Annabelle Haudry

Influence de la domestication et du système de reproduction sur la diversité et l'évolution des gènes chez les *Triticeae*

UNIVERSITE MONTPELLIER II SCIENCES ET TECHNIQUES DU LANGUEDOC

THESE

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE MONTPELLIER II

Discipline : Biologie des Populations et Ecologie Formation Doctorale : Ressources Phytogénétiques et Interactions Biologiques Ecole Doctorale : Systèmes Intégrés en Biologie, Agronomie, Géosciences, Hydrosciences, Environnement

présentée et soutenue publiquement

par

Annabelle HAUDRY

le 11 décembre 2007

Influence de la domestication et du système de reproduction sur la diversité et l'évolution des gènes chez les *Triticeae*

JURY

Isabelle OLIVIERI Maud TENAILLON Laurent DURET Rémy PETIT Jacques DAVID Sylvain GLEMIN , Examinatrice , Examinatrice , Rapporteur , Rapporteur , Directeur de Thèse

, Co-Directeur de Thèse

Remerciements

Tout d'abord je remercie le lecteur qui s'aventure dans ce document, en espérant qu'il l'intéressera. De nombreuses personnes ont permis la mise en place et l'aboutissement de ce travail, et j'aimerais les en remercier chaleureusement.

Je voudrais commencer par remercier **Isabelle Olivieri**. Merci beaucoup Isabelle, tu es la première personne à avoir eu confiance en moi, je n'aurais sûrement pas fait un DEA sans toi. C'est toi qui m'as donné les connaissances et la curiosité pour la génétique des populations, l'évolution. Merci pour m'avoir poussée, pour ne pas dire forcée, à aller faire de la génétique quantitative sur des plantes cultivées... ce qui était loin de m'enthousiasmer, en m'assurant qu'avec Thomas Bataillon et Patrice David ça serait du petit lait (ce qui s'est révélé exact). Merci encore pour avoir endosser le statut de tutrice de monitorat, et surtout pour m'avoir confier certains de tes cours, et m'avoir permis d'enseigner la génèt des pops, la coalescence... Encore merci pour avoir accepter d'être la présidente du jury, même après m'avoir confier que l'évolution moléculaire t'ennuyait, merci plus récemment pour tes recommandations lors de mes candidatures de postdoc.... J'en oublie ?

Je remercie l'ensemble des membres du jury d'avoir enrichi la présentation de mon travail de leurs commentaires. En particulier, je remercie les deux rapporteurs **Laurent Duret** et **Rémy Petit** pour leurs remarques et les ouvertures proposées sur mon travail lors de la discussion, ainsi que **Maud Tenaillon** qui n'a malheureusement pas pu assister à la soutenance mais m'a accordé une longue discussion téléphonique pour aborder notamment le concept de domestication et sa modélisation.

Merci également à **Catherine Feuillet**, **Maud** encore, **Thomas Bataillon**, **Nicolas Bierne**, **Gabriel Marais**, **Yannis Michalakis** pour leur aide dans la réflexion sur mon sujet et les choix stratégiques lors de mes comités de thèse.

De façon plus quotidienne, j'ai côtoyé un certain nombre de personnes qui m'ont accompagnée au cours de ces 4 dernières années. J'ai bien peur de ne pas pouvoir nommer tout le monde, mais je tiens tout particulièrement à remercier **Georgette** et **Alain** pour m'avoir tous les jours préparé mon petit poisson... ainsi qu'à tous les beloteurs pour les moments de détente autour du café, si importants avant de se replonger dans le boulot ! Merci **aux secrétaires** qui m'ont toujours épaulée dans ma gestion plus qu'approximative des multiples paperasses administratives. Un grand merci à toutes les personnes de Mauguio qui font de la station un endroit si convivial !

Parmi la grande famille de Mauguio, certains personnages ont joué des rôles cruciaux dans le déroulement de ma thèse.

Je pense d'abord à **Joëlle Ronfort**. « Djo » avec qui j'ai partagé le bureau pendant 3ans (du bon vieux temps où on pouvait encore fumer dedans !) et qui fait partie de mes précieuses anges gardiennes. Pendant ma course effrénée au stage de DEA, c'est toi qui m'as proposé de contacter un certain Jacques David qui revenait tout juste des US, qu'il aurait peut-être un sujet de stage à me proposer... le début d'une longue (et belle) histoire. Mais surtout merci beaucoup pour ta gentillesse, ton écoute au quotidien, ta disponibilité, pour m'avoir rassurée environ mille et une fois lors de mes « plans poubelle»... merci pour le temps que tu as consacré à répondre à toutes mes questions et à essayer de défaire les nœuds que je faisais dans ma tête. Notamment je pense au coup de stress des derniers instants : « mais au fait pourquoi théta ? », ou encore quand tu m'as installée dans la salle de réunion pour t'exposer en vrac le GCétoile, Fopétoile, oméga et BGC et

m'aider à y voir plus clair...Merci à **Christophe**, **Duane**, **Hugo** et à toi pour m'avoir accueillie plusieurs fois (voire recueillie), et quand je pétais un plomb, regarder le match France-Angleterre avec moi. Merci.

Sur les conseils de Djo, je suis donc venue un après-midi à la station pour discuter avec Jacques. Il s'est installé devant le tableau blanc pour me raconter l'histoire du blé, de sa domestication, de la quantité faramineuse de données disponibles (des séquences, cooooooool !!) et pas moins de questions auxquelles s'intéresser... c'est parti dans tous les sens, un vrai feu d'artifice ! J'ai tout de suite été séduite. Je suis finalement restée 5h pour une discussion que je pensais d'une demiheure... et Toto de s'étonner de me voir encore là « tu ne voulais pas bosser sur la symbiose et la coévolution ???!! ». Euh...si c'est vrai, d'ailleurs je remercie **Gilles Béna** qui ne m'en a pas voulu quand j'ai décidé de changer de stage...

Je voudrais remercier le grand chef **Jean-Marie Prosperi** qui m'a accueillie à la station et qui m'a accordé de précieux conseils quant à la façon d'envisager mon avenir dans le milieu de l'enseignement et de la recherche, de manière assez paternelle. Je voudrais également remercier **Pierre Roumet** qui m'a tout d'abord accueillie dans son bureau qu'il partageait avec Jacques, où j'ai plus ou moins perturbé l'ambiance masculine. Merci également Pierre pour avoir fait l'effort de relire et commenter consciencieusement tout ce que je t'ai demandé (depuis le rapport biblio de DEA à la thèse, tu as dû tout lire je crois !) et pour ta disponibilité pour répondre à mes questions concernant l'amélioration du blé (entre autres). Merci pour ta bonne humeur, pour tes blagues, même si elles ne m'ont pas toutes faite rire ;)

Je tiens à remercier toutes mes anges gardiennes, aussi précieuses les unes que les autres à la fois pour leur gentillesse sans borne et leur écoute scientifique. Merci à Marie-Hélène Muller, « MH », la première que j'ai eu le plaisir de rencontrer et que je connaissais par sa thèse... j'étais super impressionnée, et en fait je le suis toujours ! Merci d'être toi, généreuse et toujours disposée à expliquer brillamment la coalescence avec les mains ! Merci à Anne Tsitrone, « Nane », pour ta clairvoyance et ta pédagogie lors de nos discussions en BR, notamment devant les équations que plus d'un (dont moi) trouvaient effrayantes ! Merci pour tes conseils et relectures anglophones. Merci à Nathalie Chantret, « Nath » qui comptait parmi mes nombreux encadrants de DEA, et m'a appris beaucoup sur la génomique chez les Triticées. Merci aussi pour ta patience et ta gentillesse, pour les pizzas, les boîtes à chat et j'en passe... Dans la catégorie des bonnes fées, je voudrais également citer Marie-France Ostrowski, « MF », entière et passionnée : merci pour ton écoute à la fois amicale et scientifique, pour ton altruisme, pour le réconfort que tu m'as apporté, et aussi pour m'avoir fait découvert de nombreuses saveurs culinaires (j'en profite pour te demander la recette du gâteau à la carotte et jus d'orange). J'ai également eu la chance de côtoyer Muriel Tavaud-Pirra, plus connue localement en tant que « Mumax », qui n'a cessé de me rassurer et de m'encourager notamment lors de notre rituel thé de 18h me permettant une pause avant le craquage pendant la rédaction. Une autre ange gardienne a rejoint plus tardivement le groupe, de retour de postdoc, Anne-Céline Thuillet, ma prédécéceuse de choc, qui a su me prodiquer de précieux conseils sur la gestion de soi et de son temps lors de nos séances à la piscine. Une grand merci à vous toutes, mes bonnes fées, pour m'avoir soutenue et supportée (dans les deux sens du terme) lors de mes crises de confiance en moi, qui ont ponctué ces quatre dernières années.

Capitaine à bord de ce joyeux cargo, je voudrais remercier tout particulièrement le grand manitou, celui qui a « fait le ouistiti sur mon épaule » pendant tout ce temps, l'idole des jeunes, vous l'aurez reconnu, « JackiJak », **Jacques David**. Merci à toi pour m'avoir poussée à donner le meilleur de moi, à m'ouvrir à de nombreuses questions... me forçant même à m'intéresser à l'amélioration des

plantes en me montrant qu'il ne s'agissait pas forcément d'un problème industriel mais qu'on pouvait aussi faire de la science et « sauver les petits noirs ». Merci pour ta patience et tes efforts pour te rendre disponible, à discuter des flux de gènes devant une assiette de pâtes, et même ton écoute à mes problèmes de cœur ! Merci pour avoir su rester stoïque lorsque je fondais en larmes, et continuer la discussion, en comprenant que « c'était juste car je prenais les choses à cœur », courir après un paquet de kleenex et à réclamer la boîte de chocolats le lendemain ! Merci pour m'avoir guidée tout au long du chemin, en me laissant râler contre tes commentaires « tu en fais bien ce que tu veux », sachant que je finissais toujours par les prendre en compte… merci pour m'avoir débloquée de nombreuses fois grâce à tes scripts magiques, ou tes idées lumineuses. Merci pour ton soutien et tes encouragements qui m'ont permis d'y croire, et pour ton enthousiasme si communicatif.

Et un dernier pour la route (ou pas) : merci d'avoir demandé le concours de **Sylvain Glémin** pour encadrer cette thèse. Et merci Sylvain pour avoir accepté d'endosser le rôle de chef !! Merci pour ta gentillesse, ton écoute, et avoir pris le relais sur mon épaule la dernière année. Merci pour tes précieux conseils, concernant les enseignements ou la thèse, ou encore la gestion du temps... hum, et ta patience car je ne me suis pas montrée une très bonne élève ! Merci pour m'avoir empêchée de m'éparpiller, autant que possible, et illustrer magnifiquement (bien que très modestement) l'exemple à suivre... sujet de stress forcément, mais qui a surtout été un moteur pour ne pas (trop) te décevoir... j'y travaille encore. Merci pour toutes les choses que tu m'as expliquées (la liste est longue), je garde précieusement tes petits schémas si pédagogiques, et pour tout ce que tu m'as appris sur l'évolution moléculaire et le système de reproduction... bien des équations ou notions seraient restées obscures si tu ne les avais pas éclairées !!! Un énoooorme merci à tous les deux, mes deux ouistitis, pour la richesse de votre encadrement, mille fois merci d'avoir été là. MERCI !!!

Parmi les gens qui m'ont épaulée au quotidien, je pense également aux autres thésards de Mauguio. Notre illustre aîné Xavier Lacaze qui nous a montré la voie, le plus souvent avec sagesse, mais je le remercie également pour ses craquages réguliers, à savoir tous les jours aux alentours de 16h30 qui ponctuaient nos journées dans la salle informatique de franches rigolades. Quand j'ai débarqué à Mauquio, deux affreux commençaient tout juste leur thèse, et j'ai eu beaucoup de chance de passer plus de 3 ans à leur côté. En plus de leur compréhension lors des phases difficiles, ils m'ont apporté le bénéfice de leur expérience toute récente dans le parcours de la thèse et de leurs connaissances, ainsi que leur soutien. Le premier, Stéphane De Mita avec qui j'ai partagé de longs moments à débattre devant un tableau blanc (d'abord puis gribouillé) de la définition d'un qoulot d'étranglement, à décortiquer le D de Tajima, et plus largement à partager la foi en l'analyse de séquences. Merci Stef d'avoir toujours été disponible pour répondre à mes nombreuses questions, même les plus débiles (et même par skype quand tu es parti en postdoc), m'avoir dégoter nombre de papiers et pour m'avoir poussée à me poser des questions pas débiles du tout ! Le second, Mathieu Siol qui m'a très rapidement attribué le surnom de Minuscule (allez savoir pourquoi!) qui m'a suivi pendant toute la thèse... Merci Mathieu de m'avoir montré que la persévérance était payante, autant en sciences qu'au ping pong, et merci pour la bonne humeur que tu as su apporter dans nos bureaux ! Merci à vous deux pour tous les bons moments qui n'ont pas manqué (mais qui me manquent déjà) : les parties de GTA, de Wolfenstein, de pinq pong ou tout autre pétage de plomb essentiel pour se défouler, mais aussi les discussions autour de thés, bières ou pizzas, et des soirées d'hiver passées à Mauquio ensemble. Une autre thésarde contemporaine, Letizia Camus-Kulandaivelu, a partagé pendant quelques mois nos délires à la station, et m'a entraînée avec elle à la piscine, et je l'en remercie, ça m'a fait du bien. Je passe aussi un salut amical à Valérie Loywyck qui nous a fait profité de son optimisme débordant pendant quelques mois. J'en profite pour saluer **Yves Rousselle**, qui nous a malheureusement abandonnés après son DEA dans nos murs, le remercier pour son amitié, et lui souhaiter bon courage pour sa thèse. Enfin, tous mes encouragements à **Laurence Vaissayre**, qui venait de commencer sa thèse et que j'ai du effrayer lors de la rédaction !

Même si j'ai choisi d'effectuer la plupart de ma thèse à la station de Mauguio, j'ai été de nombreuses fois hébergée dans un bureau de l'ISEM, et je voudrais saluer mes cothésards de la fac qui ont toujours su me faire une petite place dans leur bureau : merci à Julien Dutheil ainsi que pour tes conseils en programmation, pour les démonstrations de réanimation ou celles plus clownesques lors des formations du CIES, à Pierre-Henri Fabre aussi pour ton humour caustique et à Benoit Nabholz pour ta gentillesse. Merci également à Nicolas Galtier et Emmanuel Douzery pour m'avoir accueillie dans leurs murs et pour les discussions, trop peu nombreuses mais riches, qu'on a pu partager. Je salue également Eric Bazin, que j'ai croisé dans les premiers mois de ma thèse, qui est maintenant en postdoc et lui souhaite une heureuse continuation.

Durant ma thèse, j'ai également passé de nombreuses heures à la fac, en dehors de l'ISEM, à donner des cours pour le monitorat. Je voudrais remercier les enseignants qui m'ont confié leurs étudiants : Sandrine Maurice, Cécile Meunier, Agnès Mignot, Marie-Hélène Muller, Isabelle Olivieri, Eric Imbert, Sylvain Glémin, David Mouillot, Christophe Petit, Paul Roiron, Marc-André Sélosse. J'en profite pour remercier chaleureusement tous les membres de l'équipe du bâtiment 4 pour leur aide et leur bonne humeur : Sylvie Lanau, Colin Durand, Olivier Guirado, Esteban Martinez et Thierry Vindolet. Je salue au passage mes étudiants en leur souhaitant une bonne continuation.

