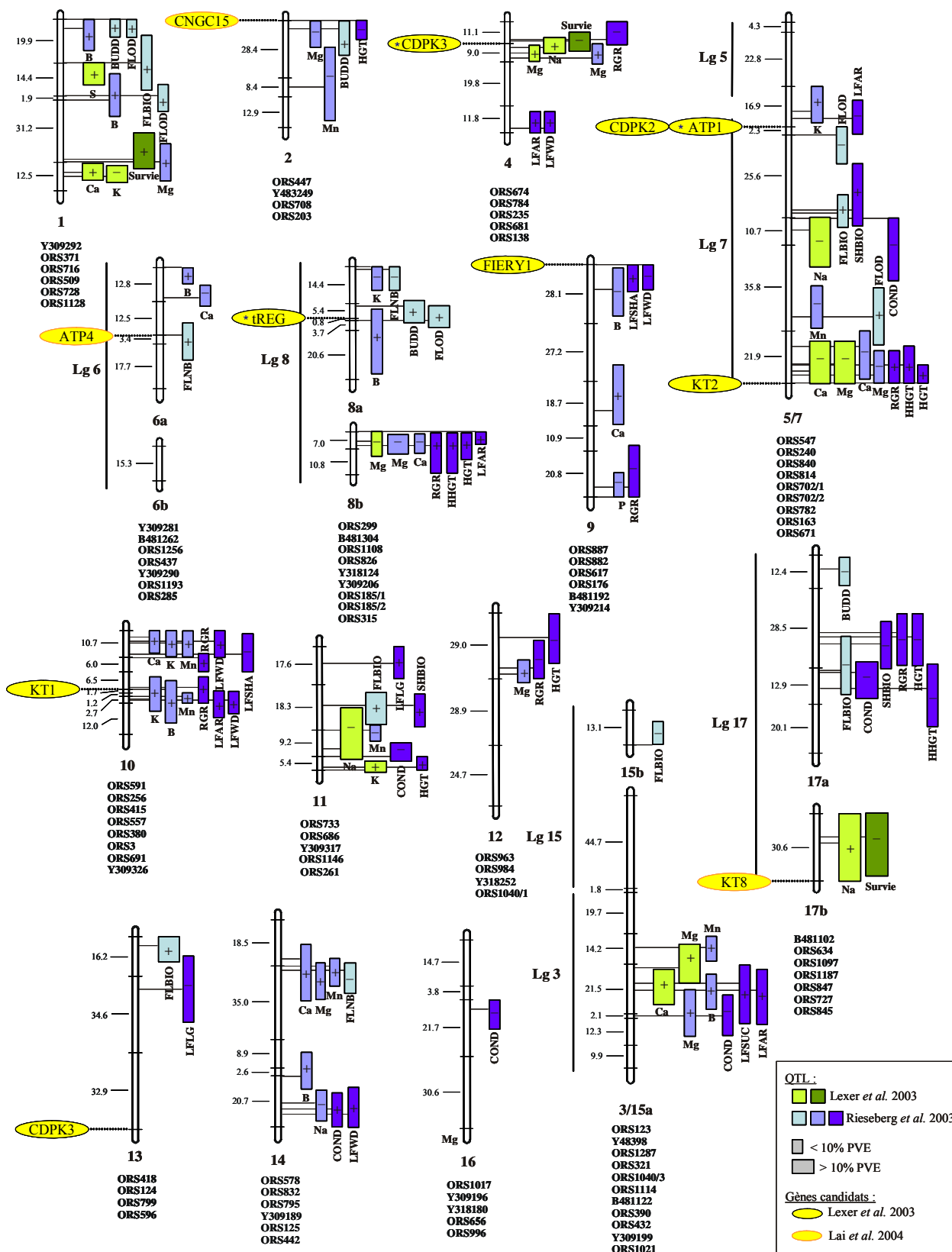
CDPK3 est co-localisé avec un QTL de survie et des QTL de teneur en ions minéraux Mg^{+} et Na^{+} préalablement identifiés à l'aide de marqueurs moléculaires chez les mêmes individus (Lexer *et al.* 2003b) (Figure 3.4). Les gènes *ATP1* et *tREG* sont cartographiés sur des groupes de liaisons pour lesquels aucun QTL de survie ou de QTL liés à la tolérance au sel n'avaient été localisés, les marqueurs au voisinage de ces deux gènes n'étaient pas associés à une meilleure survie (Lexer *et al.* 2003a). La faible densité de marqueurs utilisée lors de cette étude, dans ces deux régions, pourrait expliquer pourquoi aucun QTL d'adaptation au milieu naturel de *H. paradoxus* n'avait déjà été trouvé.

2- Dans une étude de cartographie plus récente (Lai *et al.* 2005), 243 EST contenus dans les banques de données tournesol de plus de 67 000 EST ont été localisés sur une carte obtenue pour une population de RIL de septième génération issue d'un croisement entre deux lignées de *H. annuus* RHA280 et RHA801 (Tang *et al.* 2002). Parmi ces 243 EST, 13 ont été co-localisés (1 à 10 cM) avec des QTL de caractères morphologiques et physiologiques détectés sur la population BC₂ transplantée dans le milieu naturel de *H. paradoxus* (Lexer *et al.* 2003a) ou cultivée en serre (Rieseberg *et al.* 2003), notamment des QTL de survie (Lexer *et al.* 2003a), de tolérance au sel (Lexer *et al.* 2003a, Rieseberg *et al.* 2003) et de domestication (Burke *et al.* 2002), et possèdent une fonction en accord avec le QTL. Six codent des canaux ou des protéines de transport pouvant réguler la prise d'ions minéraux et sont co-localisés avec des QTL de variation en teneur d'ions minéraux dans les feuilles.

Parmi ces 17 (11+6) gènes, nous avons retenus les 10 gènes candidats (Table 3.1, Figure 3.4) qui étaient amplifiables dans les trois espèces et dont le produit d'amplification était de taille identique pour les trois espèces. Cinq des 10 gènes (*ATP1*, *CDPK2*, *CDPK3*, *KT8* et *tREG*) ont été identifiés comme ayant une importance dans la survie en milieu salin et / ou sont co-localisés avec des QTL de survie dans le milieu de *H. paradoxus* (Figure 3.4). Les séquences des 10 gènes candidats ont été ré-analysées en octobre 2006 par BlastX de la séquence EST dans la base de données Swissprot ou sur l'ensemble des bases de données de protéines du NCBI quand aucun résultat n'était obtenu dans Swissprot (Table 3.1). L'étude de leur fonction putative montre que ces gènes codent des protéines impliquées dans les différentes voies de réponses à la salinité, la réponse au stress ionique et la réponse au stress osmotique (Figure 3.2).

Contig - amorce	Nvx noms	Taille EST (bp)	Taille (en aa)	% d'identité	Intitulé des séquences chez <i>Arabidopsis thaliana</i> (ARATH) ou <i>Medicago truncatula</i>	Fonction potentielle
QHB2d08 / <i>ATP1</i>	<i>ATP1</i>	703	234	86	ECA4_ARATH, ATPase4	Calcium-transporting ATPase, endoplasmic reticulum-type.
ctg1195A / <i>HT185</i>	<i>ATP4</i>	790	213	76	ECA4_ARATH, ATPase4	Calcium-transporting ATPase, endoplasmic reticulum-type.
ctg278 / <i>HT175</i>	<i>CNGC15</i>	670	184	86	CNG15_ARATH	Cyclic nucleotide-and calmodulin-regulated ion channel
ctg1024 / <i>HT215</i>	<i>KT8</i>	720	236	63	POT8_ARATH, AtHAK8	Potassium transporter
ctg3197	<i>KT2</i>	778	258	86	POT2_ARATH, AtKT2	Potassium transporter
QHG17C14 / <i>KTP1</i>	<i>KT1</i>	676	218	73	POT4_ARATH, AtKT4	Potassium transporter
ctg2180	<i>FIERY1</i>	515	49	73	DPNP1_ARATH, <i>FIERY1</i>	Inositol polyphosphate 1-phosphatase <i>FIERY1</i> protein
QHB15M08 / <i>CDPK2</i>	<i>CDPK2</i>	657	179	59	CDPK1_ARATH	Calcium-dependent protein kinase
QHF10M05 / <i>CDPK3</i>	<i>CDPK3</i>	735	100	63	CDPK1_ARATH	Calcium-dependent protein kinase
Octg2456	<i>tREG</i>	781	260	49	ABE87525.1,	XS and zinc finger; Transcriptional regulator

Table 3.1 : Les 10 gènes candidats et leur fonction supposée. Le nom des contigs tel qu'on le trouve dans le Compositae Genome Project (<http://cgpdb.ucdavis.edu>), et / ou le nom utilisé par Lexer *et al.* 2004 et Lai *et al.* 2005, ainsi que les nouveaux noms utilisés ici pour plus de commodité, sont indiqués. La taille en paires de bases indique la taille de l'EST blasté. La taille en acides aminés est le nombre d'acides aminés qu'il a été possible d'aligner avec la protéine la plus similaire trouvée dans la base Swissprot et le pourcentage d'identité, le pourcentage de similarité entre la protéine candidate et la protéine de la base retenue.



D'après *Lexer et al. 2003* et *Rieseberg et al. 2003*

Figure 3.4 : Carte génétique d'une population issue d'un croisement interspécifique de deuxième génération entre *H. annuus* et *H. petiolaris*. Les QTL en vert ont été détectés dans une population de BC₂ (172 individus) transplantés dans le milieu naturel de *H. paradoxus* (Lexer *et al.* 2003a), alors que ceux en bleu ont été mis en évidence dans une population BC₂ (384 individus) issus du même croisement et cultivés en serre (Rieseberg *et al.* 2003). La position des marqueurs est indiquée sur chaque groupe de liaison par des traits horizontaux avec les distances génétiques (cM) les séparant sur la gauche. Les noms des marqueurs commençant par ORS font référence à des microsatellites isolés chez *H. annuus* (Tang *et al.* 2002), les autres sont des marqueurs AFLP (données non publiées). La position des QTL est représentée par des lignes horizontales, leur magnitude et l'intervalle de LOD sont indiqués par des rectangles de largeur et hauteur variables. Le symbole +/- dans chaque rectangle donne la direction de l'effet de l'allèle de *H. annuus* dans la population de cartographie BC₂. Les gènes candidats sont positionnés au niveau du plus proche marqueur dans la région. Ils ont été cartographiés soit sur la carte BC₂ et sont représentés en ovales jaunes cerclés de noir (Lexer *et al.* 2003a), soit sur une carte de *H. annuus* et figurent ici en ovales jaunes cerclés d'orange (Lai *et al.* 2004). Les * indiquent que les gènes candidats ont été cartographiés dans des régions significativement importantes pour la survie dans le milieu de *H. paradoxus* (Lexer *et al.* 2003a).

Sont indiqués sur la carte :

- en vert pâle et bleu intermédiaire les caractères liés aux concentrations foliaires en ions (ppm) : **B** bore, **Ca** calcium, **K** potassium, **Mg** magnésium, **Mn** manganèse, **Na** sodium, **P** phosphates et **S** sulfates
- en vert foncé les QTL de survie de la population BC₂ transplantée dans le milieu naturel de *H. paradoxus*
- en bleu pâle les caractères liés à la floraison : **BUDD** pour 'days until budding' – le nombre de jours avant apparition de boutons floraux (jours), **FLBIO** pour 'flower biomass' – la biomasse florale (g), **FLNB** pour 'flower number' – nombre de fleurs et **FLOD** pour 'days until first floret' – nombre de jours avant la première fleur (jours)
- en bleu foncé les caractères concernant la masse et la croissance de la plante : **SHBIO** pour 'shoot biomass' – la biomasse de la partie aérienne de la plante (g), **HGT** pour 'height seven days after planting' – taille de la plante à sept jours (cm), **HHGT** pour 'height at final harvest' – taille à la fin de l'expérience (cm), **RGR** pour 'relative growth rate' – la croissance relative (cm/jour), **LFLG** pour 'leaf length' – la longueur de la feuille (mm), **LFWD** pour 'leaf width' – la largeur de la feuille (mm), **LFAR** pour 'leaf area' – la surface de la feuille (cm²), **LFSHA** pour 'leaf shape' – la forme de la feuille (longueur/largeur en mm/mm), **LFSUC** pour 'leaf succulence' – la succulence foliaire ((poids humide – poids sec)/la surface foliaire en mg/cm²), **COND** pour 'leaf stomatal conductance' – la conductance des stomates des feuilles (mmol/m².s).

Extraction d'ARN et quantification de l'ARNm des gènes candidats

Extraction de l'ARN total

L'extraction d'ARN total a été pratiquée à l'UMR de Génétique Végétale (Gif-sur-Yvette). Les feuilles prélevées ont été broyées dans de l'azote liquide à l'aide de billes en acier stériles et d'un agitateur à peinture par cycles de 30 secondes à 1 minute, afin d'obtenir environ 500 mg de poudre très fine pour chacun des 107 échantillons. L'extraction d'ARN a ensuite été réalisée sur ces poudres à l'aide d'un protocole au Trizol, chloroforme, isopropanol en tubes. Les ARN ont été précipités dans l'isopropanol avec du glycogène à -20°C. Après centrifugation, les culots ont été rincés par de l'éthanol à 75 %, l'éthanol a ensuite été éliminé et les culots séchés dans le speedvac. Les culots ont alors été resuspendus dans de l'eau stérile sans RNase (versol). La qualité de l'ARN extrait a été vérifiée sur gel d'agarose 1%.

Obtention des ADN complémentaires par transcription inverse

Les ADN complémentaires (ADNc) ont été obtenus par transcription inverse. La solution contenait : de la solution d'ARN extraits, des hexamères randomisés pdN6 (Amersham Pharmacia) à 25.6 μM servant d'amorces, un ARN synthétique pAW109 (Applied Biosystem) servant de témoin d'efficacité de la transcription inverse et dont la concentration est connue et de l'eau sans RNase (versol). Après incubation 5 minutes à 70°C, un volume égal a été ajouté avec comme concentration finale 1 mM de dNTP, 1 U de RNAsine (Amersham Pharmacia) pour protéger l'ARN, 10 U d'enzyme RT, du tampon RT 1X (80 μM de $(\text{NH}_4)_2\text{SO}_4$, 335 μM Tris HCl Ph 8.8, 0.5 % de tween 20) et de l'eau. La réaction de transcription inverse a été incubée à 25°C pendant 5 minutes, puis 42°C pendant une heure, avant d'être stoppée à 95°C pendant 5 minutes. Le produit de transcription inverse ainsi obtenu a été dilué 50 fois avec de l'eau sans RNase (versol) et stocké à -80°C.

Amplification par PCR semi-quantitative

L'actine et l'ubiquitine ont été choisies comme gènes de ménage pour calibrer la concentration d'ADNc sur l'ensemble des 107 individus. Une bonne corrélation existait entre les concentrations d'actine et d'ubiquitine. Cela a permis de quantifier à l'aide du logiciel Software ImageQuant TL v2005 (Amersham Bioscience) les concentrations de ces deux ADNc présents dans chacun des échantillons et de faire des dilutions afin d'homogénéiser ces concentrations sur l'ensemble des échantillons.

Pour qu'une PCR semi-quantitative soit analysable, il est nécessaire de se situer dans la partie exponentielle de la PCR et non au plateau, où la concentration en ADN est identique quelque soit la quantité d'ADNc présente au départ. De ce fait le nombre de cycles d'amplification peut varier d'un gène à un autre et il est important de pratiquer un gradient de cycles pour chaque gène. Ici 22 à 39 cycles ont été testés pour chaque gène (Table 3.2). En outre, la température d'hybridation des amorces peut varier selon leur séquence et un gradient de température a permis de trouver une température optimale d'hybridation identique de 55°C, pour l'ensemble des deux gènes de ménage et des 10 gènes candidats.

Le mélange réactionnel de PCR semi-quantitative, réalisé dans un volume total de 25 μL , contenait 5 μL du produit de transcription inverse dilué, 3 mM de MgCl_2 , 0.2 mM de chacun des dNTP, 0.5 μM de chacune des deux amorces pour le gène amplifié (Table 3.2), de la Taq polymérase produite au laboratoire, et du tampon RT 1X. Les cycles d'amplification ont été : 1 cycle d'initiation de 5 minutes à 94°C ; entre 22 et 33 cycles selon le gène considéré de 30 secondes à 94°C, 30 secondes à 55°C et 30 secondes à 72°C ; et un cycle d'extension final de 5 minutes à 72°C. Les produits ainsi obtenus ont migré sur gel d'agarose 2 % contenant 1 mg/mL de bromure d'ethidium pendant 45 minutes à 130 V.

Gènes candidats Contig - amorce	Nvx noms	Amorces		Nb cycles	Réf. Biblio.
		Forward	Reverse		
QHB2d08 / <i>ATP1</i>	<i>ATP1</i>	TAATCCGCCCGATAAATACA	TTAAACACTCGGTCTCCTGC	28	Lexer <i>et al.</i> 2004
ctg1195A / <i>HT185</i>	<i>ATP4</i>	GGTTCACTCATGCCTCATTC	TAGTGGAACGATCCCAAAAA	30	Lai <i>et al.</i> 2005
ctg278 / <i>HT175</i>	<i>CNGC15</i>	CTTCTCAAAGGACTCCCGTT	AAACAACATTTTCGTTACCG	33	Lai <i>et al.</i> 2005
ctg1024 / <i>HT215</i>	<i>KT8</i>	GATTGTATCGGTGCATAGCC	CTTTTCATCAAACACTCGAGCC	30	Lai <i>et al.</i> 2005
ctg3197	<i>KT2</i>	CCGATGTTAAGTCGGTGAAC	CGACCTGCCTTACGTCTTTA	30	Lexer <i>et al.</i> 2004
QHG17C14 / <i>KTP1</i>	<i>KT1</i>	CCCGTATCTAACAATGCACC	ATACTCGGAACTTGCCACTG	28	Lexer <i>et al.</i> 2004
ctg2180	<i>FIERY1</i>	ATCATTGTTACCAACCGGAA	AAAGCAGGAGTGTGTTTTGG	30	Lexer <i>et al.</i> 2004
QHB15M08 / <i>CDPK2</i>	<i>CDPK2</i>	CAACGATGTATCCAAGGAGG	CTCACACAACCTCCATGACCA	30	Lexer <i>et al.</i> 2004
QHF10M05 / <i>CDPK3</i>	<i>CDPK3</i>	ATGATGACAAAAGAGCAGCC	CTTGATACCCGTTCCATTTG	22	Lexer <i>et al.</i> 2004
Octg2456	<i>tREG</i>	TCGTGGTCAATGTGGTCA	CGAGCTAACCACGCGTAA	30	Lexer <i>et al.</i> 2004

Table 3.2 : Les 10 gènes candidats avec la référence bibliographique de l'étude préalable dans laquelle ils ont été identifiés. Les amorces utilisées pour les amplifier et le nombre de cycles d'amplification sont indiqués pour chacun d'entre eux.

Quantification de l'expression des gènes

La quantité d'ADNc amplifiée, reflétant la quantité d'ARNm présente dans l'échantillon pour chacun des 10 gènes, a été estimée en unité arbitraire à l'aide du logiciel ImageQuant TL v2005 (Amersham Bioscience). Ce niveau d'expression a ensuite été normalisé par la moyenne des valeurs d'expression de l'actine et de l'ubiquitine. Le nombre de cycles d'amplification variant d'un gène à un autre, les valeurs obtenues sont des rapports relatifs d'expression qui n'ont de sens que lorsqu'ils sont comparés entre individus pour un même gène.

Analyses statistiques des données

Dans notre plan d'expérience, les individus de chaque population soumis à un même traitement sont dans des blocs différents (Figure 3.3). L'existence d'une variation d'expression entre populations pour les 10 gènes candidats ne peut être testée qu'à condition qu'il n'y ait pas de différences significatives entre les trois répétitions. Nous avons donc testé sur nos données deux modèles linéaires de décomposition de la variation phénotypique (Figure 3.5) :

$$\text{Modèle 1 : } y_{ijkl} = \mu + T_i + S_j + T_i \times S_j + B_k + \varepsilon_{ijkl}$$

$$\text{Modèle 2 : } y_{ijkl} = \mu + T_i + S_j + T_i \times S_j + P_k[S_j] + \varepsilon_{ijkl}$$

y est le niveau d'expression du gène en unité arbitraire, normalisé par la moyenne du niveau d'expression de l'actine et de l'ubiquitine, $i = 1$ à 4 est le traitement salin appliqué (traitement fl, plusK, minCa ou ns), $j = 1$ à 3 est l'espèce (*H. annuus*, *H. paradoxus* ou *H. petiolaris*), $k = 1$ à 3 la population pour chacune des trois espèces et $l = 1$ à 3 les individus d'une même population soumis à un même traitement.

Dans le modèle 1 (Figure 3.5), nous cherchons à tester l'existence d'un effet bloc et chaque espèce S_j est représentée par un individu de chaque population l . Si l'effet bloc n'est pas significatif pour l'expression du gène candidat testé, alors le modèle 2 peut être testé. Dans le modèle 2, chaque espèce est représentée par trois populations k et chaque population par trois individus l , il s'agit donc d'un modèle hiérarchique.

Les graphiques représentant les résidus en fonction des valeurs prédites ont été tracés. Tous étaient homogènes ce qui nous plaçait dans les conditions d'applications des ANOVA. De même des tests de Kolmogorov-Smirnov indiquaient que les résidus étaient distribués selon une loi normale et ceci quelque soit le gène. Des comparaisons de moyennes par des tests de Student ont été effectuées quand l'effet du facteur était significatif ou proche de l'être ($P < 0.06$).

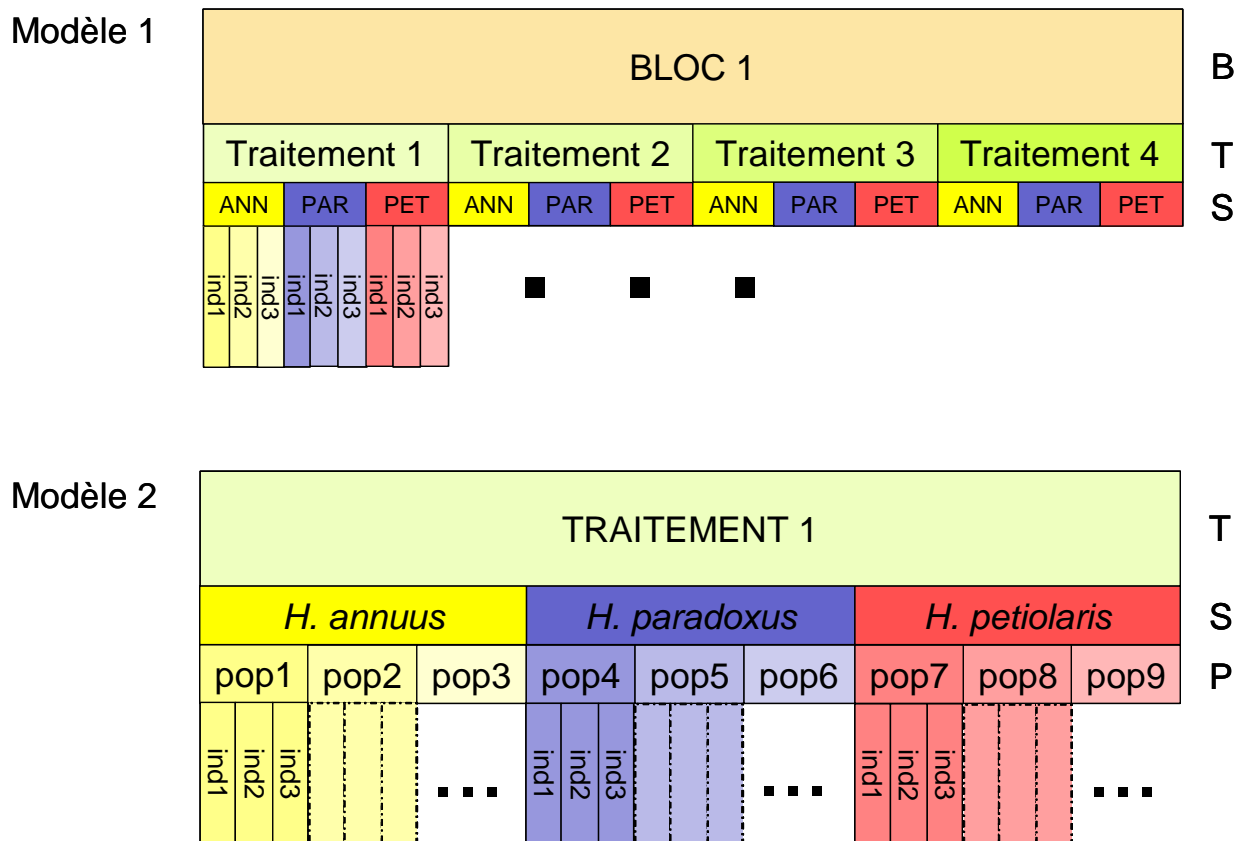


Figure 3.5 : Modèles linéaires mixtes testés. Modèle 1 : $y_{ijkl} = \mu + T_i + S_j + T_i \times S_j + P_k[S_j] + \varepsilon_{ijkl}$
 Modèle 2 : $y_{ijkl} = \mu + T_i + S_j + T_i \times S_j + I_k + \varepsilon_{ijkl}$.