On change encore de murs, pour l'ENSAM. J'ai passé quelques mois au labo de BM pour « maniper » et je voudrais remercier l'ensemble de l'équipe qui m'a chaleureusement accueillie et guidée là-bas. Je pense particulièrement à **Sylvain Santoni** qui m'a ouvert bien grandes les portes du labo, en restant toujours disponible pour discuter aussi bien technique (« mais pour ne pas perdre les billes, il faut secouer fort comment ?? ») que théorique, ainsi celles de son bureau pour m'offrir son écoute amicale. Dans les murs de l'ENSAM se cache une autre personne qui m'a énormément appris, **Alberto Cenci**. Merci Albé pour le temps que tu as passé à m'apprendre la génomique, comment reconnaître les exons et les introns, blaster en série, fouiller les bases de données, etc., si je sais lire des séquences (malheureusement pas encore en italien !), c'est grâce à toi.

Pendant mes périodes de manipes, d'autres personnes m'ont épaulée et ont participé à ma formation en BM. Je tiens tout particulièrement à remercier **Cinderella Grout** pour les sauvetages multiples, et même jusqu'à tard dans la soirée ou au téléphone, m'apprenant qu'il n'était pas bon de maniper la nuit, et me réconfortant de mes maladresses devant une assiette de pâtes, un thé ou un resto... Je remercie également **Sébastien Poirier** et **Claire Guilhaumon** a qui revient le mérite de la grande majorité des données produites, et qui m'ont facilité le travail de manière considérable. Je les remercie d'autant plus pour leur réactivité et leur compréhension devant mes demandes express (« euh, dis Claire, tu ne voudrais pas refaire tous les gènes pour le seigle aussi siteuplé ? »). Merci à **Audrey Weber** qui n'a jamais été avare de conseils qui m'ont été fort utiles pour ne pas trop faire de bourdes, ainsi qu'à **Isabelle Hochu** aussi pour la recette de la galette des rois et à **Charles Poncet** pour les nombreux divertissements qu'il m'a offert lors de mes passages au labo.

Je voudrais également remercier d'une part **Isabelle Bonnin** et **Catherine Ravel** qui ont participé au séquençage de certains gènes inclus dans l'analyse de polymorphisme du blé et d'autre part **Béatrice Ramora** pour le soin qu'elle a porté à mes petites plantes et **Jean-Claude Dusseautoir** pour les

discussions qu'on a échangées sur l'intérêt des différents croisements intra et interspécifiques chez les Triticées.

Je ne peux passer sous silence mes amis montpelliérains qui m'ont soutenu sans relâche et ont partagé des soirées de beuveries glorieuses : une mention toute spéciale pour **Matthieu**, pour avoir été là dans les bons moments comme dans les phases difficiles, m'avoir épaulée et entraînée vers le haut de la vague. Merci pour ta précieuse amitié. Je pense aussi à **Alexia** ou l'incarnation de la pêche et ses soirées à thèmes..., **Anne** ma collègue de domestication, **Claire** ma meilleure complice de « déhanchage intempestif », **Elsa** ma collègue de DEA et partenaire d'impro, **Emilie** merci aussi pour nos délires et notre trip à la run et au volcan... inoubliable, et mes autres collègues de fiesta **Ana**, **Anne-Violette**, **Eric**, **François**, **Juan**, **Lény**, **Nadir**, **Nico**, **Nitrate**, **Sébastien**, **Vanessa**, **Vincent**, **Zinzin**. Merci à tous pour toute la joie que vous m'avez communiquée !!

D'autres personnes ont travaillé à me garder les pieds sur terre pendant tout ce temps, ou l'inverse plutôt, mes amis parachutistes : Anne-Laure, Blade, Christophe, Cédric, Dom, Eric, Gaillou, Gwen, Koé, K-ro, L'Morpion, Maya, et tous ceux que j'oublie avec qui j'ai partagé des sauts... un très spécial big up pour Lolo et Christ'aile mes coachs personnels, en jump mais aussi pour le reste ! Merci aussi à Sev, hick-up, pour les heures passées ensemble au téléphone, les nombreuses fois où tu m'as accueillie chez toi et le merveilleux voyage aux Maldives dans lequel tu m'as entraînée.

En dehors de Montpellier, il y a encore quelques personnes que je souhaite remercier, qui ont été présentes pour moi pendant la thèse, mais aussi depuis bien plus longtemps. Tout d'abord je pense à ma bande de copines du collège : **Céline**, **Elisa**, **Emilie**, **Fanny**, **Hélène** et **Karen**. Merci pour votre amitié et les nombreux souvenirs que je ne pourrais lister, même sous la contrainte. Je pense aussi à mes potes de fac pour leurs encouragements et leur soutien, en plus des bons moments partagés : **Hélène**, merci pour ton accueil à Boucan, **Xavier**, pour ton accueil à Clermont et les week-ends de ski et **Yoann** pour notre trip en Pologne. J'ajoute un merci tout particulier à **Arthur** pour son soutien et la confiance qu'il arrive à me redonner.

Je termine presque, avec un énorme merci à **ma famille** pour leur indéfectible soutien depuis de longues années sans lequel je n'en serais jamais arrivée là. Merci à **mes parents** pour m'avoir toujours encouragée à faire ce qui me plaisait sans me poser de questions. Une mention spéciale à ma cousine **Milou** pour tout son amour et son soutien, merci ma puce. Une petite pensée pour la boule de poil, **Kimura**, qui a débarqué sur le parking de Mauguio un beau jour de septembre, en plein pendant la rédaction du chapitre introductif (d'où son nom...) et qui a su m'apporter un amour sans limite dont je lui suis reconnaissante.

Enfin, je dois remercier pour m'avoir accompagnée pendant la rédaction le son de Bonobo, Hightone, JMPZ, Le Peuple de l'Herbe, Mattafafix, Portishead, Vitalic et Wax Tailor, et les magnifiques crépuscules sur le Pic Saint Loup.

Et comme il est bien connu que ce que les gens préfèrent lire dans une thèse ce sont les remerciements, il ne me reste plus qu'à souhaiter une bonne lecture aux courageux qui choisiront d'aller au-delà... à noter que la rédaction de ce manuscrit ainsi que la présentation orale ont été significativement améliorées grâce aux bons soins de Djo, MF, MH, Mumax, Nane, Nath, Albé, Jak, Mathieu, Pierre, Stef et Sylvain. Merci pour votre aide précieuse.

Table des Matières

CHAPITRE 1.

DEMOGRAPHIE ET SYSTEME DE REPRODUCTION : QUEL IMPACT SUR L'EVOLUTION DES GENES ?

1.1.	QU'EST CE QUE LA DIVERSITE GENETIQUE ?	13
	1.1.1. Quel est le niveau de polymorphisme attendu dans les populations ?	14
	1.1.2. Concept d'effectif efficace	15
	1.1.3. L'hypothèse nulle : équilibre neutre mutation-dérive	17
1.2.	EFFETS DE GOULOTS D'ETRANGLEMENT : EXEMPLE DE LA DOMESTICATION	
	1.2.1. Qu'est-ce qu'un goulot d'étranglement ?	
	1.2.2. La domestication, un exemple de goulot d'étranglement récent	20
	1.2.3. La sélection : un goulot d'étranglement plus intense	24
1.3.	EFFETS DE L'AUTOGAMIE ET DES VARIATIONS DE RECOMBINAISON INTRAGENOMIQUE	ES 27
	1.3.1. Impact de l'autofécondation et de la recombinaison sur la diversité génétique	
	1.3.2. Impact de l'autogamie et de la recombinaison sur l'efficacité de la sélection	30
	1.3.3. Effet de l'autogamie et de la recombinaison sur la composition en bases	35
1.4.	LES TRITICEAE: UN COMPLEXE D'ESPECES ADAPTE A L'ETUDE DE L'IMPACT	DE LA
	DOMESTICATION, DE L'AUTOGAMIE ET DE LA RECOMBINAISON SUR L'EVOLUTI	ON DES
	GENES	39
	1.4.1. Plusieurs évènements de domestication	
	1.4.2. Polyploïdie et organisation des génomes	
	1.4.3. Un gradient de recombinaison intra chromosomique	
	1.4.4. Une large variabilité de systèmes de reproduction	
	1.4.5. Ressources genomiques publiques	40
1.5.	OBJECTIFS DE LA THESE	48
CH/	APITRE 2.	
POI	I VMORPHISME DE SEQUENCE ET INFERENCES SUR L'HISTOIRE DE LA	
	MESTICATION DU BLE	
2.1.	PROBLEMATIOUE ET SYNTHESE DES RESULTATS	
2.2.	GRINDING UP WHEAT : A MASSIVE LOSS OF NUCLEOTIDE DIVERSITY SINCE DOMESTICA	tion 56
2.3.	Discussion.	
	2.3.1. Importance de la connaissance de la population sauvage	69
	2.3.2. Un modèle alternatif de détection de sélection	72
CHA	APITRE 3.	
COI	MMENT LE SYSTEME DE REPRODUCTION ET LA RECOMBINAISON AFFEC	TENT
L'E	VOLUTION MOLECULAIRE CHEZ LES TRITICEAE ?	70
3.1.	PROBLEMATIQUE ET SYNTHESE DES RESULTATS	
3.2.	MATING SYSTEM AND RECOMBINATION AFFECT MOLECULAR EVOLUTION IN FOUR IN	RITICEAE
<u></u>	SPECIES	83 00
3.3.	DISCUSSION	
	3.3.1. Comment interpreter les patrons de GC ?	
	2.2.2. Effete de la DCC aux l'affiagaité de la célection	102
	3.3.5. Effets de la BOC sur l'éfficacité de la selection	103 104
	5.5.4. DOC et implications sur l'etuae au regime de reproduction	104
PER	RSPECTIVES	107
BIB	LIOGRAPHIE	115
ANM	NEXES	125
11		

Chapitre 1.

Démographie et Système de Reproduction : quel impact sur l'évolution des gènes ?

Chapitre 1.

Démographie et Système de Reproduction : quel impact sur l'évolution des gènes ?

I am inclined to suspect that we see, at least in some [cases], variations which are of no service to the species, and which consequently have not been seized on and rendered definite by natural selection. Variations neither useful nor injurious would not be affected by natural selection, and would be left either a fluctuating element, as perhaps we see in certain polymorphic species, or would ultimately become fixed... We may easily err in attributing importance to characters, and in believing that they have been developed through natural selection; ... many structures are now of no direct use to their possessors, and may never have been much use to their progenitors.

Darwin, 1872. Origin of Species.

1.1. Qu'est ce que la diversité génétique ?

L'extraordinaire diversité du vivant a depuis toujours piqué la curiosité de l'homme qui a cherché à l'organiser et à la comprendre, en multipliant notamment les classifications au début du XVII^e siècle. Dès l'apparition des pensées remettant en cause la fixité des espèces (fin XVIII^e), les différentes théories de l'évolution se sont évertuées à proposer des mécanismes régissant la structuration de la biodiversité au sein et/ou entre espèces. D'un point de vue évolutif, on peut différencier deux grandes catégories de diversité, dont l'importance relative a donné lieu à de longues controverses : la diversité neutre et la diversité sélectionnée.

L'histoire de cette controverse peut se résumer ainsi. Au début du XX^e, la diffusion des lois de Mendel permet d'identifier les gènes comme unités de base responsables de la variabilité morphologique. Les transformations morphologiques du vivant sont alors essentiellement perçues comme des adaptations, c'est-à-dire comme le résultat de la sélection naturelle sur les gènes. En 1931, Wright suggère que la variabilité morphologique n'est pas le résultat d'une sélection conférant un avantage ou désavantage sélectif à leurs porteurs, mais simplement le fruit du hasard. A cette époque, les désaccords portant sur la part relative du hasard et de l'adaptation dans l'évolution des espèces sont très importants (voir Provine 1986 pour discussion).

Au début des années soixante, les premiers travaux sur le polymorphisme enzymatique voient le jour (Hubby and Lewontin 1966; Lewontin and Hubby 1966). Le débat opposant la diversité neutre à la diversité adaptative se déplace alors de l'échelle morphologique à l'échelle génétique. La découverte de ce polymorphisme pose sous un nouvel angle la question du maintien de la variabilité dans les populations naturelles. L'importance relative de la sélection directionnelle est plus que jamais au cœur des discussions. En 1968, Kimura propose **la théorie neutraliste de l'évolution moléculaire** en arguant que la plupart des mutations touchant les gènes sont sélectivement neutres. Dans le cadre qu'il propose, le devenir de la majorité des variants au sein des populations n'est affecté que par des processus aléatoires. Cette théorie est encore d'actualité de nos jours pour expliquer le maintien de l'essentiel du polymorphisme génétique au sein des populations. Dans ce document, on appellera diversité « neutre » des variations de séquences d'ADN n'ayant pas d'effets directs sur la valeur sélective des individus. Toutefois, comme nous le verrons plus loin, ces variations nucléotidiques ont pu être affectées par des événements sélectifs ayant eu lieu à proximité, dans le génome.

1.1.1. Quel est le niveau de polymorphisme attendu dans les populations ?

Si l'on considère la diversité neutre, le polymorphisme génétique¹ d'une population résulte d'un équilibre dynamique entre la mutation, responsable de l'apparition de nouveaux allèles, et la dérive génétique qui entraîne la fixation aléatoire des allèles dans une population de taille finie. Dans une population de *N* individus diploïdes, Kimura a montré qu'il s'écoule en moyenne 4*N* générations entre l'apparition d'un allèle et sa fixation par dérive (chaque nouvel allèle ayant une probabilité 1/2N de se fixer). Ce temps de 4*N* générations correspond au temps total de coalescence de la population. La coalescence est un processus aléatoire par lequel on remonte au plus récent ancêtre commun d'un ensemble de séquences d'un gène. Ce processus décrit une généalogie produite par l'action de la dérive génétique (Encadré 1). Le polymorphisme attendu dans une population à l'équilibre mutation-dérive est fonction du produit du taux de mutation et du temps de coalescence de la population pendant lequel sont apparues les mutations : $\theta = 4N\mu$, paramètre correspondant au taux de mutation populationnel.

¹ Coexistence dans une population de plusieurs allèles au locus considéré.

A partir de données de séquences, plusieurs estimateurs de θ ont été définis, dont voici les deux principaux :

- la **diversité nucléotidique** π (Tajima 1983). Elle représente le nombre moyen de différences nucléotidiques entre paires de séquences. Le nombre de mutations entre deux séquences tirées au hasard dans la population vaut le produit du taux de mutation μ et du temps de coalescence entre ces séquences (2*N*, Encadré 1) multiplié par 2 (les mutations pouvant se produire sur chacune des deux branches). L'espérance de π est donc 4*N* μ .

- **l'estimateur** θ_S de Watterson (1975). Il est basé sur le nombre de sites polymorphes *S* observés dans un échantillon, dont l'espérance est fonction du nombre de mutations s'étant produites sur la longueur totale de la généalogie des gènes de l'échantillon (4*N*, Encadré 1). Il est obtenu d'après la formule suivante qui standardise *S* vis-à-vis de la taille de l'échantillon :

$$\theta_{S} = \frac{S}{\sum_{i=1}^{n-1} \frac{1}{i}}$$

où *n* est la taille de l'échantillon.