Les deux volets de cette expérience ont été mis en relation, bien que nous n'ayons pas des données phénotypiques et d'expression génique pour les mêmes individus. Les individus sont issus des mêmes populations et ont été cultivés ensemble dans les mêmes conditions. Une analyse en composantes principales peut donc être réalisée sur les moyennes estimées par populations, pour les quatre traitements salins et les trois espèces *H. annuus*, *H. paradoxus* et *H. petiolaris*, à l'aide du modèle : $\hat{y}_{ijk} = \hat{\mu} + \hat{T}_i + \hat{S}_j + \hat{P}_{jk}$, sur les caractères suivants :

- la biomasse exprimée en grammes (Biomass), la fraction racinaire qui est le rapport de la masse des racines sur la masse totale de la plante (root fraction), le taux de croissance exprimé en centimètres par jour (HGR – pour High Gross Rate) et la succulence exprimée en milligramme d'eau par centimètre carré (succulence), pour lesquels la moyenne a été effectuée sur 15 individus par traitements et populations,

- les concentrations foliaires en calcium, sodium, potassium, magnésium et sulfates exprimées en millimoles par gramme de matière sèche, avec une moyenne sur 10 individus,

Gène	Effet	Df	Sum. Sq.	valeur F	P (>F)	
KT1	trtt	3	0.8279	2.7464	0.0483	*
	espece	2	2.7738	13.8031	< 0.0001	***
	trtt:espece	6	0.3644	0.6044	0.7260	
	pop [esp]	6	0.2415	0.4006	0.8766	
	Residus	80	8.0382			
tREG	trtt	3	1.2007	2.6055	0.0575	
	espece	2	3.4966	11.3814	< 0.0001	***
	trtt:espece	6	2.0395	2.2128	0.0501	
	pop [esp]	6	2.8861	3.1313	0.0083	**
	Residus	80	12.2890			
ATP4	trtt	3	0.1414	0.2120	0.8878	
	espece	2	32.5465	73.2038	< 0.0001	***
	trtt:espece	6	1.5536	1.1648	0.3335	
	pop [esp]	6	3.2985	2.4730	0.0303	*
	Residus	80	17.7841			
CDPK3	trtt	3	1.6165	1.4492	0.2348	
	espece	2	6.2925	8.4622	0.0005	***
	trtt:espece	6	1.4143	0.6340	0.7026	
	pop [esp]	6	11.6869	5.2388	0.0001	***
	Residus	79	29.3723			
KT8	trtt	3	4.1303	12.4135	< 0.0001	***
	espece	2	1.2465	5.6194	0.0052	**
	trtt:espece	6	0.5597	0.8411	0.5420	
	pop [esp]	6	0.6677	1.0034	0.4292	
	Residus	79	8.7618			
CNGC15	trtt	3	0.3015	0.6380	0.5927	
	espece	2	0.6889	2.1866	0.1190	
	trtt:espece	6	0.6418	0.6790	0.6670	
	pop [esp]	6	2.9585	3.1302	0.0083	**
	Residus	80	12.6019			
FIERY1	trtt	3	0.8895	1.2456	0.2988	
	espece	2	1.4386	3.0219	0.0543	
	trtt:espece	6	1.9182	1.3431	0.2480	
	pop [esp]	6	1.4543	1.0183	0.4195	
	Residus	80	19.0427			
KT2	trtt	3	1.7776	2.7362	0.0492	*
	espece	2	14.8783	34.3523	< 0.0001	***
	trtt:espece	6	0.6711	0.5165	0.7941	
	pop [esp]	6	4.7809	3.6795	0.0029	***
	Residus	77	16.6747			
ATP1	trtt	3	1.1165	2.2813	0.0856	
	espece	2	5.6709	17.3802	< 0.0001	***
	trtt:espece	6	1.8988	1.9398	0.0845	
	pop [esp]	6	0.8193	0.8370	0.5450	
	Residus	79	12.8882			
CDPK2	trtt	3	0.3606	1.2236	0.3066	
	espece	2	0.2393	1.2181	0.3012	
	trtt:espece	6	0.4212	0.7145	0.6390	
	pop [esp]	6	0.5261	0.8925	0.5046	
	Residus	80	7.8594			

Table 3.3 : Analyses de variance réalisées sur les variations d'expression de chacun des 10 gènes candidats avec le modèle $y_{ijkl} = \mu + S_i + T_j + R_k + S_i \times T_j + \varepsilon_{ijkl}$. La significativité des effets est indiquée par des astérisques (* $0.010 < P < 0.050$, ** $0.001 < P < 0.010$ et *** $P < 0.001$).

- la quantité relative d'ARNm des 10 gènes candidats normalisée par la moyenne des quantités d'actine et d'ubiquitine, avec une moyenne sur trois individus.

Les valeurs moyennes des caractères physiologiques (15 individus par population et traitement) et des concentrations foliaires (sur 10 des 15 individus précédents) proviennent des données de Karrenberg *et al.* 2006 et celles des quantités d'ARNm (trois individus par population et traitement) du deuxième volet de l'étude.

Résultats

Les effets blocs n'étant pas significatifs pour l'expression des gènes candidats, nous ne présentons ici que les résultats du modèle 2.

Variation d'expression en fonction des populations

Des différences significatives d'expression entre populations d'une même espèce sont observées pour cinq des 10 gènes candidats (Table 3.3, Table 3.4). Chez *H. paradoxus* les patrons d'expression diffèrent entre populations pour trois gènes (*ATP4*, *CDPK3*, *CNCG15*). Six couples de populations impliquent la population de Grants au Nouveau Mexique (Population 4 Table 3.4). Chez *H. annuus* la population du Dakota du Sud (population Ann1, Table 3.4) diffère des deux autres pour le gène *CDPK3*. Chez *H. petiolaris* c'est la population du Montana (population Pet7, Table 3.4) qui est la plus différente, elle diffère pour trois gènes (*tREG*, *CDPK3*, *KT2*) de l'une ou l'autre des deux autres populations.

<i>CDPK3</i>	t	P	<i>KT2</i>	t	P	<i>ATP4</i>	t	P
Ann 1-2	3,98	0,0001	Pet 7-9	4,073	0,0001	Par 4-6	2,1	0,0382
Ann 1-3	3,635	0,0005	Pet 8-9	2,67	0,0093	Par 5-6	3,593	0,0006
Par 4-5	2,39	0,0192						
Par 4-6	2,03	0,0457	<i>tREG</i>			<i>CNCG15</i>		
Pet 7-8	2,28	0,0253	Pet 7-8	2,048	0,0439	Par 4-5	3,21	0,0019
			Pet 7-9	3,566	0,0006	Par 4-6	3,26	0,0016

Table 3.4 : Résultats des contrastes indiquant les différences significatives d'expression moyenne entre populations pour chaque espèce. Les populations sont indiquées par des codes. Chez *H. annuus*, Ann1 = Dakota du Sud, Ann2 = Colorado et Ann3 = Arizona, chez *H. paradoxus*, Par4 = Grants, Par5 = Santa Rosa, Par6 = Bitter Lake toutes les trois au Nouveau Mexique et chez *H. petiolaris*, Pet7 = Montana, Pet8 = Colorado, Pet9 = Texas. Cinq des 10 gènes (*CDPK3*, *KT2*, *ATP4*, *tREG* et *CNCG15*) présentent des variations d'expression selon les populations.

Variation d'expression en fonction des espèces

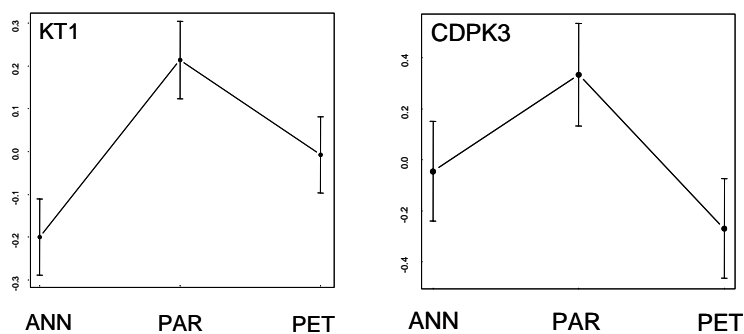
Le niveau d'expression de *tREG* ne suit pas les mêmes tendances chez l'espèce hybride et chez ses parents pour les quatre traitements ($F_{6, 0.05} = 2.2128$, $P = 0.0501$) : *tREG* est plus faiblement exprimé en milieu « non salin » chez les deux espèces parentales que chez *H. paradoxus* (résultats non montrés). Cependant aucun autre effet interaction entre les espèces et les traitements considérés n'est significatif, les variations d'expression d'une espèce à une autre sont donc globalement identiques, pour les quatre traitements, pour un gène donné.

Sur les 10 gènes candidats, cinq sont différemment exprimés chez l'espèce hybride halophyte et chez les deux espèces parentales sensibles à la salinité (Table 3.3, Figure 3.6 A, B, C). Deux sont surexprimés chez *H. paradoxus* par rapport à ses parents : *KT1* ($t_{\text{ANN-PAR}} = 5.346$, $P < 0.001$ et $t_{\text{PET-PAR}} = 2.856$, $P = 0.003$) et *CDPK3* ($t_{\text{ANN-PAR}} = 2.217$, $P = 0.015$ et $t_{\text{PET-PAR}} = 3.533$, $P < 0.001$). Les trois autres *KT2*, *tREG* et *ATP1* sont moins exprimés chez l'espèce hybride ($t_{\text{ANN-PAR}} = 5.883$, $P < 0.001$ et $t_{\text{PET-PAR}} = 7.46$, $P < 0.001$ pour *KT2*, $t_{\text{ANN-PAR}} = 4.121$, $P < 0.001$ et $t_{\text{PET-PAR}} = 2.777$, $P = 0.004$ pour *tREG*, et $t_{\text{ANN-PAR}} = 4.549$, $P < 0.001$ et $t_{\text{PET-PAR}} = 5.134$, $P < 0.001$ pour *ATP1*). Trois autres gènes sont exprimés de manière différente chez l'espèce parentale la plus sensible au sel : *H. petiolaris*. Ainsi le gène codant un facteur de signal de stress *FIERY1* est plus exprimé chez *H. petiolaris* ($t_{\text{ANN-PET}} = 1.706$, $P = 0.046$ et $t_{\text{PAR-PET}} = 2.252$, $P = 0.014$), même si l'effet espèce dans le modèle linéaire n'est pas significatif ($F_{2, 0.05} = 3.0192$, $P = 0.054$). Enfin deux gènes *ATP4* et *KT8* sont moins transcrits chez *H. petiolaris* ($t_{\text{ANN-PET}} = 10.159$, $P < 0.001$ et $t_{\text{PAR-PET}} = 10.180$, $P < 0.001$ pour *ATP4* et $t_{\text{ANN-PET}} = 3.237$, $P < 0.001$ et $t_{\text{PAR-PET}} = 2.629$, $P = 0.005$ pour *KT8*). L'expression des gènes *CNGC15* et *CDPK2* ne varie pas selon les espèces.

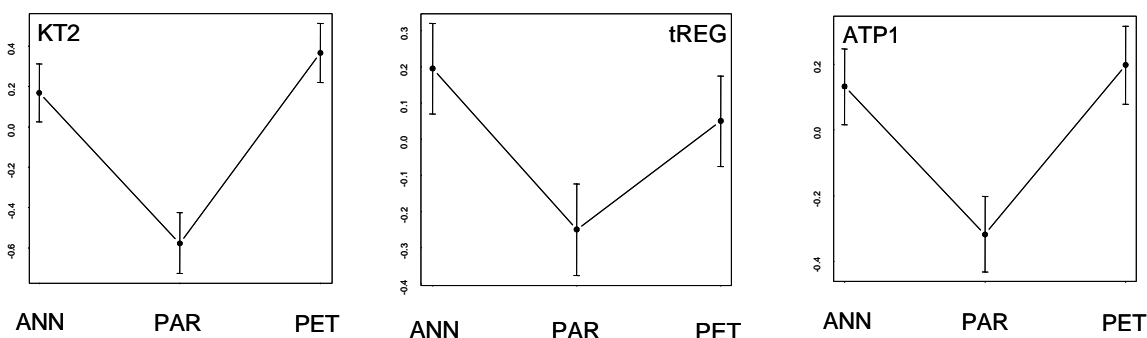
Variation d'expression en fonction des traitements

KT1 est différemment exprimé selon les milieux, significativement plus dans le milieu « field like » ($t_{\text{fl-minCa}} = 2.944$, $P = 0.003$ et $t_{\text{fl-ns}} = 2.545$, $P = 0.007$). À l'opposé, *KT2* est significativement moins fortement exprimé sous le traitement « field-like » ($t_{\text{fl-plusK}} = 1.932$, $P = 0.030$, $t_{\text{fl-minCa}} = 2.788$, $P = 0.004$ et $t_{\text{fl-ns}} = 2.114$, $P = 0.020$). Les transcrits *KT8* sont en moins grandes quantités chez les trois espèces quand les plantes ne sont pas soumises à un traitement salin ($t_{\text{ns-fl}} = 5.329$, $P < 0.001$, $t_{\text{ns-plusK}} = 5.348$, $P < 0.001$ et $t_{\text{ns-minCa}} = 3.559$, $P < 0.001$) (Figure 3.6 D).

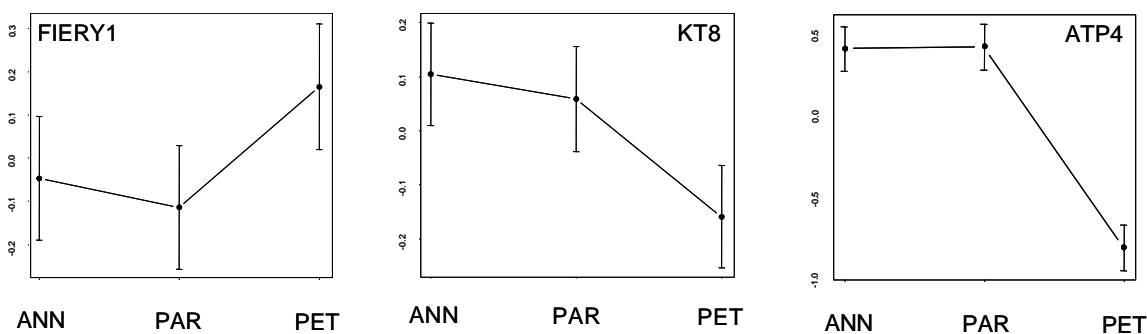
A- Deux gènes sont plus exprimés chez *H. paradoxus* par rapport à ses deux parents :



B- Trois gènes sont moins exprimés chez *H. paradoxus* par rapport à ses deux parents :



C- Trois gènes différemment exprimés chez *H. petiolaris* par rapport à *H. annuus* et *H. paradoxus* :



D- Différence d'expression selon les traitements des trois gènes codant des transporteurs à

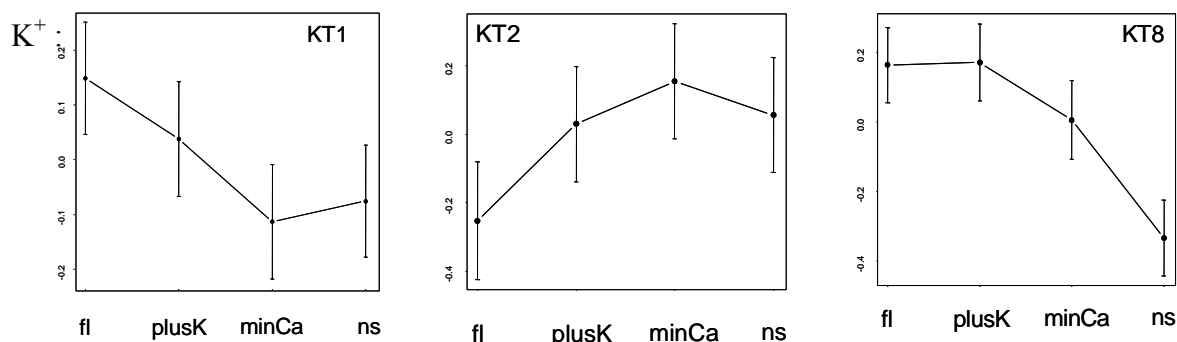


Figure 3.6 : Valeurs moyennes relatives d'expression et les erreurs types associées quand un effet espèce et parfois traitement du modèle $y_{ijkl} = \mu + T_i + S_j + T_i \times S_j + P_k[S_j] + \varepsilon_{ijkl}$ est trouvé significatif pour un gène candidat.

Relation entre les variations d'expression et les variations physiologiques

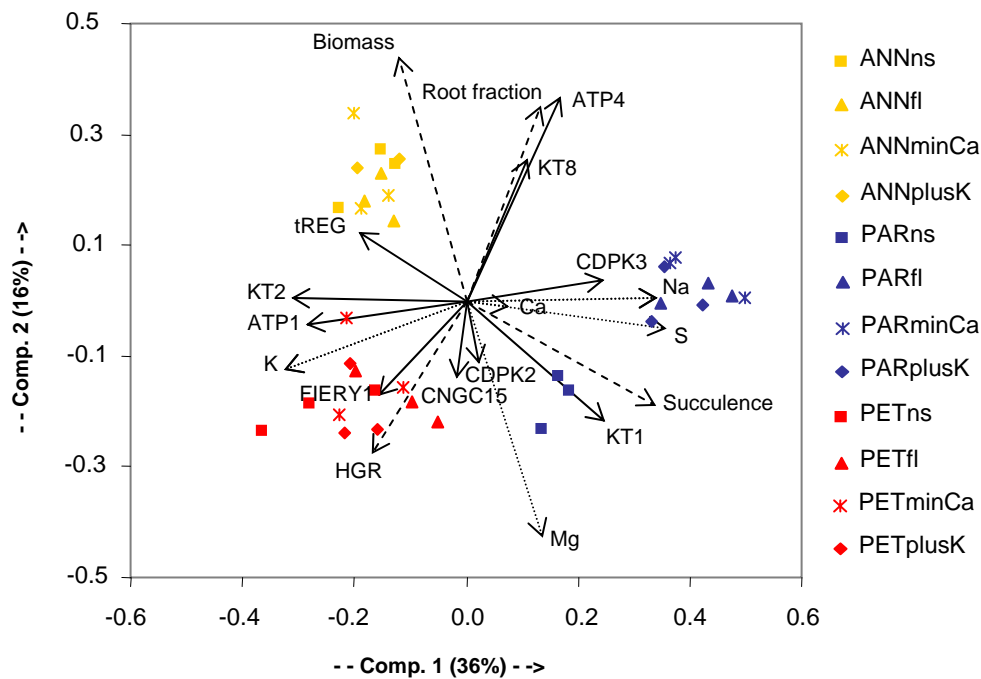
Une analyse en composantes principales (ACP) sur les variables des deux volets de l'expérience, physiologique et d'expression, a été réalisée pour visualiser la part de variation de chaque variable impliquée dans la différenciation entre espèces et traitements (Figure 3.7).

La première composante, qui explique 36% de la variation totale des variables étudiées, permet de séparer l'espèce hybride (figurés bleus) des deux espèces parentales (figurés rouges et jaunes). La deuxième composante, expliquant 16% de la variabilité, sépare les deux parents entre eux. Sur ces deux premières composantes, pour une espèce donnée tous les traitements sont confondus sauf pour *H. paradoxus* qui en milieu non salin (carrés bleus) se comporte comme *H. petiolaris* (figurés rouges) sur la deuxième composante. Sur la troisième composante, qui explique 10% de la variabilité, les figurés carrés des trois espèces correspondants aux individus soumis au traitement non salin sont projetés séparément des autres figurés, c'est-à-dire des individus soumis à un des trois traitements salins.

Les différences entre les espèces parentales et *H. paradoxus* s'expliquent par des caractères caractéristiques des plantes halophytes : une forte succulence associée à une augmentation des concentrations foliaires en Na^+ et en SO_4^{2-} et dans une moindre mesure en Ca^{2+} . L'augmentation de la concentration foliaire en Ca^{2+} est associée à l'augmentation de l'expression d'une kinase activée par le calcium, *CDPK3*, co-localisé avec un QTL de survie en milieu salin (Figure 3.4). Le transporteur de calcium *ATP1*, qui est supposé séquestrer le Ca^{2+} dans le réticulum endoplasmique, est lui plus faiblement exprimé. La faible concentration foliaire de K^+ est couplée à une faible expression relative par rapport aux parents du transporteur de potassium *KT2*. Enfin *tREG* qui est un régulateur transcriptionnel potentiel des gènes LEA impliqués dans la réponse au stress osmotique est sous exprimé chez *H. paradoxus*.

Les différences majeures entre *H. annuus* et l'espèce très sensible au sel, *H. petiolaris* s'expliquent aussi par des différences anatomiques et des variations liées aux teneurs en ions et à la synthèse d'osmoprotecteurs, mais les protéines mises en jeu pour répondre au stress salin diffèrent de celles mises en jeu par *H. paradoxus*. Les populations de *H. annuus*, tous traitements confondus, sont positivement associées à une importante biomasse et fraction racinaire et à une faible succulence, alors que *H. petiolaris* est associé à une faible fraction racinaire et à une forte croissance surtout pour les traitements salins. Chez *H. annuus* la concentration foliaire en magnésium est faible et l'expression de *KTI*, un transporteur de K^+ , est réduite. Par contre chez *H. petiolaris* la concentration foliaire en K^+ est importante et est

A- Représentation sur les deux premières composantes



B- Représentation sur les composantes 1 et 3

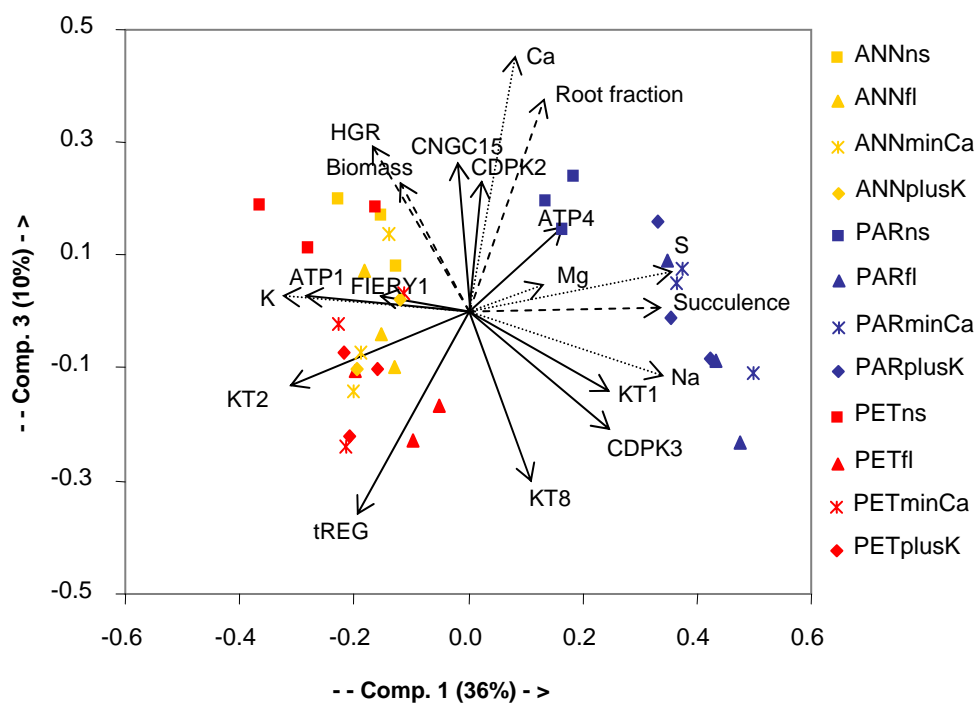


Figure 3.7 : Analyse en Composantes Principales des données phénotypiques (vecteurs pointillés) et d'expression génique (vecteurs pleins) de trois populations de chacune des trois espèces *H. annuus* (ANN en jaune), *H. paradoxus* (PAR en bleu) et *H. petiolaris* (PET en rouge) soumis à quatre traitements différents : carrés « sans sel », triangle « field-like », étoile « field-like avec moins de calcium » et losange « field-like avec un ajout de potassium ». Chaque figuré représente une population.

couplée à une faible expression du transporteur de potassium *KT8*. Le transporteur de calcium *ATP4* est faiblement exprimé, alors que *FIERY1* est très exprimé chez *H. petiolaris*.

Les différences entre traitements non salins et salins sont dues à une croissance élevée, une biomasse et une fraction racinaire importantes, une concentration de calcium importante, couplées à l'expression des gènes *CNGC15* et *CDPK2*.