1.1.2. Concept d'effectif efficace

Dans les populations naturelles, tous les individus ne participent pas forcément au processus reproductif, même dans le cas de la neutralité sélective. En général l'effectif de la population N qui détermine le rythme de la dérive génétique n'est donc pas égal à l'effectif démographique de la population. On définit alors l'effectif efficace (ou taille efficace) N_e de la population comme l'effectif d'une population idéale (de type Wright-Fisher, Encadré 1) pour laquelle les fluctuations des fréquences alléliques sont équivalentes à celles de la population étudiée. C'est donc le nombre d'individus d'une population idéale pour lequel on aurait une dérive génétique équivalente à celle de la population réelle. Il existe plusieurs types d'effectifs efficaces, selon l'effet de la dérive que l'on considère : changement du coefficient de consanguinité, de l'hétérozygotie ou de la variance des fréquences alléliques d'une génération à l'autre. On considèrera ici l'effet de la dérive sur la fluctuation des fréquences alléliques au cours du temps, soit N_e la taille d'une population idéale qui aurait les mêmes variances de fréquence allélique que la population étudiée.

Encadré 1. Théorie de la coalescence et simulations

La théorie de la coalescence est une approche rétrospective qui décrit les lois stochastiques régissant les évènements de coalescence (fusion de deux lignées alléliques) (Fig.A). Les probabilités de coalescence des gènes d'un échantillon sont fonction de la taille efficace N_e de la population dont ils sont tirés. On se place dans une population évoluant selon le **modèle Wright-Fisher**. Ce modèle formalise une population panmictique de taille finie, constante au cours de générations non chevauchantes et dont les gènes d'une génération sont issus d'un tirage avec remise parmi les gènes de la génération précédente. Soit P(t) la probabilité que deux gènes d'un échantillon de taille n aient un ancêtre commun exactement t générations dans le passé :





Fig. A. Représentation de la généalogie standard d'un échantillon (n=15) selon un modèle neutre de taille constante.

Pour *n* petit devant N_e , la loi géométrique ci-dessus peut être approximée par une loi exponentielle de même espérance. Soit T_n le temps de coalescence entre deux gènes de l'échantillon: *Tn* suit une loi exponentielle d'espérance $4N_e / (n(n-1))$.

D'où le temps de coalescence entre deux gènes $E(T_2) = 2N_e$

Le nombre de mutations survenant sur une branche de la généalogie suit une loi de Poisson d'espérance μt , le produit du taux de mutation par gène et par génération μ et la longueur de la branche t.

La généalogie de l'échantillon étant sensible aux variations de taille efficace, elle permet de formaliser l'histoire évolutive de la population. La théorie de la coalescence permet de décrire la diversité attendue dans un échantillon selon différents scénarios évolutifs.

Hudson (2002) a développé un programme « ms » qui utilise les propriétés de la théorie de la coalescence pour générer des échantillons de gènes. Le modèle de base consiste en une population Wright-Fisher, mais il est possible de complexifier le modèle en intégrant par exemple des changements démographiques au cours du temps ou de la structure génétique.

Pour chaque échantillon, le programme génère une histoire généalogique, place les mutations aléatoirement sur l'arbre et calcule les statistiques de diversité (π et θ_s).

Puisque les propriétés généalogiques sont régies par des processus aléatoires, une variabilité de généalogies peut être produite sous un seul modèle (Fig.B). La distribution des paramètres de diversité attendus selon le modèle testé peut être définie d'après les échantillons simulés (Fig.C). La comparaison d'un échantillon de séquences empirique aux échantillons simulés permet d'estimer le modèle de fonctionnement de la population étudiée. Par exemple, si la valeur de π estimée dans la population est représentée par la flèche (Fig.C), elle sort de la distribution attendue sous le modèle neutre standard, montrant un excès de diversité par rapport à l'attendu.



Fig. B. Exemples de généalogies générées à partir du modèle neutre standard à l'équilibre.



Afin d'appliquer les modèles de coalescence sur les populations naturelles, on considèrera leur taille efficace et non leur taille démographique. Le polymorphisme attendu dans une population à l'équilibre mutation-dérive s'écrit donc : $\theta = 4N_e \mu$.

1.1.3. L'hypothèse nulle : équilibre neutre mutation-dérive

Les définitions précédentes se situent dans le cadre d'un modèle nul qui se base sur deux hypothèses principales :

- la population est à l'équilibre entre les forces de mutation et de dérive génétique.
- les mutations qui ségrégent dans la population sont sélectivement neutres, *i.e.*, elles n'ont pas d'effet sur la valeur sélective des individus qui les portent.

C'est à ces conditions uniquement que les estimateurs π et θ_S atteignent $4N_e \mu$ et sont égaux entre eux. Le manquement à l'une de ces deux hypothèses se traduit par des écarts entre les valeurs des deux estimateurs π et θ_S . Des évènements démographiques entraînent des changements dans l'intensité de la dérive génétique s'exerçant sur la population, et l'équilibre entre la mutation et la dérive n'est plus conservé. La sélection, par définition, agit sur des mutations non neutres et affecte en conséquence la distribution des fréquences alléliques. De tels évènements (démographiques ou sélectifs) subis par la population modifient la forme de la généalogie des séquences par rapport à une évolution à l'équilibre ou sous neutralité sélective. Rechercher des écarts au modèle neutre standard est une voie pour faire des inférences sur l'histoire évolutive, démographique ou sélective, d'une population.

Tajima (1983) a proposé un des premiers tests d'écart à l'équilibre basé sur la diversité moléculaire d'un échantillon. Il consiste à comparer les deux estimateurs du paramètre : $\theta = 4N_e\mu$ tel que :

$$D = \frac{\pi - \theta_s}{\sqrt{Var(\pi - \theta_s)}}$$

Par construction les estimateurs π et θ_s sont affectés de manière différente par la distribution des polymorphismes dans la population, ils ne se comportent pas de la même manière face à une déviation à l'équilibre de mutation-dérive. La diversité nucléotidique est une moyenne des différences entre paires de séquences, elle est particulièrement sensible aux mutations présentes en fréquence intermédiaire dans la population. Prenons l'exemple d'un site polymorphe avec deux allèles répartis de manière équilibrée entre les individus de la

population (soient les fréquences alléliques telles que p = q = 0.5). La diversité nucléotidique à ce site est 0.5, *i.e.*, on a autant de chances de tirer deux séquences différentes que deux séquences identiques. A l'inverse si ces deux allèles sont en fréquences déséquilibrées dans la population (p=0.9 et q=0.1), π vaut 0.2. Au contraire, θ_S s'estime à partir du nombre de sites polymorphes, peu importe leur fréquence. Ainsi, les singletons² et les mutations en fréquences intermédiaires ont le même poids dans le calcul de θ_S .

Un goulot d'étranglement est un bon exemple d'une transition démographique susceptible de modifier les généalogies d'un échantillon de gènes.

1.2. Effets de goulots d'étranglement : exemple de la domestication

1.2.1. Qu'est-ce qu'un goulot d'étranglement ?

Le modèle de Wright-Fisher considère une population dont la taille reste constante au cours du temps. Or les populations connaissent des changements démographiques au cours de leur histoire : expansion ou réduction. Un goulot d'étranglement se définit comme une réduction de la taille d'une population, ce qui a pour conséquence d'augmenter les fluctuations stochastiques des fréquences alléliques d'une génération à l'autre (augmentation de la dérive génétique). Ce type d'évènement peut être associé à des évènements majeurs tels que des glaciations ou à la colonisation de nouveaux habitats par quelques migrants, représentant souvent un épisode temporaire avant une ré-expansion démographique de la population. Cette réduction de l'effectif efficace N_e de la population, entraîne une baisse du niveau de polymorphisme attendu θ ($4N_{e\mu}$). La Figure 1.1. représente l'évolution des estimateurs de diversité (π et θ_S) au cours du temps lorsqu'une population connaît un goulot d'étranglement suivi d'une ré-expansion démographique. La perte de diversité associée à un goulot d'étranglement dépend de deux paramètres : la durée *d* du goulot et l'intensité α du goulot (définie ici par le rapport de la taille efficace avant goulot sur la taille efficace pendant le goulot).

² site polymorphe pour lequel un des allèles n'est présent qu'en une seule copie dans l'échantillon



Figure 1.1. : Evolution des estimateurs de diversité lors d'un goulot d'étranglement et d'une réexpansion démographique.

La figure représente les résultats de simulations de coalescence modélisant des changements démographiques successifs au cours de l'histoire d'une population. Une espèce ancestrale de taille efficace N_{e0} à l'équilibre subit un goulot d'étranglement au temps t_1 . La taille efficace de la population est réduite d'un facteur α (intensité du goulot) pendant une durée $d = t_1 - t_2$. Au temps t_2 , la population connaît une forte expansion démographique et reprend une taille efficace Ne_1 .

Le niveau de diversité théorique de la population θ ($4N_e\mu$) est en noir, l'estimateur π (Tajima 1983) en gris clair et θ_s (Watterson 1975) en gris foncé. L'évolution du *D* de Tajima (Tajima 1989a) au cours du temps est représentée en dessous, en pointillés.

Le niveau de diversité théorique de la population θ suit l'évolution de N_e si l'on fait l'hypothèse que le taux de mutation reste constant dans le temps. Les estimateurs de diversité π et θ_S réagissent différemment aux changements démographiques. Lors d'un goulot d'étranglement les mutations rares vont plus rapidement être perdues par dérive génétique que les mutations en forte fréquence. On observe alors que θ_S , basé sur le nombre de sites polymorphes, diminue plus rapidement que π qui est particulièrement sensible aux mutations en fréquences intermédiaires. C'est l'inverse lors de l'expansion démographique qui voit l'apparition de nouvelles mutations : le nombre de sites polymorphes augmente plus vite que la fréquence des nouveaux allèles apparus, donc θ_S plus vite que π . L'évolution du test d'équilibre du *D* de Tajima représente l'écart d'estimation de la diversité par π et θ_S au cours du temps. Un scénario démographique comme celui présenté dans la Figure 1.1. illustre comment un goulot d'étranglement peut être détecté au niveau moléculaire. Pendant la durée du goulot (entre t_1 et t_2), les estimateurs π et θ_S indiquent la perte de diversité par rapport à la population ancestrale. Mais une fois la population sortie du goulot, le signal moléculaire s'atténue avec le temps, plus rapidement pour θ_S que pour π . Un goulot d'étranglement récent sera donc plus aisément détectable qu'un évènement ancien, dont les traces seront d'autant plus faibles que le retour à l'équilibre de la population est proche. Il est important de noter que si le taux de mutation du marqueur conditionne la valeur absolue de θ ($4N_e \mu$) et de ses estimateurs, il ne joue pas sur le temps qu'il faut pour atteindre l'équilibre. Ce temps dépend uniquement de la taille efficace de la population.

Soit l'espérance de π au cours du temps (Tajima 1989b) :

$$E(\pi) = \pi_1 + (\pi_0 - \pi_1) e^{-t/2N_{e1}}$$

avec π_0 la valeur de π à l'équilibre avant le changement de taille $(4N_{e0}\mu)$, π_1 la valeur de π à l'équilibre après le changement de taille $(4N_{e1}\mu)$ et *t* le temps en nombre de générations depuis le changement de taille.

Si $t/2N_{el} \ll 1$, c'est à dire au début de l'augmentation, on peut utiliser l'approximation suivante :

$$E(\pi) \approx \pi_1 + \left(\pi_0 - \pi_1\right) \left(1 - \frac{t}{2N_{e1}}\right) = \pi_0 \left(1 - \frac{t}{2N_{e1}}\right) + \pi_1 \frac{t}{2N_{e1}}$$

Or $\pi_1 = 4N_{el}\mu$ et $1 >> t/2N_{el}$, soit l'espérance de π au début de l'augmentation :

$$E(\pi) \approx \pi_0 + 2\mu t$$

1.2.2. La domestication, un exemple de goulot d'étranglement récent

Au Néolithique, lors de sa transition de chasseur-cueilleur à agriculteur, l'homme a commencé à domestiquer certaines des espèces animales ou végétales qu'il consommait. Il est admis que cette transition s'est faite à partir d'un nombre plus ou moins réduits d'individus porteurs de caractères favorables, et que pour les plantes, le ré-ensemencement des champs pouvait se faire à partir d'un nombre réduit de plantes, sélectionnées consciemment ou non,

conservées pour cet usage. Puisqu'un nombre limité de plantes a été utilisé lors de la transition de la population sauvage à la population cultivée, les évènements de domestication représentent des goulots d'étranglement. L'histoire démographique des plantes cultivées peut être schématisée selon le modèle global présenté précédemment (Figure 1.1.). A partir d'une espèce sauvage ancestrale, un échantillon d'individus a constitué le compartiment cultivé, diminuant ainsi la taille efficace : on parle de goulot d'étranglement de domestication (Eyre-Walker et al. 1998). Les plantes cultivées ont ensuite suivi les migrations humaines et ont alors connu une large expansion géographique et démographique. Les premiers mouvements migratoires ont par exemple permis l'expansion de l'amidonnier cultivé (*Triticum turgidum* ssp. *dicoccum*) de son centre de domestication dans le Croissant Fertile à l'ensemble de la région du Levant (Nesbitt and Samuel 1996).

Plusieurs disciplines se sont intéressées au processus de domestication, de l'archéologie à la génétique des populations (revues Salamini et al. 2002; Zeder et al. 2006). Les données archéologiques apportent des estimations indépendantes des données génétiques. Elles ont notamment permis la datation au carbone 14 des premiers grains de plantes cultivées situant les plus anciens épisodes de domestication aux alentours de -10 000 ans (Lev-Yadun, Gopher, and Abbo 2000), l'estimation de la durée de la domestication (plus d'un millénaire chez le blé d'après Tanno and Willcox 2006) ou encore de renseigner sur les sites de domestication (plusieurs sites de domestication pour la courge d'après Smith (Smith 1998).

La Figure 1.2. illustre un goulot d'étranglement suivi d'une ré-expansion sur des temps évolutifs courts, de l'échelle de ce que l'on peut attendre dans le cadre de la domestication pour une espèce telle que le blé. L'estimateur de diversité π augmentant plus lentement que θ_S lors de la ré-expansion, la valeur de π estimée dans les populations cultivées actuelles devrait être plus informative que θ_S quant à la taille efficace de la population pendant le goulot d'étranglement.



Figure 1.2. : Evolution des estimateurs de diversité lors d'un goulot d'étranglement et d'une réexpansion démographique correspondant à un scénario de domestication.

Des échantillons de séquences ont été simulés avec le programme ms (Encadré 1.) selon un scénario démographique correspondant à (*i*) un goulot d'étranglement simulant une réduction de la taille efficace ancestrale d'une intensité $\alpha = 10$ au temps $t_1 = 0.1$ (en unités de 4Ne générations) (*ii*) une durée du goulot d'étranglement d=0.06 (*iii*) suivi d'une expansion démographique au temps t_2 =0.04 résultant en une taille de population actuelle deux fois plus large que la population ancestrale. Voir chap.2 pour plus de détails.

Puisque les goulots d'étranglement de domestication sont des évènements très récents à l'échelle de l'évolution des espèces (environ 10 000 ans), les traces moléculaires du goulot d'étranglement doivent être détectables, malgré l'importante explosion démographique qui a suivi, le retour à l'équilibre ne se faisant pas rapidement. Une érosion de la diversité génétique chez les espèces cultivées par rapport à leurs apparentées sauvages a en effet été mise en évidence sur de nombreuses espèces (revue dans Buckler 2001). La proportion de diversité génétique perdue lors du passage de l'état « sauvage » à l'état « cultivé » dépend théoriquement de la taille efficace de la population pendant le goulot d'étranglement et de la durée de celui-ci (Eyre-Walker et al. 1998). Cette érosion de la diversité qui touche le génome dans son ensemble (Doebley 1989) a été estimée par des isozymes, des microsatellites ou des RFLPs ; en moyenne, elle représente une perte de 30% à 40% par rapport à l'espèce sauvage (Buckler 2001). Ces dernières années ont vu les études visant à estimer la perte de diversité des gènes associée à la domestication se multiplier (Figure 1.3.). La plupart des espèces cultivées étudiées ont montré un niveau de diversité nucléotidique de 30% inférieur à celui mesuré dans l'espèce sauvage apparentée (maïs, luzerne, millet, soja). Cette perte de diversité correspond à celle mesurée dans les études précédentes utilisant d'autres marqueurs (Buckler

2001). Cependant certaines espèces ont montré une érosion de diversité nucléotidique pouvant atteindre 60% chez le tournesol (Liu and Burke 2006) ou l'orge (Caldwell et al. 2006; Kilian et al. 2006).