Discussion

Les gènes candidats et leur expression chez les trois espèces Helianthus

La comparaison des séquences des 10 protéines obtenues à partir d'EST clonés chez *H. annuus*, avec les séquences protéiques d'*Arabidopsis thaliana* (sauf *tREG* dont les seuls homologues trouvés sont un gène de *Medicago truncatula* et un gène de *Zea mays*) a permis d'attribuer une fonction putative à toutes les protéines codées par ces gènes. Six de ces gènes appartiennent à des familles de transporteurs d'ions. C'est le cas de *KT2*, *KT4* et *KT8* qui appartiennent à la famille de transporteur de K^+ « *KUP/HAK/KT* » (Mäser *et al.* 2001), de *CNGC15* qui appartient à la famille des « cyclic nucleotide-gated channel » (Mäser *et al.* 2001) et de *ATP1* et *ATP4* qui font partie de la famille des transporteurs de Ca^{2+} dans le reticulum endoplasmique (Axelsen et Pallgren 1998). Pour ces transporteurs, l'ensemble de la séquence a pu être alignée et le pourcentage d'identité entre les séquences de *H. paradoxus* et celles de *A. thaliana* varie entre 63 à 86%. Il est parfois plus important entre orthologues qu'entre paralogues. C'est le cas de *CNGC15* dont le pourcentage d'identité avec la protéine de *A. thaliana* est de 86% (Table 3.1) alors que les protéines codées par la famille des *CNGC* chez *A. thaliana* ont des pourcentages d'identité entre elles compris entre 55% et 83% (Mäser *et al.* 2001). Pour la famille des protéines kinases *CDPK*, pour *FIERY1* et *tREG*, le pourcentage d'identité entre la séquence de *H. paradoxus* et celle de *A. thaliana* est moins élevé. Pour *CDPK3*, 100 acides aminés sur les 245 qui forment la protéine, soit 41% de la protéine, sont alignés. Pour *CDPK2*, 82% de la séquence est bien alignée mais le pourcentage d'acides aminés identiques n'est que de 59%. Pour *FIERY1*, seulement 29% de la séquence est alignée (Table 3.1). Enfin, la séquence prédite en acides aminés de *tREG* chez *H. annuus* n'a seulement que 49% d'identité avec celle de *M. truncatula*. Si *tREG* est associée à une protéine à doigts de zinc chez *M. truncatula*, sa fonction est inconnue, mais est proche d'un régulateur transcriptionnel chez *Z. mays*. L'association de *tREG* à une protéine Bzip-TF reste donc à confirmer.

Sur les 10 gènes candidats étudiés, cinq sont différentiellement exprimés entre *H. paradoxus* et ses parents et trois supplémentaires entre *H. petiolaris* et ses deux congénères. *H. petiolaris* est de loin l'espèce la plus sensible à la salinité (Welch et Rieseberg 2002 ; Lexer *et al.* 2003a et b) et celle qui, logiquement, a le patron d'expression le plus éloigné de *H. paradoxus*. De même dans notre expérience, *H. petiolaris* différait beaucoup, pour un certain nombre de traits d'histoire de vie des deux autres espèces. À la fin de l'expérience, *H. petiolaris* était à un stade de développement plus avancé que ses congénères : certaines plantes, principalement sous stress salin, avaient atteint le stade de floraison. Cette précocité relative est probablement une stratégie pour assurer sa descendance. Plusieurs études expérimentales montrent en effet qu'il existe des tradeoffs entre traits d'histoire de vie quand les ressources sont limitantes (revue dans Stearns 1992 ; Reznick *et al.* 2000 ; Sgrö et Hoffmann 2004). Les ressources sont allouées à une fonction plutôt qu'une autre, comme ici où elles sont allouées à la reproduction plutôt qu'à la croissance.

Une réponse génétique au stress salin

En comparant les caractères physiologiques et l'expression des gènes candidats, il est possible de discuter le rôle de chacun des gènes et donc d'avoir une idée sur la réponse génétique au stress salin dans notre système d'étude.

L'adaptation au milieu salin de *H. paradoxus* est associée à une absorption plus importante en Na^+ et SO_4^{2-} et donc à une tolérance plus forte au Na^+ . L'augmentation de Na^+ dans les tissus est associée à une baisse de K^+ (Karrenberg *et al.* 2006 et Figure 3.7) caractéristique des plantes halophytes. Pour autant cela ne signifie pas qu'il y a modification de la balance K^+/Na^+ intracellulaire, les concentrations foliaires ne rendant pas compte de la compartimentation ionique intracellulaire. L'expression plus élevée de *KT1* chez *H. paradoxus* que chez ses parents et *KT8* chez l'espèce hybride et *H. annuus* que chez *H. petiolaris*, tous deux transporteurs à K^+ , suggère que *H. paradoxus* et dans une moindre mesure *H. annuus* utilisent bien ce type de transporteurs pour contrecarrer l'afflux trop important de Na^+ dans la cellule et ainsi maintenir un rapport K^+/Na^+ intracellulaire raisonnable. Les variations d'expression de *KT8*, localisé sur la carte BC₂ dans une région (LG17) où se situe un QTL de survie et un QTL d'ion Na^+ , confirme son rôle potentiel dans l'adaptation de *H. paradoxus* aux marais salins.

Comme nous l'avons vu, le Ca^{2+} joue un rôle primordial dans la séquestration et l'exclusion du Na^+ ainsi que dans la synthèse d'osmoprotecteurs comme les LEA. Il doit donc resté libre dans le cytoplasme pour assurer son rôle. Le transporteur de Ca^{2+} *ATP1*, impliqué

dans la séquestration du Ca^{2+} à l'intérieur du réticulum endoplasmique, et donc dans la balance K^+/Na^+ via le système SOS, est négativement corrélé à la quantité de Ca^{2+} (Figure 3.7) et est logiquement moins exprimé chez *H. paradoxus* que chez ses parents, évitant ainsi la séquestration du Ca^{2+} chez l'espèce hybride. *ATP1* se situe à proximité d'un QTL d'ion K^+ , d'un QTL de date de floraison et d'aire foliaire et a été localisé dans une région impliquée dans la survie de l'espèce hybride (Lexer *et al.* 2004). Les résultats d'expression confirment qu'*ATP1* est un candidat intéressant impliqué dans la réponse au stress salin et la survie de *H. paradoxus*. De même les variations d'expression de *FIERY1* peuvent s'expliquer par son rôle dans la signalisation calcique conduisant à une induction des réponses ioniques et osmotiques. *FIERY1*, une inositol polyphosphate phosphatase dégradant l'IP3 et freinant du même coup la libération de Ca^{2+} intracellulaire est plus exprimée chez *H. petiolaris*, la défavorisant probablement dans sa réponse au stress salin.

La réponse au stress salin chez *H. paradoxus* passe également, comme attendu, par la stimulation de la synthèse d'osmoprotecteur. Le gène *CDPK3*, induit par le Ca^{2+} et qui appartient à une famille de 34 membres intervenant dans les stress abiotiques (hydrique, thermique ou salin) chez *A. thaliana* (Harmon *et al.* 2001), est plus exprimé chez l'espèce hybride. Ces variations d'expression confirment le rôle de *CDPK3* dans l'adaptation à un environnement salin et probablement dans la spéciation de *H. paradoxus*. *CDPK3* est, en effet, co-localisé à un QTL de survie en milieu naturel salin, ainsi qu'à un QTL de croissance et à des QTL d'ions minéraux Mg^{2+} et Na^+ . *CDPK3* est le seul gène dont l'expression varie selon les populations chez les trois espèces étudiées de *Helianthus*. Il serait intéressant d'analyser les caractéristiques des milieux dans lesquelles les populations se situent et tester les corrélations entre variations d'expression et teneurs en ions minéraux.

Les profils d'expression des gènes *KT2*, *ATP4* et *tREG* sont plus surprenants. A l'inverse de *KT1* et *KT8*, *KT2* qui code également un transporteur à K^+ est sous exprimé chez *H. paradoxus* par rapport à ses parents. En outre, *KT2* est significativement moins exprimé dans le traitement « fl » et « plus K » (Figure 3.6), réagissant négativement à la présence de K^+ , ce qui n'est pas le cas de *KT1* et *KT8*, plus exprimés dans les traitements « fl » et « plus K ». Aucune donnée présente dans les bases de données et la littérature, ne nous permet d'expliquer les différences d'expression observées pour ces trois gènes codant des protéines dont la fonction putative est identique. Il est possible que ces canaux K^+ présentent des différences d'efficacité, certains étant plus ou moins perméables au Na^+ . De plus, la famille des gènes KT est très importante et il est possible que certains de ces gènes soient spécifiques de certains tissus, de certaines cellules ou même de certaines membranes (Kim *et al.* 1998 et

Mäser *et al.* 2001). Des échantillons de racines de nos plantes ont également été prélevés, mais nous n'avons pas réussi à en extraire l'ARN, ce qui ne nous a pas permis de comparer les différences d'expression entre tissus. Bien qu'*ATP4* ait une fonction putative identique à celle *ATP1*, son patron d'expression n'est pas identique. En effet, *ATP4* est sous exprimé chez *H. petiolaris* par rapport à *H. annuus* et *H. paradoxus*. Les voies de réponses privilégiées d'une espèce à l'autre ne sont peut-être pas les mêmes, ou les fonctions des gènes sont peut être plus complexes. Quant à *tREG*, il est localisé dans une région importante pour la survie d'individus d'une population BC₂ issue d'un croisement entre *H. annuus* et *H. petiolaris*, transplantée dans le milieu naturel salin de *H. paradoxus* (Lexer *et al.* 2004). Notre expérience confirme son rôle (composante 1, Figure 3.7), mais pas sa fonction. En effet si *tREG* est un facteur transcriptionnel induisant les protéines LEA, on s'attendrait à ce qu'il soit surexprimé chez *H. paradoxus* par rapport à ses parents et non sous exprimé comme c'est le cas. Notons que l'expression de ces trois gènes, *KT2*, *tREG* et *ATP4* varie selon les populations de *H. petiolaris* (*KT2*, *tREG*) ou de *H. paradoxus* (*ATP4*). Pour ces trois gènes, la sélection n'a pas homogénéisé les profils d'expression au sein de l'espèce.

Une réponse plastique au stress salin

10% de la variation sépare le traitement « ns » des trois traitements salins (composante 3, Figure 3.7). En milieu salin – les trois traitements confondus – la réponse générale des trois espèces est une réduction de la fraction racinaire, une faible concentration en Ca²⁺ associées à une sous expression de *CDPK2*, régulateur des gènes LEA et *CNGC15*. Le Ca²⁺ a donc un rôle central dans cette réponse plastique. Les *CDPK* interviennent dans différents stress abiotiques, tels que le stress thermique, salin et hydrique (Saijo *et al.* 2000), et certaines pourraient être spécialisées dans un stress en particulier (Urao *et al.* 1994, Xiong *et al.* 2002). *CDPK2* est co-localisé avec *ATP1*, qui semble avoir une importance dans la composante génétique de réponse au stress salin et *CNGC15* est proche de QTL d'ion Mg²⁺, de croissance et de floraison.

H. paradoxus en milieu « ns » se comporte en partie comme *H. petiolaris* (composante 2, Figure 3.7), alors qu'aucune différence n'est à noter entre les parents. L'adaptation aux marais salins de *H. paradoxus* est peut-être due à une plasticité plus importante chez l'espèce hybride que chez ses parents. On peut également imaginer que *H. paradoxus*, qui est une espèce jeune comparativement à ses parents, n'a pas encore eu le temps d'être « canalisée ». Ainsi même si *H. paradoxus* semble être une halophyte vraie, c'est-à-dire que sa production de biomasse est stimulée par la présence de sel, elle est tout à fait capable de s'adapter à un

milieu ne contenant que peu de sels, mais n'aurait peut être pas pu s'établir en compétition avec les deux espèces parentales.

Ainsi, chez *H. paradoxus* la réponse au stress salin est associée à des caractères propres aux plantes halophytes : succulence importante, concentrations foliaires élevées en Na^+ et SO_4^{2-} et plus faible en K^+ . *H. paradoxus* d'après notre expérience est plus plastique que les deux espèces parentales. Plusieurs gènes candidats, particulièrement *ATP1*, *CDPK3* et *KT8*, qui semblent être impliqués, ou localisés dans des régions impliquées, dans la fitness de *H. paradoxus* (Lexer *et al.* 2003a, 2004 ; Lai *et al.* 2005), ont des patrons d'expression différents chez *H. paradoxus* et les deux espèces parentales. Ils apparaissent donc comme de bons gènes candidats potentiels de la spéciation, et cette spéciation par adaptation à un environnement extrême ne serait pas le résultat de la mise en place d'une unique voie de réponse à la salinité, mais d'une réponse plus globale impliquant plusieurs voies. L'ensemble des gènes candidats est en cours de séquençage sur les individus de cette étude dans le laboratoire de Loren Rieseberg. Nous espérons pouvoir associer les différences d'expression observées à des mutations précises.