Figure 1.3. : Illustration de la réduction de diversité nucléotidique liée à la domestication. Le graphe représente la diversité nucléotidique, mesurée sur les sites non codant ou synonymes (π silent), estimée au sein du compartiment sauvage (en rouge) et cultivé (en orange) pour plusieurs espèces (Gaut and Clegg 1993; Tenaillon et al. 2004; Wright et al. 2005; Caldwell et al. 2006 ; Hyten et al. 2006; Kilian et al. 2006; Liu and Burke 2006; Muller et al. 2006; Haudry et al. 2007).

Caractéristiques des goulots d'étranglement de domestication

Puisque la domestication est un évènement évolutif récent, les espèces cultivées n'ont théoriquement pas eu le temps de revenir à équilibre mutation-dérive. Par conséquent, étudier les impacts de la domestication sur le génome nécessite le recours à des outils intégrant les écarts à l'équilibre mutation-dérive. Les séquences d'ADN nous renseignent sur la diversité présente au sein de la population, mais le patron de diversité nous informe également sur les propriétés généalogiques de la population. La théorie de la coalescence (Hudson 1990, voir Encadré 1) permet de reconstruire des généalogies de gènes échantillonnés dans une population selon différents scénarios évolutifs. La comparaison de la diversité de séquences entre les populations théoriques simulées et les populations réelles permet ensuite d'approcher les gammes de scénario les plus vraisemblables pour expliquer le patron de diversité observé (Nordborg 2003). La théorie de la coalescence a permis de caractériser les paramètres de goulots d'étranglement (intensité, durée), notamment chez le maïs (Eyre-Walker et al. 1998; Tenaillon et al. 2004; Wright et al. 2005), la luzerne (Muller et al. 2006) et plus récemment le

blé (Chap. 2), mais aussi chez l'homme (Wakeley and Hey 1997; Stajich and Hahn 2005) et la drosophile (Glinka et al. 2003 ; Ometto et al. 2005).

L'atout principal de l'étude de la domestication des plantes cultivées pour quantifier l'impact d'un goulot d'étranglement sur la diversité des gènes est de disposer de l'espèce sauvage apparentée actuelle. En effet, elle est généralement considérée comme un bon prédicteur de la situation ancestrale de l'espèce avant sa domestication. Sous l'hypothèse que la population sauvage est à l'équilibre mutation-dérive, sa diversité nous permet l'estimation de l'état initial ($4N_{e0}\mu$). De plus l'accès à la population sauvage permet d'identifier des modifications morphologiques liées à la domestication. Ainsi, les grains trouvés sur les sites archéologiques sont attribués à une catégorie sauvage ou cultivée, permettant une estimation de la date de domestication (t_1) indépendante des données génétiques. De manière similaire, les populations africaines de la drosophile peuvent être considérées comme populations de référence par rapport aux populations non africaines afin d'étudier le goulot d'étranglement de *Drosophila melanogaster* lors de la migration de l'Afrique vers l'Europe, il y a 10 000 à 15 000 ans (Glinka et al. 2003; Thornton and Andolfatto 2006).

1.2.3. La sélection : un goulot d'étranglement plus intense

La formation des espèces domestiquées s'est également accompagnée d'un choix (conscient ou non, voir à ce sujet la discussion dans Ross-Ibarra, Morrell, and Gaut 2007) de caractères avantageux pour la culture. En d'autres termes, la domestication d'une espèce correspond à un goulot d'étranglement agissant sur l'ensemble du génome et une sélection plus intense sur certains gènes cibles. Ces deux phénomènes participent chacun à diminuer la diversité présente dans la population cultivée mais font que les gènes n'ont pas tous été affectés de la même manière. Les gènes qui ne sont pas impliqués dans l'expression de phénotypes favorisés lors du passage à l'agriculture (neutres vis-à-vis de la domestication) ont subi une simple réduction d'effectif efficace due à la dérive génétique créée par le goulot d'étranglement de domestication (Figure 1.4.A, partie gauche). Les gènes codant pour des caractères agronomiquement avantageux impliqués dans la domestication (sélectionnés directement par l'homme ou du fait d'une adaptation au système de culture) ont quant à eux

subi une réduction de diversité supplémentaire, puisque étant les cibles directes d'une sélection positive (Figure 1.4.A, partie droite).



Figure 1.4. : Représentation schématique d'un goulot d'étranglement et de son impact au niveau d'un gène neutre et d'un gène sous sélection. Figure adaptée de (Ross-Ibarra, Morrell, and Gaut 2007).

- A. La diversité génétique est représentée par des cercles grisés avant et après la domestication au niveau d'un gène neutre et d'un gène sous sélection lors de la domestication. Le goulot d'étranglement réduit le nombre de génotypes (différents niveaux de gris) dans le compartiment domestiqué par rapport à la population sauvage. Cette réduction de diversité est plus forte pour un gène sous sélection, pour lequel un seul génotype persiste après la domestication. Cependant, un goulot d'étranglement très fort pourrait également anéantir la diversité au niveau d'un gène neutre
- **B.** La diversité est représentée par des haplotypes schématisés avant et après la domestication, au niveau d'un gène neutre et d'un gène sous sélection lors de la domestication. Les traits représentent la séquence d'ADN et les ronds les sites polymorphes (variants) ; les ronds blancs sont neutres, tandis que le rond noir correspond au site directement impliqué dans la sélection du gène. Tandis que plusieurs haplotypes du gène neutre traversent la grille de la domestication, la sélection du site noir entraîne la fixation de cet haplotype uniquement, avec les sites neutres qu'il porte (« selective sweep » ou « balayage sélectif »).

Le niveau de diversité résiduel à chaque locus dans le compartiment cultivé est donc attendu inversement proportionnel à l'importance adaptative du locus lors de la domestication. Les gènes qui ont joué un rôle majeur peuvent avoir perdu toute variabilité génétique dans le compartiment cultivé (Whitt et al. 2002). La sélection que l'homme a exercée sur les plantes lors de la domestication peut s'illustrer comme un goulot d'étranglement qui s'ajoute au goulot d'étranglement démographique dû au simple effet d'échantillonnage. Il est cependant important de noter qu'un site sous une forte pression de sélection tend à se fixer rapidement dans la population. Il entraîne alors avec lui les sites neutres qui lui sont liés génétiquement, réduisant ainsi localement la diversité génétique (Figure 1.4.B, partie droite). Ce processus d'entraînement des gènes neutres lors d'un évènement de sélection, nommé «selective sweep » ou « balayage sélectif » (Maynard Smith and Haigh 1974), dépend du patron de recombinaison local (paragraphe 1.3.1.). Un balayage sélectif des sites sélectionnés lors de la domestication entraîne théoriquement l'érosion de la diversité neutre au niveau de sites et de gènes liés génétiquement.

Cependant ce schéma sélectif n'est pas universel. Les formes sauvages sont adaptées aux conditions climatiques et environnementales associées à leur aire de répartition. La diffusion de l'agriculture jusqu'en Europe il y a entre 5000 et 8500 ans a confronté les espèces domestiquées à des nouveaux environnements, auxquels elles ont dû s'adapter localement. L'expansion démographique post-domestication pourrait avoir favorisé l'apparition et la sélection de nouveaux allèles chez les formes cultivées, créant de la diversité génétique au niveau des sites concernés. La présence d'allèles dérivés en fréquence importante a notamment été observée au niveau de locus de résistance au mildiou chez l'orge (Piffanelli et al. 2004) et le blé (Yahiaoui, Brunner, and Keller 2006). De même des allèles impliqués dans la réponse à la vernalisation et à la photopériode chez l'orge semblent être apparus après la domestication (Cockram et al. 2007). L'approche de sélection directionnelle présentée dans la Figure 1.4. ne s'applique pas à ce type de sélection favorisant le niveau de diversité des formes cultivées, par sélection balancée ou à l'échelle de l'espèce par l'adaptation locale. Les gènes impliqués dans ces deux types de sélection devraient au contraire présenter une intensité de goulot d'étranglement moins forte que les gènes soumis au goulot d'étranglement démographique de la domestication.

En résumé, nous avons vu que la domestication s'est accompagnée d'un goulot d'étranglement réduisant la diversité génétique des espèces cultivées. L'analyse de données de polymorphisme et couplée à l'utilisation de la théorie de la coalescence permettent de faire des inférences sur l'histoire évolutive des espèces.

Un des objectifs de cette thèse consiste à retracer l'histoire de la domestication des blés cultivés et notamment évaluer l'intensité des goulots d'étranglement.

1.3. Effets de l'autogamie et des variations de recombinaison intragénomiques

Le système de reproduction joue un rôle fondamental sur la quantité et la distribution du polymorphisme au sein des populations (Charlesworth and Wright 2001). Parce qu'il contrôle le mode de transmission des gènes d'une génération à la suivante, le système de reproduction peut affecter à la fois les patrons de diversité et de recombinaison. En effet, une transition de régime de reproduction vers l'autogamie s'accompagne d'une rapide augmentation du taux d'homozygotie dans la population, ce qui a deux effets majeurs :

- <u>une réduction de la taille efficace de la population.</u> En diminuant le nombre de gamètes échantillonnés indépendamment pour participer à la génération suivante, l'augmentation de l'homozygotie a pour conséquence de diminuer N_e à taille démographique constante, et donc d'augmenter la dérive génétique. Dans le cas d'autogamie stricte, N_e est réduit de moitié (Pollak 1987). Pour des taux d'autofécondation intermédiaires, $N_e = \frac{N}{1+Fis}$ avec N la taille démographique de la population et *Fis* l'indice de fixation de Wright mesurant le déficit intra-population en hétérozygotes du au régime de reproduction, $Fis = \frac{S}{2-S}$ où S est le taux d'autofécondation. Par ailleurs, les espèces autogames, présentant souvent une forte structure de populations et de faibles taux de migration, sont plus sujettes à des phénomènes d'extinction et recolonisation, ce qui a pour conséquence de réduire d'autant plus la taille efficace de l'espèce (Ingvarsson 2002).

- <u>une diminution du taux de recombinaison efficace</u>. Dans les populations autogames la forte fréquence d'homozygotie diminue le nombre d'événements de recombinaison qui se font en situation hétérozygote, base d'une recombinaison dite « efficace » (Nordborg 2000). On peut définir r_e le taux de recombinaison efficace tel que : $r_e = r(1 - Fis)$, r étant le taux de recombinaison et (1-Fis) mesurant la proportion de la population ne souffrant pas de déficit en hétérozygote. Les crossing-overs se produisant entre des sites homozygotes ne sont pas efficaces génétiquement, *i.e.*, ils ne participent pas à la création de nouveaux haplotypes. La réduction du taux de recombinaison efficace augmente les effets d'entraînements entre les gènes, ce qui a pour conséquence de réduire encore plus la taille efficace au voisinage des gènes sélectionnés (cf. paragraphe suivant).

Au sein même des espèces, il existe un niveau de variation supplémentaire sur l'efficacité de la recombinaison à l'échelle des chromosomes. En effet le taux de recombinaison d'une espèce n'est pas homogène sur tout le génome : des gradients le long des chromosomes et plus localement des pics ou « hot spots » de recombinaison ont été observés chez les Mammifères (Nachman and Churchill 1996; Kong et al. 2002; Jensen-Seaman et al. 2004; Coop and Przeworski 2007), la levure (Gerton et al. 2000), *Arabidopsis thaliana* (Kim et al. 2007).

Les conséquences de l'autofécondation et d'une recombinaison réduite sont similaires sur un certain nombre de caractéristiques génomiques. Dans les sections suivantes, les effets de l'autofécondation et de la recombinaison sur la distribution de la diversité dans les génomes seront traités parallèlement et illustrés par les comparaisons (*i*) entre espèces autogames et allogames (*ii*) entre régions génomiques présentant différents niveaux de recombinaison.

1.3.1. Impact de l'autofécondation et de la recombinaison sur la diversité génétique

La réduction de N_e associée à une transition vers l'autogamie devrait se traduire par une réduction du polymorphisme au sein des populations. En effet, nous avons vu dans la section 1.1. que le polymorphisme attendu dans une population était fonction de la taille efficace de la population ($\theta = 4N_e\mu$ dans une population panmictique à l'équilibre mutationdérive).

La réduction de la recombinaison efficace chez les autogames participe également à diminuer localement l'effectif efficace, et donc la diversité, dans le génome des populations autogames. En effet, la recombinaison influence la distribution de la variation génétique localement par des effets d'entraînement de la diversité neutre liée aux sites sous sélection (Hill and Robertson 1966). On différencie trois catégories d'effets d'entraînement :

 les balayages sélectifs lors de sélection directionnelle sur des mutations avantageuses (Maynard Smith and Haigh 1974; voir revue, Braverman et al. 1995).

- la sélection d'arrière-plan contre les mutations délétères (Charlesworth, Morgan, and Charlesworth 1993).
- les interactions Hill-Robertson entre mutations faiblement délétères co-ségrégeant dans une population (McVean and Charlesworth 2000).

La quantité de diversité neutre perdue et la longueur du fragment chromosomique touchée par ces effets d'entraînement sont fortement dépendantes du taux de recombinaison local. Un locus neutre perdra moins de diversité sous la pression de sélection s'exerçant sur ses voisins si le taux de recombinaison local est fort et vient briser le déséquilibre de liaison³ créé par la sélection ou la dérive entre les sites. Par conséquent une corrélation positive est attendue entre le niveau de polymorphisme et le taux de recombinaison local.

L'effet théorique de l'autogamie sur la diversité génétique est globalement confirmé par les données empiriques. De nombreuses études se sont intéressées à la comparaison de la variabilité génétique observée chez des espèces autogames et allogames. La Figure 1.5. présente les niveaux de diversité génétique estimés par différents types de marqueurs moléculaires (allozymes, RAPD, microsatellites et séquences) dans de nombreuses espèces (Hamrick and Godt 1996; Nybom 2004; Glémin, Bazin, and Charlesworth 2006). Les taux d'hétérozygotie attendue (hétérozygotie de Nei) sont significativement supérieurs dans les populations allogames par rapport aux autogames (p < 5%, Figure 1.5.). Ces résultats montrent des niveaux de diversité intermédiaires pour les espèces présentant un régime de reproduction mixte. La réduction de la diversité génétique attendue dans les populations pratiquant l'autofécondation est confirmée par les données et participe en partie à la variation des niveaux de diversité observée entre espèces ($R^2 = 6-8\%$, Glémin, Bazin, and Charlesworth 2006).

Figure 1.5. : Diversité génétique en fonction du système de reproduction.

L'histogramme représente les niveaux de diversité génétique estimés au sein de populations allogames (bleu foncé), autogames (bleu clair) et de régime de reproduction intermédiaire (bleu turquoise).

Les estimations sont issues de différents types de marqueurs moléculaires : allozymes (Hamrick and Godt 1990), séquences nucléotidiques Glemin, Bazin, and Charlesworth 2006), RAPD et microsatellites (Nybom 2004).



³ Association non aléatoire d'allèles appartenant à des locus différents

A l'échelle intra-génomique, il est possible de contraster la diversité mesurée dans des régions fortement et faiblement recombinantes. Une corrélation positive entre la diversité génétique et différents estimateurs du taux de recombinaison a été mise en évidence empiriquement chez de nombreuses espèces animales : drosophile (Kaplan, Hudson, and Langley 1989; Begun and Aquadro 1992), homme (Nachman et al. 1998; Nachman 2001), souris domestique (Nachman 1997) ou végétales : tomate (Stephan and Langley 1998; Roselius, Stephan, and Stadler 2005), betterave maritime (Kraft et al. 1998), Aegilops (Dvorak, Luo, and Yang 1998). Cependant, il semblerait que généralement seules les régions avec un taux de recombinaison très faible montrent une réduction de diversité détectable (Hudson and Kaplan 1995; Charlesworth 1996; Innan and Nordborg 2003; Wright et al. 2006). Chez certains organismes la recombinaison apparaît corrélée à la fois aux niveaux de polymorphisme et de divergence, ne permettant pas de distinguer les effets de la sélection et de la mutation (Wright et al. 2006). La corrélation positive entre la diversité nucléotidique et le taux de recombinaison inféré d'après le patron de déséquilibre de liaison observée chez le maïs et Arabidopsis lyrata pourrait être due à des événements démographiques et de balayages sélectifs (Tenaillon et al. 2002; Wright et al. 2005; Wright et al. 2006).

1.3.2. Impact de l'autogamie et de la recombinaison sur l'efficacité de la sélection

L'efficacité de la sélection agissant au sein des populations dépend du produit de N_e et du coefficient de sélection *s* de l'allèle muté (où s est le coefficient de sélection d'un allèle mutant par rapport à l'allèle sauvage, Kimura 1983). Une diminution de N_e augmentant l'effet de la dérive par rapport à la sélection devrait par conséquent entraîner l'augmentation de la probabilité de fixation des mutations faiblement délétères (*s* petit). Réciproquement, des mutations faiblement avantageuses ont une probabilité de fixation attendue plus faible dans les populations de taille efficace réduite, où elles sont plus facilement perdues par dérive. D'après la théorie presque neutre de l'évolution, la plupart des mutations étant faiblement délétères, des différences de taille efficace, se traduisant par des différences d'efficacité de la sélection, devraient jouer sur l'évolution des génomes essentiellement par l'accumulation de mutations délétères (Ohta 1992). Il est possible d'étudier l'évolution d'un gène et d'évaluer les contraintes sélectives qui agissent sur lui au travers des patrons de divergence synonyme

Encadré 2. Evolution des codons et force de sélection.

Selon la fonction de la protéine pour laquelle ils codent, les gènes peuvent être soumis à différents types de contraintes sélectives. Ici on dissociera notamment les effets de la sélection positive (ou adaptative) des effets de la sélection purificatrice sur l'évolution des gènes. Les mutations qui apparaissent sur la séquence codante peuvent être classées en deux catégories selon leur impact sur la séquence protéique traduite :

- les mutations « synonymes » sont celles qui n'entraînent pas de changement dans la protéine du fait de la redondance du code génétique (*i.e.*, le fait que le même acide aminé est codé par plusieurs codons différents). Puisqu'elles n'ont pas d'effet sur la protéine, les mutations synonymes sont généralement considérées comme neutres. Notons que ce n'est pas toujours le cas (Chamary, Parmley, and Hurst 2006, pour revue), des codons préférés ayant pu être identifiés dans de nombreuses espèces, définissant ainsi le biais d'usage de codon.
- les mutations « non synonymes » qui elles, modifient la séquence protéique. On distingue alors différents effets de ces mutations sur l'individu qui les porte. Elles peuvent être neutres (le changement d'acide aminé n'a pas d'impact), délétères (le changement d'acide aminé confère un désavantage sélectif) ou avantageuses (le changement d'acide aminé confère un avantage sélectif).

Deux taux de substitution correspondant à ces deux catégories de mutation sont alors définis : d_s le taux de substitution synonyme et d_N le taux de substitution non synonyme. Le rapport entre les taux de substitution non

synonymes sur synonyme, $\omega = \frac{d_N}{d_S}$, peut alors prendre différentes valeurs en fonction des pressions de

sélection exercées sur le gène :

- $\omega = 1$ dans le cas ou aucune pression sélective ne s'exerce (exemple d'un pseudogène), les mutations non synonymes sont neutres.
- $\omega < 1$ dans le cas où le gène subit une pression purificatrice. Les mutations non synonymes sont délétères, la sélection tend à les éliminer.
- ω >1 dans le cas où le gène subit une pression de sélection positive. Les mutations non synonymes sont avantageuses, leur fixation est favorisée par la sélection.

CODEML est un programme développé par Yang (Yang 1997) qui permet de modéliser l'évolution des codons le long d'une phylogénie (Goldman and Yang 1994). De nombreux modèles permettent d'estimer les ω sur une phylogénie par maximum de vraisemblance (Yang and Bielawski 2000, pour revue). Il est notamment possible de contraster différentes valeurs de ω en fonction (A) des branches d'un arbre phylogénétique et/ou (B) des différentes positions (codons) d'une séquence.



(d_S , vitesse d'évolution considérée comme neutre, reflétant l'interaction entre les forces de mutation et de dérive) et non synonyme (d_N , vitesse d'évolution de la protéine, soumis à la force de la sélection ; Encadré 2). Ainsi, une réduction de l'efficacité de la sélection contre les mutations délétères doit se solder par une augmentation de la fixation de mutations non synonymes, relativement au taux de mutations synonymes, soit une augmentation du ratio d_N/d_S . Conformément aux prédictions du modèle d'évolution presque neutre, l'effet de la taille efficace des populations sur le taux d'évolution des protéines a été mis en évidence chez les Mammifères (Keightley and Eyre-Walker 2000), la drosophile (DeSalle and Templeton 1988) et les Oiseaux (Johnson and Seger 2001).

En abaissant fortement N_e , l'autogamie devrait donc affecter l'évolution des génomes des organismes autogames, en réduisant l'efficacité de la sélection naturelle (Charlesworth and Wright 2001), permettant l'accumulation de mutations faiblement délétères et limitant la fixation d'allèles avantageux (Charlesworth 1994). Cependant la probabilité de fixation de mutations avantageuses partiellement récessives est plus grande dans les populations autogames du fait de l'augmentation des taux d'homozygotie (Charlesworth 1992). Les différences de patron de substitutions adaptatives attendues entre autogames et allogames sont donc dépendantes de la dominance des allèles. La probabilité de fixation des mutations délétères est, elle, peu sensible aux effets de dominance (Charlesworth 1994) et majoritairement dépendante de N_e . Cependant, pour des mutations codominantes, une réduction de N_e de plus de moitié chez les autogames est nécessaire à une augmentation de la probabilité de fixation des allèles délétères (Charlesworth 1992; Glémin 2007). On s'attend donc à observer une réduction de l'efficacité de la sélection chez les autogames si les effets démographiques et d'entraînement accentuent la réduction de N_e due à l'échantillonnage des gamètes.

En plus des variations du taux de recombinaison à l'échelle intra-génomique, la réduction de recombinaison efficace chez les autogames accentue la réduction de N_e localement par les effets d'entraînement (Hill and Robertson 1966) et affecte l'efficacité de la sélection indirectement (Encadré 3). En créant de nouvelles combinaisons haplotypiques, la recombinaison permet d'associer plusieurs mutations favorables dans le même individu et leur fixation conjointe dans la population par la sélection naturelle. L'apparition aléatoire des mutations dans des populations de taille finie créé du déséquilibre de liaison entre les allèles (Ohta and Kimura 1969). Prenons l'exemple de deux mutations avantageuses apparaissant à deux locus différents, sur le même chromosome, chez deux individus distincts (Encadré 3). Si



Représentation schématique des modèles de sélection avec et sans recombinaison. Figure adaptée de Marais (2003).

A et B sont des sites soumis à sélection, C est un site neutre avec deux allèles C_1 et C_2 (en vert et en jaune) qui ségrégent dans la population. Les deux haplotypes $A_1B_1C_1$ et $A_1B_1C_2$ sont présents dans la population. Deux nouveaux allèles apparaissent par mutation aux sites sous sélection A et B dans deux individus séparés. Les mutations A_2 et B_2 (en rouge) confèrent un avantage sélectif à l'individu qui les porte (+10% et +0.1% respectivement). Il y a alors quatre haplotypes qui ségrégent dans la population : $A_1B_1C_1$, $A_1B_1C_2$, $A_2B_1C_1$ et $A_1B_2C_2$. La sélection va favoriser les individus porteurs des mutations avantageuses.

<u>En absence de recombinaison</u> (à gauche), les allèles A_2 et B_2 ne peuvent se retrouver dans le même haplotype, ils vont rentrer en compétition. Puisque l'allèle A_2 apporte un avantage sélectif (+10%) beaucoup plus fort que l'allèle B_2 (+0.1%), c'est l'allèle A_2 qui va se fixer dans la population, entraînant dans sa course folle (par balayage sélectif) l'haplotype $A_2B_1C_1$. On notera que la fixation de l'allèle favorable A_2 s'accompagne de la fixation de l'allèle faiblement délétère B_1 .

<u>Avec de la recombinaison</u> (à droite), des crossing-over peuvent briser les associations d'allèles résultant de l'apparition aléatoire des mutations. De nouveaux haplotypes sont ainsi formés, notamment des combinaisons associant les allèles avantageux : $A_2B_2C_1$ et $A_2B_2C_2$. La sélection est plus efficace qu'en absence de recombinaison, il n'y a pas de compétition entre les allèles favorables. La recombinaison permet le maintien de polymorphisme au site neutre.

La recombinaison va ainsi affecter la trace génomique laissée par la sélection. En l'absence de recombinaison entre l'allèle favorable A_2 qui va se fixer dans la population et un site neutre lié C_1 (à gauche), le balayage sélectif est « complet ». Un seul haplotype se fixe lors de la sélection, puis de nouveaux allèles vont apparaître par mutation (en faible fréquence), ce qui se traduira par un D de Tajima négatif. A l'inverse, si un évènement de recombinaison brise l'association entre les allèles A_2 et C_1 (à droite), la sélection va également entraîner des allèles C_2 . Les haplotypes $A_2B_2C_1$ et $A_2B_2C_2$ présenteront des fréquences intermédiaires, se traduisant par un D de Tajima positif.

On peut décliner ce modèle selon différents scénarios sélectifs :

- Sélection purificatrice : les mutations qui apparaissent (A₂ et B₂) seront délétères dans ce cas (-10% et -0.1% respectivement). L'allèle A₂ sera fortement contre-sélectionné, entraînant la perte de l'haplotype A₂B₁C₁ et donc de l'allèle avantageux B₁ en l'absence de recombinaison.
- <u>Effet Hill-Robertson</u> : les prédictions sont les mêmes. Mais dans ce cas, l'allèle A_2 sera soumis à faible sélection, avec peu de différences entre les haplotypes $A_2B_1C_1$ et $A_1B_2C_2$. En l'absence de recombinaison, l'une des combinaisons sera fixée (ou éliminée) aléatoirement, par dérive.

un évènement de recombinaison survient entre les deux locus, les allèles favorables se retrouvent dans le même fond génétique, leur fixation n'en sera que plus rapide et le déséquilibre de liaison est réduit par la recombinaison. A l'inverse, en l'absence d'évènement de recombinaison entre les deux locus, les deux allèles favorables ne pourront jamais se retrouver dans le même haplotype ni être fixées simultanément. Dans les régions fortement recombinantes, les mutations avantageuses vont se fixer rapidement tandis que les mutations délétères vont être purgées efficacement. A l'inverse, un taux de recombinaison faible rend la sélection positive et purificatrice moins efficaces, et cette situation conduit à la fixation d'allèles faiblement délétères dans la population. La probabilité de fixation d'une mutation favorable ne dépend donc pas uniquement de l'avantage qu'elle confère à l'individu qui la porte, mais également de son environnement génomique (taux de recombinaison local et pression de sélection aux locus voisins).

Quand la théorie se confronte aux données...

Nous avons vu qu'une réduction de l'efficacité de la sélection est attendue à la fois dans les espèces asexuées qui ne recombinent pas, dans les espèces autogames pour lesquelles la recombinaison efficace est réduite et dans les régions du génome non ou faiblement recombinantes.

Les prédictions concernant l'effet du système de reproduction sur l'efficacité de la sélection ne sont pas toujours soutenues par les données. De récentes études ont cherché à mettre en évidence une réduction de l'efficacité de la sélection naturelle chez les espèces asexuées (voir pour revue, de Visser and Elena 2007).

Une accumulation de mutations délétères a été mise en évidence dans les mitochondries des lignées de daphnies asexuées par rapport à des lignées sexuées (Paland and Lynch 2006). Dans une étude de l'impact du système de reproduction sur les patrons de diversité génétique des plantes à fleurs, Glémin et al. (2006) ont trouvé un signal faible de réduction de l'efficacité de la sélection positive et purificatrice chez les espèces autogames par rapport aux allogames. A l'inverse, l'étude du couple d'espèces apparentées *Arabidopsis lyrata* (allogame) et *A. thaliana* (fortement autogame) n'a révélé aucun signal d'effet du système de reproduction sur l'efficacité de la sélection (Wright, Lauga, and Charlesworth 2002). Des développements théoriques récents montrent que le relâchement de la sélection sur
les espèces autogames devrait pourtant être détectable, avec un signal plus fort sur le patron de substitution entre espèces que sur le patron de polymorphisme (Glémin 2007).

Quelques études se sont également intéressées au potentiel adaptatif des espèces en fonction de leur système de reproduction, notamment chez la levure *Saccharomyces cerevisiae*. Les résultats de ces études sont variables mais tendent à montrer une plus grande efficacité de la sélection chez les espèces allofécondantes (de Visser and Elena 2007). Elles se sont cependant limitées à l'étude de peu d'espèces, toutes unicellulaires.

En contrastant des régions génomiques fortement et faiblement recombinantes, Haddrill et al. (2007) ont examiné l'impact du taux de recombinaison local sur les patrons de substitution entre *Drosophila melanogaster* et *D. yakuba*. Tous les patrons observés sont concordants avec une réduction de l'efficacité de la sélection en absence de recombinaison, se traduisant par une accumulation de mutations non synonymes. Cependant, ces observations se limitent aux régions non recombinantes, les régions avec un taux de recombinaison faible, intermédiaire ou fort ont des patrons de substitutions similaires. Seules les régions avec un très faible taux de recombinaison pourraient subir un relâchement de l'efficacité de la sélection.

L'absence d'un signal clair de l'impact de l'autofécondation sur l'efficacité de la sélection dans les données pourrait donc s'expliquer par le maintien d'un taux de recombinaison minimal suffisant à maintenir une sélection efficace.