Références bibliographiques :

- Ashraf M, O'Leary JW (1995) Distribution of cations in leaves of salt-tolerant and salt-sensitive lines of sunflower under saline conditions. *Journal of Plant Nutrition*, **18**.
- Axelsen KB, Palmgren MG (1998) Evolution of substrate specificities in the P-type ATPase superfamily. *Journal of Molecular Evolution*, **46**, 84-101.
- Ballesteros E, Blumwald E, Donaire JP, Belver A (1997) Na⁺/H⁺ antiport activity in tonoplast vesicles isolated from sunflower roots induced by NaCl stress. *Physiologia Plantarum*, **99**, 328-334.
- Bohnert HJ, Ayoubi P, Borchert C, *et al.* (2001) A genomics approach towards salt stress tolerance. *Plant Physiology and Biochemistry*, **39**, 295-311.
- Burke JM, Tang S, Knapp SJ, Rieseberg LH (2002) Genetic analysis of sunflower domestication. *Genetics*, **161**, 1257-1267.
- Chinnusamy I, Jagendorf A, Zhu JK (2005) Understanding and improving salt tolerance in plants. *Crop Science*, **45**, 437-448.
- Flowers TJ, Hajibagheri MA, Clipson NJW (1986) Halophytes. *Quarterly Review of Biology*, **61**, 313-337.
- Flowers TJ, Koyama ML, Flowers SA, *et al.* (2000) QTL: their place in engineering tolerance of rice to salinity. *Journal of Experimental Botany*, **51**, 99-106.
- Francois LE (1996) Salinity effects on four sunflower hybrids. *Agronomy Journal*, **88**, 215-219.
- Harmon AC, Gribskov M, Gubrium E, Harper JF (2001) The CDPK superfamily of protein kinase. *New Phytologist*, **151**, 175-183.
- Karrenberg S, Edelist C, Lexer C, Rieseberg L (2006) Response to salinity in the homoploid hybrid species *Helianthus paradoxus* and its progenitors *H. annuus* and *H. petiolaris*. *New Phytologist*, **170**, 615-629.
- Kim EJ, Kwak JM, Uozumi N, Schroeder JI (1998) *AtKUP1*: an *Arabidopsis* gene encoding high-affinity potassium transport activity. *Plant Cell*, **10**, 51-62.
- Koyama ML, Levesley A, Koebner RM, Flowers TJ, Yeo AR (2001) Quantitative trait loci for component physiological traits determining salt tolerance in rice. *Plant Physiology*, **125**, 406-422.
- Lai Z, Livingstone K, Zou Y, *et al.* (2005) Identification and mapping of SNPs from ESTs in sunflower. *Theoretical and Applied Genetics*, **111**, 1532-1544.
- Lexer C, Lai Z, Rieseberg LH (2004) Candidate gene polymorphisms associated with salt tolerance in wild sunflower hybrids: implications for the origin of *Helianthus paradoxus*, a diploid hybrid species. *New Phytologist*, **161**, 225-233.
- Lexer C, Welch ME, Durphy JL, Rieseberg LH (2003a) Natural selection for salt tolerance quantitative trait loci (QTLs) in wild sunflower hybrids: implications for the origin of *Helianthus paradoxus*, a diploid hybrid species. *Molecular Ecology*, **12**, 1225-1235.
- Lexer C, Welch ME, Raymond O, Rieseberg LH (2003b) The origin of ecological divergence in *Helianthus paradoxus* (Asteraceae): selection on transgressive characters in a novel hybrid habitat. *Evolution*, **57**, 1989-2000.
- Maathuis FJM, Amtmann A (1999) K⁺ nutrition and Na⁺ toxicity : the basis of cellular K⁺/Na⁺ ratios. *Annals of Botany*, **84**.
- Mäser P, Thomine S, Schroeder JI, *et al.* (2001) Phylogenetic relationships within cation transporter families of *Arabidopsis*. *Plant Physiology*, **126**, 1646-1667.
- Reznick D, Nunney L, Tessier A (2000) Big houses, big cars, superfleas and the costs of reproduction. *Trends in Ecology and Evolution*, **15**, 421-425.
- Rieseberg LH, Raymond O, Rosenthal DM, *et al.* (2003) Major ecological transitions in wild sunflowers facilitated by hybridization. *Science*, **301**, 1211-1216.

- Sahi C, Agarwal M, Reddy MK, Sopory SK, Grover A (2003) Isolation and expression analysis of salt stress-associated ESTs from contrasting rice cultivars using a PCR-based subtraction method. *Theoretical and Applied Genetics*, **106**, 620-628.
- Sahi C, Singh A, Kumar K, Blumwald E, Grover A (2006) Salt stress response in rice: genetics, molecular biology, and comparative genomics. *Functional and Integrative Genomics*, **6**, 263-284.
- Saijo Y, Hata S, Kyojuka J, Shimamoto K, Izui K (2000) Over-expression of a single Ca²⁺-dependent protein kinase confers both cold and salt/drought tolerance on rice plants. *Plant Journal*, **23**, 319-327.
- Sgrö CM, Hoffmann AA (2004) Genetic correlations, tradeoffs and environmental variation. *Heredity*, **93**, 241-248.
- Shiozaki N, Yamada M, Yoshida Y (2005) Analysis of salt-stress-inducible ESTs isolated by PCR-subtraction in salt-tolerant rice. *Theoretical and Applied Genetics*, **110**, 1177-1186.
- Stearns SC (1992) *The evolution of life histories*.
- Taji T, Seki M, Satou M, *et al.* (2004) Comparative genomics in salt tolerance between *Arabidopsis* and *Arabidopsis*-related halophyte salt cress using *Arabidopsis* microarray. *Plant Physiology*, **135**, 1697-1709.
- Tang S, Yu JK, Slabaugh B, Shintani K, Knapp J (2002) Simple sequence repeat map of the sunflower genome. *Theoretical and Applied Genetics*, **105**, 1124-1136.
- Urao T, Katagiri T, Mizoguchi T, *et al.* (1994) Two genes that encode Ca²⁺-dependent protein kinases are induced by drought and high-salt stresses in *Arabidopsis thaliana*. *Molecular and General Genetics*, **244**, 331-340.
- Welch ME, Rieseberg LH (2002) Habitat divergence between a homoploid hybrid sunflower species, *Helianthus paradoxus* (Asteraceae), and its progenitors. *American Journal of Botany*, **89**, 472-478.
- Xiong L, Lee H, Ishitani M, Zhu JK (2002) Regulation of osmotic stress-responsive gene expression by the *LOS6/ABAI* locus in *Arabidopsis*. *Journal of Biological Chemistry*, **277**, 8588-8596.

DISCUSSION ET PERSPECTIVES

Discussion et perspectives

Nos études ont permis de confirmer le rôle dans l'adaptation de *H. paradoxus* à son environnement salin, de trois régions géniques et de l'une plus particulièrement. Ces régions ont été détectées au départ dans une population BC₂ d'hybrides synthétiques. Tout d'abord nous confirmons qu'il est possible de déceler une baisse de diversité au niveau de marqueurs neutres du fait du balayage sélectif intervenant dans une région soumise à sélection. Ce résultat est cependant loin d'être généralisable et peut-être dû à la structure et à la démographie particulière de notre espèce. Les populations de *H. paradoxus* sont majoritairement isolées, avec probablement une petite taille efficace. Toutes ces caractéristiques font de *H. paradoxus* un modèle idéal pour détecter des signatures de la sélection. Récemment Cano *et al.* (2006) ont étudié la différence de diversité entre microsatellites localisés dans des régions putativement neutres et un microsatellite lié au gène *EDA* (Ectodyplasin), un QTL majeur impliqué dans le nombre de plaques latérales osseuses défensives dans des populations d'épinoches, *Gasterosteus aculeatus*, de mer, de rivière et de lacs en Europe ayant divergé après la dernière glaciation il y a environ 10 000 ans. Les épinoches ayant colonisé les lacs ont des plaques latérales moindres du fait du risque plus faible de prédation. Ils n'ont détecté aucune déviation à la neutralité sur l'ensemble des marqueurs, ni de balayage sélectif au niveau du microsatellite lié au QTL en utilisant l'indice $\ln RH$ et concluent qu'utiliser des marqueurs liés à des QTL majeurs n'est pas une stratégie performante pour détecter des balayages sélectifs dans des populations naturelles. Contrairement à notre étude, la comparaison se fait à l'échelle de populations et non d'espèces, mais surtout les populations d'épinoches sont probablement moins isolées et plus mobiles, avec des tailles de populations efficaces plus conséquentes.

Les régions où les QTL de survie ont été détectées, présentent un fort déséquilibre de liaison avec certains marqueurs cartographiés dans la même région, ce qui appuie la thèse selon laquelle des balayages sélectifs peuvent s'étendre sur de grandes distances. La cartographie par DL développée ici, semble être un outil intéressant, particulièrement bien adapté aux régions sous sélection. Le long des groupes de liaison, des groupes de déséquilibre de liaison existent et pour optimiser les chances d'ordonner un maximum de marqueurs il est nécessaire que les distances ne soient pas trop importantes entre marqueurs.

Les questions qui restent donc maintenant à élucider sont :

1- Existe-t-il une règle permettant de prédire l'étendue des zones balayées ? Si oui est-elle fonction du régime de reproduction ? De la taille efficace de la population ?

La taille efficace dans les populations naturelles peut être estimée grâce à la distribution temporelle des variations de fréquences alléliques (Goldringer et Bataillon 2004). Il faut pour cela des données échantillonnées sur plusieurs années, ce que nous n'avons malheureusement pas.

2- Les balayages sélectifs sont étudiés soit à l'échelle intra-génique, soit à l'échelle de la cartographie génétique (centiMorgans). Est-ce qu'il existe des signatures communes et simultanées à ces deux échelles ?

Enfin, nous avons montré l'intérêt des approches de types « analyses d'expression » pour confirmer le rôle de gènes candidats. Le choix des gènes était bien ciblé, puisque l'on observe des différences d'expression entre espèces pour huit des 10 gènes candidats. Lai *et al.* (2006) dans une étude de microarrays sur plus de 3 000 EST entre *H. deserticola* et ses deux espèces parentales *H. annuus* et *H. petiolaris*, ont mis en évidence un pourcentage limité, voisin de 13%, de gènes s'exprimant différenciellement chez *H. deserticola*. Ils ont noté que la catégorie des transporteurs était sur-représentée dans les gènes différenciellement exprimés, renforçant la pertinence de nos résultats. Dans notre étude cinq des 10 gènes candidats sont des transporteurs et sont tous différenciellement exprimés entre les trois espèces confirmant le rôle de ce type de gènes dans l'hybridation homoploïde du groupe *Helianthus*. On retrouve ces transporteurs dans différentes voies de réponses au stress. La spéciation de *H. paradoxus* et son adaptation à des marais salins impliquent de nombreux gènes et leur sous- ou sur-expression par rapport aux espèces parentales concordent avec l'hypothèse de ségrégation transgressive mise en place lors de l'hybridation homoploïde.

Les trois approches développées ici, confirment l'importance d'une région sous sélection chez *H. paradoxus* au niveau du groupe de liaison 4 : une moindre diversité est observée au niveau de marqueurs microsatellites liés au QTL de survie. L'un de ces microsatellites (ORS784) est en fort déséquilibre de liaison avec d'autres marqueurs microsatellites dans la région et l'ordre le plus vraisemblable pour ces marqueurs est cohérent avec les données de cartes génétiques déjà existantes. En outre le gène CDPK3, intervenant dans la réponse au stress osmotique, est co-localisé avec ORS784, un gène plus exprimé chez l'espèce hybride que chez les espèces parentales et qui semble particulièrement adapté dans la fitness de *H. paradoxus*. Il sera intéressant de voir s'il existe des différences de séquence au

niveau de ce gène entre espèces pour mieux comprendre son importance et sa fonction qui est encore floue.

Il est intéressant de noter que les deux ACP, l'une effectuée sur des marqueurs neutres, des microsatellites (chapitre 1), et l'autre sur des caractères physiologiques, concentrations ioniques et expressions géniques (chapitre 3), conduisent à la même séparation des trois espèces *H. annuus*, *H. petiolaris* et *H. paradoxus*. Les 12 marqueurs, utilisés pour construire la première ACP, sont pour moitié localisés dans des régions potentiellement soumises à sélection. Néanmoins aucune des deux catégories de marqueurs (potentiellement soumise et non soumise) ne contribuait plus que l'autre aux deux premières composantes, et deux ACP séparées sur chacun des six marqueurs (non montrées ici) conduisent au même résultat. Ainsi des caractères supposés pour partie adaptatifs structurent de la même manière les différentes espèces que les marqueurs dits neutres. La divergence entre marqueurs n'est pas indépendante de l'histoire évolutive des espèces et explique leur séparation de façon similaire aux caractères adaptatifs. C'est un résultat qui a déjà été observé Camus-Kulandaivelu *et al.* (2006) dans une étude d'investigation de la base génétique de l'adaptation du maïs aux climats tempérés sur plusieurs centaines de lignées et cultivars de maïs ; ils ont montré que les groupes établis sur la bases de 55 marqueurs microsatellites répartis sur tout le génome étaient fortement différenciés pour le caractère date de floraison, un caractère important pour l'adaptation à un climat donné.

La séparation des deux espèces parentales de l'espèce hybride sur la première composante de l'ACP n'était pas triviale, étant donné que le taux de fertilité entre l'espèce hybride et chacune des deux espèces parentales est plus important qu'entre les deux espèces parentales elles-mêmes (voir Figure 8 de l'introduction). Elle confirme l'existence et l'importance de flux de gènes entre espèces parentales, due à la répartition géographique partiellement en sympatrie de *H. annuus* et *H. petiolaris* et à un flux de gènes réduit de *H. paradoxus* isolé dans sa propre niche écologique.

La spéciation a beaucoup été étudiée par la recherche de barrières à la reproduction entre deux espèces, des gènes ont été trouvés mais ils n'ont pas de fonction particulière (Coyne et Orr 2004, pp 313-319). Pour comprendre le rôle de l'adaptation dans la spéciation, l'intérêt des approches écologiques est particulièrement probant. Dans les populations naturelles, l'histoire des populations peut être étudiée grâce à la diversité génétique au niveau des marqueurs moléculaires et être mise en liaison avec la diversité phénotypique.

Ici nous avons montré que l'adaptation de *H. paradoxus* à son milieu est à la fois liée à un mécanisme génétique et à une réponse plastique. Il existe une variabilité phénotypique qui n'est pas stable et dépend du fond génétique et de l'environnement. La variabilité plastique permet l'adaptation rapide à un nouveau milieu. C'est aussi ce qui est observé dans le cas de la domestication du maïs à partir de la téosinte (Lauter et Doebley 2002). Les allèles cultivés sont présents dans les populations de téosintes actuelles, mais ne sont 'allumés' que dans des conditions particulières. Cela montre l'importance des interactions épistatiques et de la variation plastique qui permet une adaptation rapide, et qui laisse du temps avant l'apparition d'un spectre adaptatif plus réduit.

REFERENCES

Références

- Abbott R (1992) Plant invasions, interspecific hybridization and the evolution of new plant taxa. *Trends in Ecology and Evolution*, **7**, 401-405.
- Andolfatto P (2001) Adaptive hitchhiking effects on genome variability. *Current Opinion in Genetics & Development*, **11**, 635-641.
- Arnold ML (1997) *Natural hybridization and evolution* Oxford University Press, New York.
- Arnold ML, Hamrick JL, Bennett BD (1990) Allozyme variation in Louisiana rises: a test for introgression and hybrid speciation. *Heredity*, **65**, 297-306.
- Barbash DA, Awadalla P, Tarone AM (2004) Functional divergence caused by ancient positive selection of a *Drosophila* hybrid incompatibility locus. *PLoS Biology*, **2**, e142.
- Bateson W (1922) Evolutionary faith and modern doubts. *Science*, **55**, 55-61.
- Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society B: Biological Sciences*, **263**, 1619-1626.
- Bello Y, Waxman D (2006) Near-periodic substitution and the genetic variance induced by environmental change. *Journal of Theoretical Biology*, **239**, 152-160.
- Bennett JH (1954) On the theory of random mating. *Annals of Eugenics*, **18**, 311-317.
- Bowcock AM, Kidd JR, Mountain JL, *et al.* (1991) Drift, admixture, and selection in human evolution: a study with DNA polymorphisms. *Proceedings of the National Academy of Sciences, U S A*, **88**, 839-843.
- Bradshaw HD, Schemske DW (2003) Allele substitution at a flower colour locus produces a pollinator shift in monkeyflowers. *Nature*, **426**, 176-178.
- Brideau NJ, Flores HA, Wang J, *et al.* (2006) Two Dobzhansky-Muller genes interact to cause hybrid lethality in *Drosophila*. *Science*, **314**, 1292-1295.
- Buerkle CA, Morris RJ, Asmussen MA, Rieseberg LH (2000) The likelihood of homoploid hybrid speciation. *Heredity*, **84 (Pt 4)**, 441-451.
- Camus-Kulandaivelu L, Veyrieras JB, Madur D, *et al.* (2006) Maize adaptation to temperate climate: relationship between population structure and polymorphism in the *Dwarf8* gene. *Genetics*, **172**, 2449-2463.

- Cano JM, Matsuba C, Makinen H, Merila J (2006) The utility of QTL-Linked markers to detect selective sweeps in natural populations - a case study of the *EDA* gene and a linked marker in threespine stickleback. *Molecular Ecology*, **15**, 4613-4621.
- Cavalli-Sforza LL (1966) Population structure and human evolution. *Proceedings of the Royal Society B: Biological Sciences*, **164**, 362-379.
- Chardon F, Virlon B, Moreau L, *et al.* (2004) Genetic architecture of flowering time in maize as inferred from quantitative trait loci meta-analysis and synteny conservation with the rice genome. *Genetics*, **168**, 2169-2185.
- Colosimo PF, Hosemann KE, Balabhadra S, *et al.* (2005) Widespread parallel evolution in sticklebacks by repeated fixation of *Ectodysplasin* alleles. *Science*, **307**, 1928-1933.
- Coyne JA, Orr HA (1998) The evolutionary genetics of speciation. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **353**, 287-305.
- Coyne JA, Orr HA (2004) *Speciation*, Sunderland, MA.
- Damerval C, Maurice A, Josse JM, de Vienne D (1994) Quantitative trait loci underlying gene product variation: a novel perspective for analyzing regulation of genome expression. *Genetics*, **137**, 289-301.
- Darwin C (1859) *On the origin of species by means of natural selection or the preservation of favored races in the struggle for life*, London.
- Davis AW, Wu CI (1996) The broom of the sorcerer's apprentice: the fine structure of a chromosomal region causing reproductive isolation between two sibling species of *Drosophila*. *Genetics*, **143**, 1287-1298.
- DeVries H (1906) *Species and varieties, their origin by mutation*, Chicago, IL.
- Dieckmann U, Doebeli M (1999) On the origin of species by sympatric speciation. *Nature*, **400**, 354-357.
- Dobzhansky T (1935) A critique of the species concept. *Philosophy of Science*, **2**, 344-355.
- Dobzhansky T (1937) *Genetics and the origin of species* Columbia university press, New York.
- Doebeli M, Dieckmann U (2003) Speciation along environmental gradients. *Nature*, **421**, 259-264.
- Drossel B, McKane A (2000) Competitive speciation in quantitative genetic models. *Journal of Theoretical Biology*, **204**, 467-478.
- Ehrenreich IM, Purugganan MD (2006) The molecular genetic basis of plant adaptation. *American Journal of Botany*, **93**, 953-962.

- Farnir F, Coppieters W, Arranz JJ, *et al.* (2000) Extensive genome-wide linkage disequilibrium in cattle. *Genome Research*, **10**, 220-227.
- Fay JC, Wu CI (2000) Hitchhiking under positive Darwinian selection. *Genetics*, **155**, 1405-1413.
- Fisher RA (1930) *The genetical theory of natural selection*, Oxford.
- Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. *Genetics*, **133**, 693-709.
- Gallez GP, Gottlieb LD (1982) Genetic evidence for the hybrid origin of the diploid plant *Stephanomeria diegensis*. *Evolution*, **36**, 1158-1167.
- Gibson G, Weir B (2005) The quantitative genetics of transcription. *Trends in Genetics*, **21**, 616-623.
- Gillespie JH (1991) *The causes of molecular evolution*, Oxford.
- Goldringer I, Bataillon T (2004) On the distribution of temporal variations in allele frequency: consequences for the estimation of effective population size and the detection of loci undergoing selection. *Genetics*, **168**, 563-568.
- Grant V (1958) The regulation of recombination in plants. *Cold Spring Harbor Symposia on Quantitative Biology*, **23**, 337-363.
- Grant V (1981) *Plant speciation*, New York, USA.
- Gross BL, Rieseberg LH (2005) The ecological genetics of homoploid hybrid speciation. *Journal of Heredity*, **96**, 241-252.
- Gross BL, Schwarzbach AE, Rieseberg LH (2003) Origin(s) of the diploid hybrid species *Helianthus deserticola* (Asteraceae). *American Journal of Botany*, **90**, 1708-1719.
- Haldane JBS (1922) Sex ratio and unisexual sterility in animal hybrids. *Journal of Genetics*, **12**, 101-109.
- Hatfield T, Schluter D (1999) Ecological speciation in sticklebacks: environment-dependent hybrid fitness. *Evolution*, **53**, 866-873.
- Hawthorne DJ, Via S (2001) Genetic linkage of ecological specialization and reproductive isolation in pea aphids. *Nature*, **412**, 904-907.
- Heiser CB (1947) Hybridization between the sunflower species *Helianthus annuus* and *H. petiolaris*. *Evolution*, **1**, 249-262.
- Heiser CB (1949) Study in the evolution of the sunflower species *Helianthus annuus* and *H. bolanderi*. *University of California Publications in Botany*, **23**, 157-196.
- Heiser CB (1958) Three new annual sunflowers (*Helianthus*) from the southwestern United States. *Rhodora*, **60**, 272-283.

- Hey J (2001) The mind of the species problem. *Trends in Ecology and Evolution*, **16**, 326-329.
- Higashi M, Takimoto G, Yamamura N (1999) Sympatric speciation by sexual selection. *Nature*, **402**, 523-526.
- Hirai H, Taguchi T, Saitoh Y, *et al.* (2000) Chromosomal differentiation of the *Schistosoma japonicum* complex. *International Journal for Parasitology*, **30**, 441-452.
- Hollocher H, Wu CI (1996) The genetics of reproductive isolation in the *Drosophila simulans* clade: X vs. autosomal effects and male vs. female effects. *Genetics*, **143**, 1243-1255.
- Hudson RR (2001) Two-locus sampling distributions and their application. *Genetics*, **159**, 1805-1817.
- Hudson RR, Kreitman M, Aguade M (1987) A test of neutral molecular evolution based on nucleotide data. *Genetics*, **116**, 153-159.
- Kaplan NL, Hudson RR, Langley CH (1989) The "hitchhiking effect" revisited. *Genetics*, **123**, 887-899.
- Karrenberg S, Edelist C, Lexer C, Rieseberg L (2006) Response to salinity in the homoploid hybrid species *Helianthus paradoxus* and its progenitors *H. annuus* and *H. petiolaris*. *New Phytologist*, **170**, 615-629.
- Kim Y, Orr HA (2005) Adaptation in sexuals vs. asexuals: clonal interference and the Fisher-Muller model. *Genetics*, **171**, 1377-1386.
- Kim Y, Stephan W (2000) Joint effects of genetic hitchhiking and background selection on neutral variation. *Genetics*, **155**, 1415-1427.
- Kim Y, Stephan W (2002) Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics*, **160**, 765-777.
- Kimura M (1983) *The neutral theory of molecular evolution*, Cambridge.
- Kreitman M (2000) Methods to detect selection in populations with applications to the human. *Annual Review of Genomics and Human Genetics*, **1**, 539-559.
- Lai Z, Gross BL, Zou Y, Andrews J, Rieseberg LH (2006) Microarray analysis reveals differential gene expression in hybrid sunflower species. *Molecular Ecology*, **15**, 1213-1227.
- Lai Z, Nakazato T, Salmaso M, *et al.* (2005) Extensive chromosomal repatterning and the evolution of sterility barriers in hybrid sunflower species. *Genetics*, **171**, 291-303.
- Lauter N, Doebley J (2002) Genetic variation for phenotypically invariant traits detected in teosinte: implications for the evolution of novel forms. *Genetics*, **160**, 333-342.
- Levin DA (1983) Polyploidy and novelty in flowering plants. *American Naturalist*, **122**, 1-25.

- Levin DA (2002) *The role of chromosomal change in plant evolution*, Oxford.
- Lewontin RC (1964) The Interaction of Selection and Linkage II. Optimum Models. *Genetics*, **50**, 757-782.
- Lewontin RC (1974) The problem of genetic diversity. *Harvey Lectures*, **70 Series**, 1-20.
- Lewontin RC, Birch LC (1966) Hybridisation as a source of variation for adaptation to new environments. *Evolution*, **20**, 315-336.
- Lewontin RC, Krakauer J (1973) Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics*, **74**, 175-195.
- Lexer C, Welch ME, Durphy JL, Rieseberg LH (2003) Natural selection for salt tolerance quantitative trait loci (QTLs) in wild sunflower hybrids: implications for the origin of *Helianthus paradoxus*, a diploid hybrid species. *Molecular Ecology*, **12**, 1225-1235.
- Lu G, Bernatchez L (1999) Correlated trophic specialization and genetic divergence in sympatric lake whitefish ecotypes (*Coregonus clupeaformis*): support for the ecological speciation hypothesis. *Evolution*, **53**, 1491-1505.
- Machado CA, Kliman RM, Markert JA, Hey J (2002) Inferring the history of speciation from multilocus DNA sequence data: the case of *Drosophila pseudoobscura* and close relatives. *Molecular Biology and Evolution*, **19**, 472-488.
- Mackay TF (2001) The genetic architecture of quantitative traits. *Annual Review of Genetics*, **35**, 303-339.
- Maynard Smith J, Haigh J (1974) The hitchhiking effect of a favourable gene. *Genetical Research*, **23**, 23-35.
- Mayr E (1942) *Systematics and the origin of species*, New York.
- Mayr E (1963) *Animal species and evolution*, Cambridge, MA.
- McCarthy EM, Asmussen MA, Anderson WW (1995) A theoretical assessment of recombinational speciation. *Heredity*, **74**, 502-509.
- McDonald JH, Kreitman M (1991) Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature*, **351**, 652-654.
- Moran PAP (1975) Wandering distributions and the electrophoretical profile. *Theoretical Population Biology*, **8**, 318-330.
- Muller HJ (1925) Why polyploidy is rarer in animals than in plants. *American Naturalist*, **59**, 346-353.
- Muller HJ (1940) *Bearing of the Drosophila work on systematics*, Oxford.
- Nagel L, Schluter D (1998) Body size, natural selection, and speciation in sticklebacks. *Evolution*, **52**, 209-218.