1.3.3. Effet de l'autogamie et de la recombinaison sur la composition en bases

L'autogamie devrait également réduire le phénomène de conversion génique biaisée vers GC (BGC) *via* la diminution d'évènements de recombinaison en situation hétérozygote. La conversion génique est un mécanisme moléculaire associé à la recombinaison entraînant la copie de gènes entre fragments homologues (Lamb 1984). Les événements de conversion génique au niveau de sites hétérozygotes entraînent la création de mésappariements, dont la réparation est biaisée vers GC (Galtier et al. 2001). Le modèle de conversion génique biaisée vers GC prédit que la teneur en GC doit augmenter localement avec le taux de recombinaison (revue Marais 2003). C'est un processus neutre puisqu'il n'a pas d'effet sur la valeur sélective des individus ; cependant la signature moléculaire associée à la BGC est de type sélectif, du

fait de l'avantage qu'il confère aux allèles G et C. Prenons l'exemple d'un site hétérozygote à deux allèles A et G, ne conférant l'un l'autre aucun avantage sélectif à l'individu qui les porte. Lorsque ce site est impliqué dans un évènement de conversion génique, la réparation biaisée du mésappariement favorise l'allèle G par rapport à l'allèle A. La probabilité de fixation des allèles ne dépend donc pas uniquement de la force de la dérive : la BGC augmente les probabilités de fixation des allèles G et C et diminue celles des allèles A et T.

C'est parce que la dynamique de la BGC mime les effets de la sélection qu'il est souvent difficile de dissocier l'impact de ces deux processus sur l'évolution des gènes.

Une corrélation positive a été mise en évidence entre la teneur en bases G et C et le taux de recombinaison local chez la levure, les mammifères, la drosophile et *C. elegans* (Eyre-Walker and Hurst 2001; Marais, Mouchiroud, and Duret 2001; Yu et al. 2001; Birdsell 2002; Marais 2003). Meunier et Duret (2004) ont mis en évidence une forte corrélation entre la teneur d'équilibre en GC et le taux de recombinaison dans le génome de l'homme, suggérant que la dynamique d'enrichissement en GC dépend du patron local de recombinaison.

Chez les plantes, la teneur en GC distingue fortement les Monocotylédones des Dicotylédones, les premières apparaissant globalement plus riches en GC que les dernières (Carels and Bernardi 2000). De plus, et là encore en fort contraste avec les Dicotylédones, chez le riz et d'autres Graminacées s'observe un gradient décroissant pour la teneur en GC à la troisième position des codons (GC3) à partir de la partie non transcrite en 5' vers la partie codante en 3' (Wong et al. 2002). Typiquement cette différence peut aller jusqu'à 25% de différence entre la partie 5' et la partie 3' des gènes. Cette décroissance en GC3 s'accompagne d'un gradient identique dans le biais d'usage du code. Pour expliquer un tel gradient dans les gènes de Monocotylédones, deux mécanismes, non mutuellement exclusifs peuvent être évoqués : une sélection sur les codons favoris et imaginer que cette sélection a de plus forts coefficients dans la partie amont du gène (pour démarrer correctement une traduction par exemple) ou bien une concentration des effets de conversion génique biaisée vers GC dans la partie 5' par une accumulation dans cette même zone des évènements de recombinaison. Le fait que les introns montrent également ce type de gradient (Wong et al. 2002) et que les espèces allogames aient des gradients de GC plus marqués que les autogames (Akhunov et al. 2003a) suggère une distribution non homogène de la recombinaison sur les gènes. Toutefois, il est à noter que les allogames pourraient également sélectionner les codons favorables plus efficacement. Ce phénomène ne se produirait pas chez les Dicotylédones. Enfin, une plus grande intensité de recombinaison en 5' permettrait aussi d'améliorer l'efficacité de la sélection dans cette zone, renforçant la probabilité de fixer les codons favorables par rapport à la zone 3' (voir discussion sur les effets d'entraînement).

Sous l'hypothèse que l'intensité de BGC est positivement corrélée au taux de crossingover, le modèle de BGC pourrait expliquer les corrélations observées entre recombinaison et teneur en GC dans les génomes. Dans ce cas, une transition de système de reproduction vers l'autogamie devrait également entraîner une réduction de l'efficacité de la BGC.

Peut-on détecter un effet du système de reproduction sur la teneur en GC des génomes ?

Wright et al. (2007) ont comparé la composition en bases de deux espèces allogames *Arabidopsis lyrata* et *Brassica oleraceae* et d'une espèce autogame apparentée, *A. thaliana*. Les auteurs ont montré que les pressions de sélection sur l'usage du code sont constantes entre espèces ; la composition en base au sein de la famille des *Brassicaceae* serait donc la résultante d'un processus neutre, tel que la BGC ou un biais mutationnel.

Marais et al. (2004) ont cherché à mettre en évidence l'effet de l'autofécondation sur l'évolution de la composition en bases. Ils ont intégré le taux d'autofécondation au modèle de BGC (Nagylaki 1983), formalisé comme un biais méiotique agissant uniquement aux sites hétérozygotes. Leurs démonstrations théoriques, en accord avec les observations des données de séquences sur *A. thaliana*, montrent que la BGC est inefficace dans les espèces fortement autogames en comparaison avec des espèces allogames. En effet, chez *A. thaliana*, aucune corrélation entre recombinaison et teneur en GC n'a été mise en évidence (Marais, Charlesworth, and Wright 2004). Des différences d'intensités de BGC pourraient expliquer les variations de teneur en GC observées au sein de la famille des *Triticeae*. En effet, les espèces autogames de la famille montrent des teneurs en GC plus faibles que les allogames (Akhunov et al. 2003a; Glémin, Bazin, and Charlesworth 2006).

L'étude au sein d'un complexe d'espèces proches, couvrant des régimes de reproduction contrastés, d'un ensemble de gènes orthologues échantillonnés dans des zones génomiques *a priori* connues pour avoir des intensité de recombinaison variable permettrait

d'avancer sur ces questions. La Figure 1.6. résume les effets de l'autofécondation et de la recombinaison présentés ci-dessus.



Figure 1.6.: Représentation schématique des prédictions théoriques concernant l'interaction de l'autofécondation et de la recombinaison sur le polymorphisme, l'efficacité de la sélection et la composition en bases des gènes.

Le deuxième objectif de cette thèse a été d'étudier l'impact du régime de reproduction et des variations du taux de recombinaison local sur l'efficacité de la sélection et le patron d'évolution de la teneur en GC chez les Triticeae.

1.4. Les *Triticeae* : un complexe d'espèces adapté à l'étude de l'impact de la domestication, de l'autogamie et de la recombinaison sur l'évolution des gènes

Etudier les complexes d'espèces sauvage-cultivé permet d'approfondir les connaissances fondamentales sur l'évolution des plantes (Darwin 1883) car leur contexte historique nous permet de mieux comprendre la manière dont s'établissent les pressions de sélection (direction, intensité). La comparaison sauvage/cultivé illustre comment l'interaction de différents facteurs évolutifs agit sur leur évolution. Identifier des sources de diversité et comprendre l'impact du système de reproduction et de la démographie sur l'évolution des plantes cultivées sont également des étapes indispensables pour la gestion des ressources génétiques (Frankel, Brown, and Burdon 1995). D'un point de vue agronomique, les *Triticeae* représentent un complexe d'espèces particulièrement stratégique. En effet, cette tribu appartenant à la famille des *Poaceae* comprend les principales espèces domestiquées dans la région du Croissant Fertile que sont le blé, l'orge et le seigle ainsi que des herbes fourragères. Ces trois espèces représentent la majeure partie de la consommation mondiale en céréales, avec en 2005 une production mondiale sur 273 millions d'ha, dont 213 de blé, soit une production plus importante que le riz (Source FAO http://faostat.fao.org/).

L'enjeu économique lié à l'étude de ces espèces cultivées a encouragé un fort développement d'outils moléculaires couplé à une longue tradition de l'étude de leur histoire évolutive.

1.4.1. Plusieurs évènements de domestication

L'histoire de la domestication au sein des *Triticeae* est assez bien documentée grâce aux études combinées de génétique et d'archéologie (voir la revue de Salamini et al. 2002). L'aire de répartition des *Triticeae* sauvages se situe au Moyen-Orient, un des principaux centres de domestication avec l'Amérique Centrale et l'Asie. La domestication implique des formes sauvages de blés (*Triticum monococcum* et *T. turgidum dicoccoides*), d'orge



Figure 1.7.: Représentation schématique de l'histoire évolutive des principales espèces cultivées de Triticeae. Adaptée de (Chantret et al. 2005).

Les espèces sauvages et cultivées sont représentées respectivement dans des cercles et des cadres. Les espèces ancestrales ou inconnues sont entourées en pointillés, les autres ont des représentants contemporains. Les estimations de dates de divergence, de polyploïdisation et de domestication sont issues de (Feldman et al., 1995; Nesbitt and Samuel, 1995; (Allaby, Banerjee, and Brown 1999; Eckardt 2001; Huang et al. 2002a; Salamini et al. 2002).

(Hordeum spontaneum) et de seigle (Secale vavilovii). Parmi les formes cultivées, l'histoire évolutive des blés (Triticum sp.) est la plus complexe. Elle comprend plusieurs épisodes de domestication et un épisode de polyploïdisation (Figure 1.7.). La première espèce de blé a avoir été domestiquée est l'engrain, il y a plus de 10 000 ans, l'espèce diploïde T. *monococcum* (génome A^mA^m); sa culture a été délaissée à partir de l'âge de bronze (Salamini et al. 2002). Les blés polyploïdes domestiqués ont connu une culture plus intensive, probablement du fait de l'apparition de grains nus facilitant l'agriculture. Les blés tétraploïdes (dont T. turgidum ssp. dicoccum et T. t. ssp. durum) ont été domestiqués à partir de l'amidonnier sauvage T. t. ssp. dicoccoides (génome AABB), issu d'une allopolyploïdisation entre T. urartu et une espèce Aegilops porteuse du génome B il y a 0.25-1.3 Ma (Figure 1.7.). Une espèce très proche morphologiquement de T. t. ssp. dicoccoides, T. timopheevii (AAGG), semble issue d'un événement de polyploïdisation indépendant, impliquant des diploïdes apparentés proches de T. urartu et d'Aegilops speltoides (Dvorak and Zhang 1990; Miyashita, Mori, and Tsunewaki 1994; Mori, Liu, and Tsunewaki 1995; Kilian et al. 2007; Cenci et al. in prep.). Le blé tendre T. aestivum (génome AABBDD) est également le fruit d'une allopolyploïdisation, entre une sous-espèce tétraploïde Triticum turgidum avec génomes AABB et l'espèce diploïde Ae. tauschii (génome DD). L'histoire de la domestication des blés tétraploïdes et hexaploïde est détaillée dans le Chapitre 2.

1.4.2. Polyploïdie et organisation des génomes

Le nombre chromosomique de base des *Triticeae* est 2x=2n=14 bâti sur 7 paires de chromosomes dont l'homéologie est relativement bien conservée entre les espèces. Des espèces allo-polyploïdes sont fréquemment observées avec des niveaux de ploïdie allant de 4 à 6 avec des espèces allotétraploides (2n=4X=28) ou bien hexaploïdes (2n=6X=42). Ainsi, parmi les 31 espèces des genres *Triticum* et *Aegilops*, 18 espèces sont allopolyploïdes (Huang et al. 2002b). L'hérédité est disomique chez les allopolyploïdes, les appariements entre chromosomes homéologues étant réprimés. Chez le blé, ce contrôle de l'appariement est sous contrôle génétique (Griffiths et al. 2006).

Les génomes constitutifs des polyploïdes étant proches et ayant peu divergé, la redondance génétique due à la présence de copies homéologues au sein du même individu rend difficile l'étude du polymorphisme populationnel à un locus donné. Pour parvenir à étudier le polymorphisme de façon non confuse tout en pratiquant un séquençage direct par

PCR, il convient donc de dessiner des amorces permettant l'amplification spécifique d'un locus, sur une seule copie (un seul des locus homéologues) (Ravel et al. 2006). Mais la redondance peut présenter des avantages pour l'étude des espèces polyploïdes. Par exemple, le caractère allopolyploïde du blé tendre (6X) lui permet de supporter la disparition de chromosomes entiers et cela a permis la création de séries de lignées viables pour lesquelles une partie connue du génome est manquante. Ainsi, les lignées nullisomiques sont délestées d'une paire de chromosomes (2n=40) tandis que les lignées ditélosomiques ont perdu un bras de chromosome (Sears and Sears 1978).

Ces lignées ont été largement utilisées pour l'assignation de caractères, de marqueurs et de gènes aux différents chromosomes des blés. Plus récemment une série de lignées de délétion a été produite, pour lesquelles de plus petits fragments de chromosomes sont manquants (« bins de délétion »), permettant l'affinement de la localisation des caractères (Endo and Gill 1996; Qi et al. 2003). Au total, 436 lignées de délétions couvrant la totalité du génome ont été identifiées par C banding (Endo and Gill 1996); environ 1/3 de ces lignées sont viables. Celles-ci permettent l'établissement de véritables cartes physiques des chromosomes, bien que la résolution ne reste pas très élevée et que certaines zones ne peuvent être prises en compte (certaines délétions ne pouvant être fixées). Ainsi, à l'aide d'une série de 101 lignées de délétion, Sorrells et al. (2003) ont comparé la localisation chromosomique de 4485 ESTs (expressed sequence tags) entre blé et riz (Figure 1.8.). Bien que la radiation des Poaceae ait été estimée à plus de 65 Ma, on a observé une bonne conservation dans l'ordre des gènes entre le blé et le riz au niveau de grands fragments chromosomiques. Tandis que certains chromosomes comme le chromosome 4A semblent avoir subi plusieurs réarrangements chromosomiques chez le blé tendre par rapport au riz, le chromosome 3B pour lequel la synténie avec le chromosome 1 du riz est très largement conservée. Si des délétions ou pertes de colinéarité peuvent exister à une échelle plus fine entre les chromosomes 3B et 1 du riz, le riz apparaît comme un bon prédicteur de la localisation des gènes sur le chromosome 3B du blé tendre (Chap. 3 pour plus de détails).



Figure 1.8.: Relations entre les génomes du blé et du riz.

Les chromosomes du blé tendre sont représentés à gauche. A droite, chaque bin de délétion est représenté et coloré en accord avec le chromosome du riz pour lequel la meilleure homologie de séquences des ESTs a été trouvée. Figure extraite de Sorrells et al. (2003).

1.4.3. Un gradient de recombinaison intra chromosomique

L'assignation physique de marqueurs dans les bins de délétions du blé tendre a également permis l'étude de l'organisation du génome du blé le long des chromosomes. Grâce à la cartographie d'ESTs, Akhunov et al. (2003b) ont mis en évidence que la densité en gènes et le taux de recombinaison augmentent avec la distance relative au centromère du bin de délétion analysé (Figure 1.9.). Des analyses de la distribution de la recombinaison sur l'orge (Kunzel, Korzun, and Meister 2000) et le seigle (Heslop-Harrison 1991) suggèrent que la réduction du taux de recombinaison dans les régions centromériques des chromosomes est une caractéristique touchant l'ensemble des *Triticeae*. Ce gradient intra chromosomique du taux de recombinaison au sein de la tribu en fait un modèle intéressant pour l'étude de l'impact des variations du taux local de recombinaison sur l'évolution des gènes, à travers la diversité génétique, la composition en bases ou l'efficacité de la sélection.

Fréquence de recombinaison (cM) par unité de longueur (Mb)



Figure 1.9. : Gradient de recombinaison chez le blé.

Les estimations du taux de recombinaison (cM/Mb) de chaque bin sont représentées par rapport à la distance physique du bin au centromère du chromosome. Sur l'axe horizontal, 0.0 représente le centromère et 1.0 le télomère. La courbe de tendance et son équation sont précisées. Figure extraite de (Akhunov et al. 2003b).