- Navarro A, Barton NH (2003) Accumulating postzygotic isolation genes in parapatry: a new twist on chromosomal speciation. *Evolution*, **57**, 447-459.
- Nielsen R (2001) Statistical tests of selective neutrality in the age of genomics. *Heredity*, **86**, 641-647.
- Noor MA, Grams KL, Bertucci LA, *et al.* (2001a) The genetics of reproductive isolation and the potential for gene exchange between *Drosophila pseudoobscura* and *D. persimilis* via backcross hybrid males. *Evolution*, **55**, 512-521.
- Noor MA, Grams KL, Bertucci LA, Reiland J (2001b) Chromosomal inversions and the reproductive isolation of species. *Proceedings of the National Academy of Sciences, U S A*, **98**, 12084-12088.
- Ogden R, Thorpe RS (2002) Molecular evidence for ecological speciation in tropical habitats. *Proceedings of the National Academy of Sciences, U S A*, **99**, 13612-13615.
- Ohta T, Kimura M (1973) A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genetical Research*, **22**, 201-204.
- Orr HA (1993) Haldane's rule has multiple genetic causes. *Nature*, **361**, 532-533.
- Orr HA (1998) Testing natural selection vs. genetic drift in phenotypic evolution using quantitative trait locus data. *Genetics*, **149**, 2099-2104.
- Orr HA (2001) The genetics of species difference. *Trends in Ecology and Evolution*, **16**, 343-350.
- Orr HA (2005a) The genetic theory of adaptation: a brief history. *Nature Reviews. Genetics.*, **6**, 119-127.
- Orr HA (2005b) Theories of adaptation: what they do and don't say. *Genetica*, **123**, 3-13.
- Otto SP, Whitton J (2000) Polyploid incidence and evolution. *Annual Review of Genetics*, **34**, 401-437.
- Phillips PC (2005) Testing hypotheses regarding the genetics of adaptation. *Genetica*, **123**, 15-24.
- Podos J (2001) Correlated evolution of morphology and vocal signal structure in Darwin's finches. *Nature*, **409**, 185-188.
- Presgraves DC (2003) A fine-scale genetic analysis of hybrid incompatibilities in *Drosophila*. *Genetics*, **163**, 955-972.
- 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.

- Ratcliffe LM, Grant PR (1983) Species recognition in Darwin's finches (*Geospiza*, Gould). II: geographic variation in mate preference. *Animal Behaviour*, **31**, 1154-1165.
- Remington DL, Thornsberry JM, Matsuoka Y, *et al.* (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proceedings of the National Academy of Sciences, U S A*, **98**, 11479-11484.
- Rieseberg LH (1991) Homoploid reticulate evolution in *Helianthus* (Asteraceae) : evidence from ribosomal genes. *American Journal of Botany*, **78**, 1218-1237.
- Rieseberg LH (1997) Hybrid origin of plant species. *Annual Review of Ecology and Systematics*, **28**, 359-389.
- Rieseberg LH (2001) Chromosomal rearrangements and speciation. *Trends in Ecology and Evolution*, **16**, 351-358.
- Rieseberg LH, Carter R, Zona S (1990) Molecular tests of the hypothesized hybrid origin of two diploid *Helianthus* species (Asteraceae). *Evolution*, **44**, 1498-1511.
- Rieseberg LH, Linder CR, Seiler GJ (1995) Chromosomal and genic barriers to introgression in *Helianthus*. *Genetics*, **141**, 1163-1171.
- Rieseberg LH, Raymond O, Rosenthal DM, *et al.* (2003) Major ecological transitions in wild sunflowers facilitated by hybridization. *Science*, **301**, 1211-1216.
- Rieseberg LH, Archer MA, Wayne RK (1999) Transgressive segregation, adaptation and speciation. *Heredity*, **83 (Pt 4)**, 363-372.
- Rieseberg LH, Wood TE, Baack EJ (2006) The nature of plant species. *Nature*, **440**, 524-527.
- Rodriguez DJ (1996) A model for the establishment of polyploidy in plants. *American Naturalist*, **147**, 33-46.
- Rosenthal DM, Rieseberg LH, Donovan LA (2005) Re-creating ancient hybrid species' complex phenotypes from early-generation synthetic hybrids: three examples using wild sunflowers. *American Naturalist*, **166**, 26-41.
- Rundle HD, Nagel L, Wenrick Boughman J, Schluter D (2000) Natural selection and parallel speciation in sympatric sticklebacks. *Science*, **287**, 306-308.
- Sawamura K, Davis AW, Wu CI (2000) Genetic analysis of speciation by means of introgression into *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences, U S A*, **97**, 2652-2655.
- Schemske DW (2000) Understanding the origin of species. *Evolution*, **54**, 1069-1073.
- Schilling EE, Heiser CB (1981) An infrageneric classification of *Helianthus*. Compositae. *Taxon*, **30**, 393-403.

- Schlötterer C (2002) A microsatellite-based multilocus screen for the identification of local selective sweeps. *Genetics*, **160**, 753-763.
- Schlötterer C, Dieringer D (2005) A novel test statistic for the identification of local selective sweeps based on microsatellite gene diversity. In: *Selective Sweep* (ed. Nurminsky D), pp. 55-64. Landes Bioscience, New York.
- Schlötterer C, Vogl C, Tautz D (1997) Polymorphism and locus-specific effects on polymorphism at microsatellite loci in natural *Drosophila melanogaster* populations. *Genetics*, **146**, 309-320.
- Schluter D (2001) Ecology and the origin of species. *Trends in Ecology and Evolution*, **16**, 372-380.
- Schwarzbach AE, Donovan LA, Rieseberg LH (2001) Transgressive character expression in a hybrid sunflower species. *American Journal of Botany*, **88**, 270-277.
- Schwarzbach AE, Rieseberg LH (2002) Likely multiple origins of a diploid hybrid sunflower species. *Molecular Ecology*, **11**, 1703-1715.
- Slatkin M (1995) Hitchhiking and associative overdominance at a microsatellite locus. *Molecular Biology and Evolution*, **12**, 473-480.
- Soltis DE, Soltis PS (1999) Polyploidy: recurrent formation and genome evolution. *Trends in Ecology and Evolution*, **14**, 348-352.
- Stebbins GL (1950) *Variation and evolution in plants*, New York.
- Stebbins GL (1957) Self-fertilization and population variability in the higher plants. *American Naturalist*, **91**, 337-354.
- Stephan W, Mitchell SJ (1992) The effect of strongly selected substitutions on neutral polymorphism: analytical results based on diffusion theory. *Theoretical Population Biology*, **41**, 237-254.
- Storz JF (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology*, **14**, 671-688.
- Sucena E, Stern DL (2000) Divergence of larval morphology between *Drosophila sechellia* and its sibling species caused by cis-regulatory evolution of ovo/shaven-baby. *Proceedings of the National Academy of Sciences, U S A*, **97**, 4530-4534.
- Sweigart AL, Fishman L, Willis JH (2006) A simple genetic incompatibility causes hybrid male sterility in mimulus. *Genetics*, **172**, 2465-2479.
- Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585-595.

- Takahashi A, Tsaur SC, Coyne JA, Wu CI (2001) The nucleotide changes governing cuticular hydrocarbon variation and their evolution in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences, U S A*, **98**, 3920-3925.
- Tang S, Yu JK, Slabaugh B, Shintani K, Knapp J (2002) Simple sequence repeat map of the sunflower genome. *Theoretical and Applied Genetics*, **105**, 1124-1136.
- Tao Y, Chen S, Hartl DL, Laurie CC (2003a) Genetic dissection of hybrid incompatibilities between *Drosophila simulans* and *D. mauritiana*. I. Differential accumulation of hybrid male sterility effects on the X and autosomes. *Genetics*, **164**, 1383-1397.
- Tao Y, Hartl DL (2003) Genetic dissection of hybrid incompatibilities between *Drosophila simulans* and *D. mauritiana*. III. Heterogeneous accumulation of hybrid incompatibilities, degree of dominance, and implications for Haldane's rule. *Evolution*, **57**, 2580-2598.
- Tao Y, Zeng ZB, Li J, Hartl DL, Laurie CC (2003b) Genetic dissection of hybrid incompatibilities between *Drosophila simulans* and *D. mauritiana*. II. Mapping hybrid male sterility loci on the third chromosome. *Genetics*, **164**, 1399-1418.
- Templeton AR (1980) The theory of speciation via the founder principle. *Genetics*, **94**, 1011-1038.
- Tenesa A, Wright AF, Knott SA, *et al.* (2004) Extent of linkage disequilibrium in a Sardinian sub-isolate: sampling and methodological considerations. *Human Molecular Genetics*, **13**, 25-33.
- Thompson JD, Lumaret R (1992) The evolutionary dynamics of polyploid plants: origins, establishment and persistence. *Trends in Ecology and Evolution*, **7**, 302-307.
- Ting CT, Tsaur SC, Wu ML, Wu CI (1998) A rapidly evolving homeobox at the site of a hybrid sterility gene. *Science*, **282**, 1501-1504.
- True JR, Weir BS, Laurie CC (1996) A genome-wide survey of hybrid incompatibility factors by the introgression of marked segments of *Drosophila mauritiana* chromosomes into *Drosophila simulans*. *Genetics*, **142**, 819-837.
- Turelli M, Orr HA (1995) The dominance theory of Haldane's rule. *Genetics*, **140**, 389-402.
- Turelli M, Orr HA (2000) Dominance, epistasis and the genetics of postzygotic isolation. *Genetics*, **154**, 1663-1679.
- Turner BL (1981) Letter to G. Seiler., Austin, TX.
- Vitalis R, Dawson K, Boursot P (2001) Interpretation of variation across marker loci as evidence of selection. *Genetics*, **158**, 1811-1823.

- Wang RL, Hey J (1996) The speciation history of *Drosophila pseudoobscura* and close relatives: inferences from DNA sequence variation at the period locus. *Genetics*, **144**, 1113-1126.
- Weinig C, Dorn LA, Kane NC, *et al.* (2003) Heterogeneous selection at specific loci in natural environments in *Arabidopsis thaliana*. *Genetics*, **165**, 321-329.
- Welch ME, Rieseberg LH (2002a) Habitat divergence between a homoploid hybrid sunflower species, *Helianthus paradoxus* (Asteraceae), and its progenitors. *American Journal of Botany*, **89**, 472-478.
- Welch ME, Rieseberg LH (2002b) Patterns of genetic variation suggest a single, ancient origin for the diploid hybrid species *Helianthus paradoxus*. *Evolution*, **56**, 2126-2137.
- Wiehe T (1998) The effect of selective sweeps on the variance of the allele distribution of a linked multiallele locus: hitchhiking of microsatellites. *Theoretical Population Biology*, **53**, 272-283.
- Wright S (1932) The role of mutation, inbreeding, crossbreeding and selection in evolution. **1**, 356-366.
- Wright S (1943) Isolation by distance. *Genetics*, **28**, 114-138.
- Wright S (1951) The genetical structure of populations. *Annals of Eugenics*, **15**, 323-354.
- Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. *Computer Applications in the Biosciences*, **13**, 555-556.
- Yang Z, Nielsen R (2002) Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Molecular Biology and Evolution*, **19**, 908-917.
- Zhivotovsky LA (2001) Estimating divergence time with the use of microsatellite genetic distances: impacts of population growth and gene flow. *Molecular Biology and Evolution*, **18**, 700-709.

ANNEXE

Article :

Karrenberg S., Edelist C., Lexer C. and Rieseberg L. H. 2006. Response to salinity in the homoploid hybrid species *Helianthus paradoxus* and its progenitors *H. annuus* and *H. petiolaris*. *New Phytologist*. 170: 615-629