1.4.4. Une large variabilité de systèmes de reproduction

Une autre caractéristique intéressante des Triticeae est la diversité de régimes de reproduction représentée au sein des espèces de la tribu. Le seigle est une espèce allogame auto-incompatible (Lundqvist 1954; Hackauf and Wehling 2005) alors que l'orge est fortement autogame avec un taux d'autofécondation estimé à 98.4% (Brown, Zohary, and Nevo 1978). Le continuum entre l'allogamie et l'autogamie stricte observé au sein de la tribu des Triticeae peut par exemple être illustré par les ratios estimés entre les nombres de grains de pollen et nombre d'ovules (Figure 1.10.). Cet indice se fonde sur le principe selon lequel les plantes autogames investissent moins de ressources énergétiques que les espèces allogames dans la dispersion de leurs gamètes (Cruden 1977). Si cette hypothèse est controversée concernant les espèces entomophiles, une corrélation positive entre la valeur sélective et le nombre de gamètes mâles produits est attendu chez les espèces anémophiles. La Figure 1.10. représente la variabilité observée pour cet indice dans un échantillon de Triticeae. Le seigle, par exemple, est caractérisé par un ratio pollen/ovule d'environ 70 000, alors que d'autres espèces plus autogames présentent un ratio très inférieur à 1000 (Zamora 2005). La Figure 1.10. illustre également les différences morphologiques associées au système de reproduction, notamment au niveau de la taille des anthères, par des photos des épis en fleur d'une espèce allogame (Dasypyrum villosum, en haut) et d'une espèce autogame (*Eremopyrum triticeum*, en bas).

Cette diversité de régime de reproduction parmi des espèces phylogénétiquement proches offre un véritable terrain d'investigation pour mieux comprendre l'impact de l'autofécondation sur l'évolution des gènes.



Figure 1.10. : Variabilité des systèmes de reproduction chez les Triticeae.

Le graphique représente des ratios pollen/ovule (P/O) estimés chez 24 espèces de *Triticeae*. Les espèces présentant les P/O les plus faibles (en bas) produisent peu de pollen proportionnellement au nombre d'ovules. Elles présentent généralement de petites anthères (photo du bas, *Eremopyrum triticeum*) par rapport aux espèces à fort P/O (*Dasypyrum villosum*, en haut). Les espèces à fort P/O (en haut) investissent beaucoup dans leur fonction mâle en produisant une importante quantité de pollen relativement à leur quantité d'ovules (Zamora 2005).

1.4.5. Ressources génomiques publiques

Afin d'étudier l'évolution moléculaire d'une ou de plusieurs espèces, un élément indispensable est.... d'avoir des données moléculaires ! Les plantes cultivées offrent, de ce point de vue, des sujets d'étude particulièrement bien documentés. Du fait de l'importance économique du blé, de l'orge et du seigle, des programmes de séquençage se sont développés autour de ces espèces ainsi que des espèces sauvages apparentées. Le Tableau 1.1. récapitule l'état actuel des données de séquences disponibles dans les bases de données publiques pour plusieurs espèces de *Triticeae*, et les met en regard avec les informations disponibles chez l'homme ou d'autres espèces modèles. Chez les *Triticeae*, plusieurs espèces sont intensivement étudiées, essentiellement le blé et l'orge. A ces ressources, s'ajoute la séquence

du génome du riz dont le génome est 40 fois plus petit que celui du blé et depuis octobre 2007, le génome de *Brachypodium distachyon*, une *Poaceae* située juste à l'extérieur du phylum des *Triticeae* est également disponible (http://www.brachypodium.org/).

Les informations concernant les gènes du riz (nombre d'exons, taille des introns, séquences) peuvent donc servir de référence pour rechercher et caractériser des gènes chez les *Triticeae* (http://www.gramene.org/).

Tableau 1.1. : Quantité de séquences disponibles dans les bases de données publiques (mise à jour du 20/11/2007).

Le tableau précise la taille du génome (en Giga bases), le nombre de nucléotides total en paires de bases (tous types de séquences confondus), d'ESTs (Expresses Sequence Tags) et de gènes reconstitués à partir d'assemblages d'ESTs. Le classement de ces espèces dans la base de données est précisé entre parenthèses. Les espèces dont la séquence du génome est disponible sont indiquées par une étoile.

Organismes	Taille génome (Gb)	Nucléotides (pb)	(No).	EST	(No).	UniGene
Homo sapiens (homme) *	3.5	12 643 419 141	(1)	8 134 112	(1)	122 036
Zea mays (maïs)	2.5	3 175 151 267	(5)	1 159 264	(9)	57 447
Oryza sativa (riz) *	0.45	1 823 316 858	(8)	1 214 083	(8)	40 272
Arabidopsis thaliana *	0.15	877 468 654	(15)	1 276 692	(6)	29 918
Brachypodium distachyon *	0.35	14 513 246	(242)	20 449		
Triticum aestivum (blé tendre)	16.5	647 594 094	(20)	1 050 932	(10)	41 277
Hordeum vulgare ssp. vulgare	5.5	238 113 086	(49)	437 713	(17)	21 418
Hordeum vulgare ssp.	5.5	13 698 238	(270)	24 161		
Triticum monococcum	6.2	8 634 702	(364)	10 139		
Triticum turgidum ssp.durum	12	5 486 683	(527)	8 924		
(ble dur) Aegilops tauschii	?	5 391 913	(538)	116		
Secale cereale (seigle)	8	4 860 151	(624)	9 293		
Triticum turgidum	12	3 770 441	(781)	1 938		
Aegilops speltoides	5	2 512 343	(985)	4 315		
Arabidopsis lyrata	0.23	?	(>1000)	561		

données extraites de Genbank (EST et Unigene) <u>http://www.ncbi.nlm.nih.gov</u>, de la base de données japonaise DDBJ (nucléotides) <u>http://www.ddbj.nig.ac.jp/</u>, Bennett MD, Leitch IJ. 2004. Angiosperm DNA C-values database (release 5.0, Dec. 2004) /cvalues/homepage.html (tailles de génomes).

1.5. Objectifs de la thèse

Au cours de ce document, deux principales études sont détaillées.

Dans un premier temps, j'ai cherché à évaluer l'impact de la domestication sur le niveau de diversité génétique des formes cultivées de blé. Le polymorphisme de séquence de 21 gènes a été estimé pour les quatre principaux représentants de l'histoire de la domestication du blé : l'espèce sauvage, l'amidonnier domestiqué, le blé dur et le blé tendre. Grâce à la comparaison des données de séquences avec des échantillons réalisés par simulations de coalescence, l'intensité des goulots d'étranglement associés à la domestication a pu être estimée. Des variants au modèle démographique moyen défini ont été recherchés, dans le but de détecter des gènes potentiellement impliqués dans les épisodes sélectifs de la domestication. Cette étude est détaillée dans le Chapitre 2.

Dans un second temps, les prédictions théoriques concernant l'impact de l'autofécondation et les variations du taux de recombinaison local ont été testées sur un échantillon d'espèces de *Triticeae*. En contrastant deux espèces autogames et deux espèces allogames d'une part, et des régions génomiques faiblement et fortement recombinantes d'autre part, l'efficacité de la sélection et le patron d'enrichissement en GC ont été estimés sur la phylogénie reconstruite pour 52 gènes. Cette étude est détaillée dans le Chapitre 3.

Chapitre 2.

Polymorphisme de séquence et inférences sur l'histoire de la domestication du blé



Figure 2.1. : Schéma simplifié récapitulant l'évolution des blés domestiqués et cultivés

1. Tétraploïdisation. L'espèce sauvage *T. turgidum dicoccoides* est allopolyploïde : elle possède deux génomes homéologues A et B. Cette espèce est le résultat d'une hybridation spontanée entre le blé diploïde *T. urartu* (génome AA, Dvorak et al. 1993; Dvorak et al. 1998) et une espèce diploïde inconnue, porteuse du génome BB, dont l'espèce la plus proche actuellement identifiée est *Aegilops speltoides* (génome SS, Dvorak and Zhang 1990; Daud and Gustafson 1996). Des données moléculaires suggèrent que l'événement de polyploïdisation ayant donné naissance à l'espèce *T. turgidum* se serait produit entre 0.25 et 1.3 Ma (Mori, Liu, and Tsunewaki 1995; Huang et al. 2002a; Dvorak and Akhunov 2005).

2. Domestication. À partir de l'amidonnier sauvage *T. t.* ssp *dicoccoides* a été domestiquée l'espèce *T. t.* ssp *dicoccum* (amidonnier domestiqué) il y a 10-12 000 ans (Nesbitt and Samuel 1998; Tanno and Willcox 2006) sur le principal critère de rachis solide, limitant par conséquent la dispersion des graines (Figure 2.2.). *T. t.* ssp *dicoccum* s'est répandue à travers l'Europe et l'Asie suivant les migrations humaines. Elle représentait la principale céréale du Croissant Fertile jusqu'à l'âge de Bronze (-10 000ans, Bar-Yosef 1998).

3. Sélection de *T.durum*. Du Néolithique au 2éme millénaire avant JC, l'amidonnier domestiqué a progressivement été remplacé par une nouvelle forme de blé tétraploïde, *T. t.* ssp *durum*. La transition entre les deux formes réside essentiellement sur le passage de grains vêtus à grains nus (diminution de difficulté de battage des grains, Figure 2.2.).

4. Hexaploïdisation : le blé tendre (*T. aestivum*) est l'espèce de blé la plus largement cultivée aujourd'hui. Elle résulte d'une hybridation entre une sous-espèce domestiquée de *T. turgidum* (génome AABB) et une espèce sauvage diploïde (génome DD), *Aegilops tauschii* ssp *strangulata* (Kihara 1944; McFadden and Sears 1946; Dvorak et al. 1998). Cette polyploïdisation aurait eu lieu après les origines de l'agriculture, il y a moins de 8000 ans (Nesbitt and Samuel 1996).

Chapitre 2.

Polymorphisme de séquence et inférences sur l'histoire de la domestication du blé

2.1. Problématique et synthèse des résultats

Le blé a été l'une des premières espèces domestiquées, il y a environ 12 000 ans, dans la région du Croissant Fertile (Nesbitt and Samuel 1998; Tanno and Willcox 2006). L'amidonnier sauvage Triticum turgidum dicoccoides est l'espèce sauvage identifiée comme la représentante actuelle de l'espèce ancestrale à l'origine de la domestication du blé. Cette espèce est allotétraploïde (génome AABB), résultant d'une hybridation spontanée récente (0.25-1.3 Ma entre deux espèces diploïdes Triticum urartu (génome AA) et une espèce proche de Aegilops speltoides donneuse du génome BB (Figure 2.1.1.). Des transitions morphologiques successives s'associent aux différents épisodes de domestication (Figures 2.1. et 2.2.). Un premier évènement de domestication a donné naissance à l'amidonnier domestiqué, la sous-espèce T. t. ssp dicoccum (Figure 2.1.2.). Cette transition vers l'agriculture aurait pris plus d'un millier d'années d'après les dernières études archéologiques (Tanno and Willcox 2006), avec un remplacement progressif de l'amidonnier sauvage par l'amidonnier domestiqué. Suivant les mouvements des populations humaines, l'amidonnier s'est répandu à travers l'Europe et l'Asie jusqu'à l'âge de Bronze (-10 000ans, Bar-Yosef 1998) avant d'être progressivement remplacé par une nouvelle sous-espèce tétraploïde T. t. ssp durum, le blé dur, cultivé encore de nos jours (Figure 2.1.3.). Thuillet et al. (2005) ont estimé les effectifs efficaces au sein des sous-espèces de T. turgidum d'après la diversité mesurée à l'aide de marqueurs microsatellites dont le taux de mutation était estimé (Figure 2.3.). Ils ont ainsi mis en évidence d'importantes pertes de diversité génétique résultant des transitions morphologiques (Figure 2.2.) au cours de l'histoire des blés tétraploïdes. Le blé tendre, T. aestivum, qui représente à lui seul 95% de la consommation mondiale de blé est une espèce hexaploïde (génome AABBDD). Elle est le fruit d'une hybridation interspécifique entre un blé tétraploïde non identifié entre les sous espèces T. t. ssp. durum, carthlicum et dicoccum et l'espèce diploïde Aegilops tauschii var. strangulata (Figure 2.1.4).



Figure 2.2. : Principales caractéristiques morphologiques liées à la domestication des blés.

(A) l'amidonnier sauvage (T. t. ssp dicoccoides); (B) l'amidonnier domestiqué (T. t. ssp dicoccum); (C) le blé dur (T. t. ssp durum) et (D) le blé tendre (T. aestivum) Photos by C. Uauy (Dubcovsky and Dvorak 2007). Les barres d'échelle blanches représentent 1cm. Les lettres en bas à droite indiquent la composition de chaque génome.

La domestication repose principalement sur le passage de rachis cassant (A) à rachis non cassant (B à D), limitant par conséquent la dispersion des graines. Mais d'autres caractères tels que l'augmentation de la taille des grains, la perte de la dormance et l'homogénéisation de la date de floraison accompagnent la domestication. Le passage de grains vêtus (glumes indurées, grain enchâssé ; A et B) à des grains nus (glumes fines et rondes ; C et D) correspond également à une importante modification morphologique, facilitant le battage des grains.



Figure 2.3. : Variations d'effectif efficace au cours de l'histoire des blés tétraploïdes

Le graphique montre les estimations d'effectifs efficaces chez l'amidonnier sauvage *T. t.* ssp *dicoccoides* (A), l'amidonnier domestiqué *T. t.* ssp *dicoccum* (B), les variétés de pays de blé dur (C1) et les variétés élites de blé dur (C2). Les effectifs efficaces ont été calculés d'après la relation $\theta = 4Ne\mu$. Les estimations se basent sur l'indice de diversité mesuré par l'hétérozygotie de Nei sur 15 marqueurs microsatellites pour lesquels le taux de mutation avait été estimé (Thuillet et al. 2005).

Il existe d'autres espèces de blés tétraploïdes qui portent également les génomes A et B, cependant les espèces présentées dans la Figure 2.1. sont les principales représentantes des étapes successives de l'histoire de la domestication du blé. Pourtant peu d'études ont été menées sur la diversité nucléotidique de ces espèces, notamment du fait de la complexité de séquencer des gènes chez un organisme polyploïde récent (du fait de la présence des gènes en plusieurs copies ayant peu divergé et de la difficulté de dessiner des amorces pour l'amplification spécifique par PCR pour un séquençage direct).

Le but de cette étude consiste à estimer les pertes de diversité associées à la domestication et de comprendre l'impact des épisodes successifs de goulot d'étranglement et de sélection sur le génome des formes cultivées. Nous avons étudié le polymorphisme de séquence de 21 gènes au sein des populations *T. t.* ssp *dicoccoides*, *T. t.* ssp *dicoccum*, *T. t.* ssp *durum* et *T. aestivum*. La comparaison des données avec des échantillons issues de simulations de coalescence sous différents scénarios a été utilisée pour estimer les paramètres d'intensité de réduction d'effectif efficace lors des goulots de domestication. La définition du cadre démographique à l'origine de chaque forme cultivée a permis la recherche de locus potentiellement soumis à sélection positive lors de la transition vers les formes cultivées.

Les analyses de diversité ont mis en évidence que la population sauvage T. t. ssp *dicoccoides* contenait relativement peu de diversité au regard d'autres espèces sauvages autogames. Un plus faible niveau de diversité est attendu chez les espèces autogames que chez les allogames (attendu théorique développé en introduction), cependant T. t. ssp *dicoccoides* possède une diversité nucléotidique totale de 0.27 %, inférieure celle observée en moyenne chez les espèces autogames (0.6%, Glémin, Bazin, and Charlesworth 2006). Un effet de fondation dû à la polyploïdisation à l'origine de cette espèce pourrait en partie expliquer la faible diversité observée. La formation du polyploïde T. t. ssp *dicoccoides* étant récente (Mori, Liu, and Tsunewaki 1995; Huang et al. 2002a; Dvorak and Akhunov 2005), on peut s'interroger sur le retour à l'équilibre mutation-dérive de cette espèce. Conformément à cette hypothèse, le niveau de polymorphisme de T. t. ssp *dicoccoides* est supérieur à celui de T. *armeniacum*, une espèce polyploïde sœur issue de l'hybridation des mêmes espèces polyploïdes, d'origine plus récente (Cenci et al. in prep.).

Les épisodes successifs de domestication ont entraîné une forte érosion de cette diversité présente dans la population sauvage sur l'ensemble du génome. En moyenne sur les

21 gènes analysés dans cette étude (couvrant 21 720pb), les populations *T. t.* ssp *dicoccum* et *T. aestivum* ont perdu 69% de la diversité nucléotidique totale de la population sauvage, tandis que *T. t.* ssp *durum* en a perdu plus de 80% !

En utilisant des méthodes de coalescence, des populations théoriques ont été construites selon différents scénarios de réduction de taille efficace des populations (Encadré 1). Afin d'estimer l'intensité des goulots d'étranglement liés à la domestication, les indices de diversité mesurés dans les populations échantillonnées ont été comparés à ceux attendus dans des populations ayant subi différentes réductions de taille efficace par rapport à la population sauvage. Afin de décrire la population sauvage ancestrale, nous nous sommes basés sur l'estimation de la taille efficace de la population sauvage réalisée à partir de quatorze marqueurs microsatellites, calibrés pour leur taux de mutation (Thuillet et al. 2005). Le modèle de simulation que nous avons utilisé consiste en une réduction instantanée de la taille efficace, sans ré-expansion. D'après des simulations, cette approximation du scénario démographique de domestication semble avoir peu d'impact sur le niveau de diversité actuel, étant donné l'échelle de temps très réduite de l'expansion post-domestication (Eyre-Walker et al. 1998).

Pour chaque locus et chaque population cultivée, la vraisemblance des données observées a été estimée pour 50 valeurs d'intensité α de réduction de N_e . En multipliant les vraisemblances entre locus, une intensité de goulot moyenne à l'origine de chacun des compartiments cultivés a pu être estimée en maximisant la vraisemblance multilocus. En moyenne, l'effectif efficace a été réduit par rapport à la forme sauvage d'un facteur de 3.15 lors du premier épisode de domestication (de *T. t.* ssp *dicoccoides* à *T. t.* ssp *dicoccum*), de 4.20 lors de la formation du blé tendre (de *T. t.* ssp *dicoccoides* à *T. aestivum*) et de 5.83 lors du passage au blé dur (de *T. t.* ssp *dicoccoides* à *T. t.* ssp *durum*).

Les pertes de diversité nucléotidique dans les populations cultivées de blé par rapport à la population sauvage sont les plus importantes recensées parmi les espèces domestiquées étudiées à ce jour (Tableau 3 de l'article). Nos estimations étant congruentes avec les ratios d'effectifs efficaces estimés sur les marqueurs microsatellites (Figure 3, Thuillet et al. 2005), les scénarios moyens définis peuvent être considérés comme de bons estimateurs des évènements démographiques subis par les populations cultivées. Tandis que les gènes non sélectionnés sont uniquement soumis aux goulots d'étranglements démographiques, les gènes sous sélection positive pendant les épisodes de domestication sous soumis à des goulots d'étranglement plus sévères (Figure 1.4.). Il est possible d'utiliser le scénario moyen défini pour une population comme « cadre » démographique pour rechercher des gènes sous sélection, présentant des intensités de goulots plus fortes. Ce type d'approche de détection de gènes sous sélection est détaillé dans le paragraphe 3.3.2.

Wright et al. (2005) ont ainsi pu déterminer que 2 à 4 % des gènes chez le maïs auraient été soumis à sélection au cours de la domestication. Pour l'ensemble des gènes étudiés, le niveau de diversité moyen chez le maïs montre est réduit de moitié par rapport à la téosinte (Figure 2.4.). Certains gènes (en gris sur Figure 2.4.) ont perdu significativement plus de diversité que ce que prédit le modèle démographique moyen, ils ont été soumis à un goulot d'une intensité plus forte qui pourrait être due à la sélection lors de la domestication (paragraphe 1.2.3.).



Figure 2.4. : Réduction de la diversité génétique chez le maïs par rapport à la téosinte

Le graphique représente pour chaque gène le niveau de diversité chez le maïs et la téosinte. La ligne en pointillés a une pente de 1, elle représente l'attendu en l'absence de goulot d'étranglement. La ligne pleine a une pente de 0.57, elle représente l'attendu sous le scénario démographique moyen. Les points noirs sont les gènes inclus dans la distribution du goulot démographique, les points gris les gènes candidats

Dans des proportions similaires, parmi les 21 locus analysés dans notre étude, un locus s'écarte significativement du modèle démographique neutre. L'intensité particulièrement forte du goulot d'étranglement subi par ce gène suggère qu'il pourrait avoir été soumis à sélection directe ou indirecte (par entraînement) durant la domestication. Ce gène, comme d'autres parmi l'échantillon, est monomorphe dans les formes cultivées tandis que son niveau de diversité était particulièrement élevé dans la population sauvage. Il montre donc une intensité de goulot d'étranglement plus importante que les gènes dont la diversité était déjà faible avant la domestication. Cependant, parmi les 38 sites polymorphes dans la population sauvage, 7 parmi les sites fixés dans les formes cultivés étaient des variants rares au départ. La fixation des allèles rares est peu probable sous l'action de la dérive, alors que cette situation est attendue dans un scénario sélectif. Cette région chromosomique apparaît comme une zone

soumise à une forte sélection et la diversité présente dans l'espèce sauvage pourrait présenter un fort intérêt pour des études de génétique d'association, afin d'identifier la fonction ou le phénotype en jeu.

Le manuscrit reproduit ci-après détaille cette étude et ses résultats.

2.2. Grinding up Wheat: A Massive Loss of Nucleotide Diversity Since Domestication

Grinding up Wheat: A Massive Loss of Nucleotide Diversity Since Domestication

A. Haudry,* A. Cenci,* C. Ravel,† T. Bataillon,*‡ D. Brunel,§ C. Poncet,* I. Hochu,* S. Poirier,* S. Santoni,* S. Glémin, and J. David*

*UMR Diversité et Adaptation des Plantes Cultivées, Montpellier, France; †UMR Amélioration et santé des plantes, INRA Site de Crouëlle, Clermont-Ferrand, France; ‡BiRC-Bioinformatics Research Center, Department of Genetics and Ecology, University of Aarhus, Aarhus, Denmark; §INRA, Centre National de Génotypage, Evry, France; and ||Institut des Sciences de l'Evolution, Université Montpellier, Montpellier, France

Several demographic and selective events occurred during the domestication of wheat from the allotetraploid wild emmer (Triticum turgidum ssp. dicoccoides). Cultivated wheat has since been affected by other historical events. We analyzed nucleotide diversity at 21 loci in a sample of 101 individuals representing 4 taxa corresponding to representative steps in the recent evolution of wheat (wild, domesticated, cultivated durum, and bread wheats) to unravel the evolutionary history of cultivated wheats and to quantify its impact on genetic diversity. Sequence relationships are consistent with a single domestication event and identify 2 genetically different groups of bread wheat. The wild group is not highly polymorphic, with only 212 polymorphic sites among the 21,720 bp sequenced, and, during domestication, diversity was further reduced in cultivated forms—by 69% in bread wheat and 84% in durum wheat—with considerable differences between loci, some retaining no polymorphism at all. Coalescent simulations were performed and compared with our data to estimate the intensity of the bottlenecks associated with domestication and subsequent selection. Based on our 21locus analysis, the average intensity of domestication bottleneck was estimated at about 3-giving a population size for the domesticated form about one third that of wild dicoccoides. The most severe bottleneck, with an intensity of about 6, occurred in the evolution of durum wheat. We investigated whether some of the genes departed from the empirical distribution of most loci, suggesting that they might have been selected during domestication or breeding. We detected a departure from the null model of demographic bottleneck for the hypothetical gene HgA. However, the atypical pattern of polymorphism at this locus might reveal selection on the linked locus GspIA, which may affect grain softness--an important trait for end-use quality in wheat.

Introduction

Domestication events provide good examples of dramatic morphological and genetic modifications occurring on a short evolutionary time scale. These changes reflect demographic and selective events during the adaptation of crops to a wide range of environments, sometimes very different from those of their native area. Small initial population sizes and intense human selection for agronomic traits are thought to have decreased the available genetic diversity of most crop plants (Tanksley and McCouch 1997). Thus, domestication can be seen as a population bottleneck in most crop species (Buckler et al. 2001). Molecular marker-based studies of crop domestication have increased our understanding of the current genetic status of crop species (Salamini et al. 2002), making it possible to identify agronomically useful genes in wild relatives and to introduce these genes into the cultivated gene pool (Septiningsih et al. 2003) and to identify genes involved in the domestication process or in subsequent selection events (Wright et al. 2005).

Wheat was among the first crop to be domesticated 12,000 years ago in the Fertile Crescent (Nesbitt and Samuel 1998; Tanno and Willcox 2006). Tetraploid forms of current domesticated wheats are derived from a wild tetraploid progenitor, identified as the wild emmer *Triticum turgidum* ssp. *dicoccoides* (referred to as *dicoccoides*). This species has an allotetraploid genome (AABB) resulting from spontaneous amphiploidization between the diploid

Key words: wheat, domestication, nucleotide diversity, bottleneck, coalescence.

E-mail: haudry@supagro.inra.fr.

Mol. Biol. Evol. 24(7):1506–1517. 2007 doi:10.1093/molbev/msm077 Advance Access publication April 18, 2007

© The Author 2007. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. All rights reserved. For permissions, please e-mail: journals.permissions@oxfordjournals.org

wild wheat Triticum urartu (AA genome, Dvorak et al. 1993, 1998) and an unidentified diploid Aegilops species (BB genome), the closest current relative of which is Ae. speltoides (Dvorak and Zhang 1990; Daud and Gustafson 1996; Khlestkina and Salina 2001). Molecular data suggest that dicoccoides is a recent allopolyploid, originating between 0.25 and 1.3 MYA (Mori et al. 1995; Huang et al. 2002; Dvorak and Akhunov 2005). There are still dicoccoides populations in the Fertile Crescent and these populations have been studied with amplified fragment length polymorphism (AFLP) and microsatellite markers (Ozkan et al. 2002; Sasanuma et al. 2002; Thuillet et al. 2005). A recent study based on AFLP data identified 2 different genetic taxa within dicoccoides-a Western race (Israel, Jordan, Lebanon, and Syria) and a Central-Eastern race (Iran, Iraq, and Turkey) (Ozkan et al. 2005)—but the level of genetic differentiation of these 2 races was not estimated. The dicoccoides genotypes from the Central-Eastern group are more closely related to cultivated populations than those of the Western group, suggesting that only this group contributed to the germplasm of domesticated wheat (Ozkan et al. 2002; Mori 2003). Tetraploid wheat domestication seems to have occurred at a single location, in south-eastern Turkey (Mori 2003; Ozkan et al. 2005). This area has been identified as a cradle of crop domestication in the Neolithic era and a probable site for the beginnings of western agriculture (Heun et al. 1997; Nesbitt and Samuel 1998; Lev-Yadun et al. 2000; Salamini et al. 2002).

The first domesticated tetraploid wheat emmer (*Triticum turgidum* ssp. *dicoccum*, referred to as *dicoccum*) has a nonbrittle rachis and a uniform flowering time, lacks grain dormancy, and has larger kernels than the wild *dicoccoides*. Emmer was spread with human migration throughout Europe and Asia and was the most important crop in the

Fertile Crescent until the early Bronze Age, 10,000 BC (Bar-Yosef 1998). Emmer was gradually replaced by a new form of tetraploid wheat (Triticum turgidum ssp. durum, referred to as durum) considered to be the ancestral form of current macaroni wheat. The transition from emmer to modern durum wheat involved the acquisition of free threshing. Major losses of neutral genetic diversity occurred at successive stages in the history of Triticum turgidum ssp. In diversity surveys based on microsatellite loci calibrated for their mutation rate, the wild dicoccoides was found to have an average effective population size (N_e) of 32,500 (Thuillet et al. 2005). This size corresponds to the effective number of breeders in an ideal Wright-Fisher population. The estimated effective population size of emmer (*dicoccum*) is only half ($N_e = 12,000$) and that for cultivated *durum* is only a fifth of this value ($N_e = 6,000$ in old landraces and 1,300 in the most recently improved varieties). These marked decreases in N_e during wheat improvement history illustrate the intensity of the successive bottlenecks in tetraploid wheat evolution.

Zohary (1999) investigated the number of times that the wild progenitors of Neolithic agriculture were domesticated in the Near East. Based on polymorphism and taxonomic information, he concluded that emmer wheat was domesticated only once, consistent with the monophyletic origin of emmer. This domestication event may have continued over a millennium, during which time wild wheat persisted in cultivated fields (Tanno and Willcox 2006). However, Ozkan et al. (2005) recently argued that the origins of domesticated tetraploid wheat are consistent with a scenario involving 2 major lineages still found in *durum* and *dicoccum*.

The history of tetraploid wheat domestication is well documented, but that of common wheat remains incomplete. Bread wheat (Triticum aestivum referred to hereafter as *aestivum*), the most widely cultivated wheat today, is a hexaploid form of free-threshing wheat (genome AABBDD). It is thought to have resulted from recent hybridization (no more than 8,000 years ago, according to Nesbitt and Samuel [1996]) between an allotetraploid wheat (AABB) and the diploid (DD) Aegilops tauschii var. strangulata (Kihara 1944; McFadden and Sears 1946; Dvorak et al. 1998). The sources of the tetraploid AB genomes of aestivum remain a matter of debate. If T. aestivum shares its A and B genomes with the T. turgidum spp. wheats, its allotetraploid progenitor is currently not identified although it is hypothesized that a domesticated form was involved in this cross because the current distribution range of Ae. tauschii does not overlap with the distribution of the wild *dicoccoides* (Nesbitt and Samuel 1996). Zohary and Hopf (2000) suggested that the tetraploid dicoccum might be the progenitor of aestivum, with a Caspian origin for the hybridization with Ae. tauschii generating a hulled hexaploid wheat. This hybridization was then followed by the rapid emergence of freethreshing forms. However, as pointed out by Nesbitt and Samuel (1996), several lines of archaeological evidence, including the lack of remains of hulled hexaploid wheats in this area, are inconsistent with this hypothesis. It is therefore possible that free-threshing in hexaploids was directly inherited from free-threshing tetraploids, consistent with genetic evidence for the rapid emergence of free-threshing tetraploids (Salamini et al. 2002). The A and B genomes of *durum* and *aestivum* not only show extended conservation but also have marked differences (Isidore et al. 2005). However, combined polymorphism analyses of tetraploid and hexaploid wheats have not yet been carried out. No conclusive study has yet identified unambiguously the sources of the A and B genomes of bread wheat among tetraploid potential donors. Based on the D genome polymorphism in bread wheat, amphiploidization with *Ae. tauschii* is thought to have occurred at least twice (Dvorak et al. 1998; Giles and Brown 2006), so there may have been at least 2 different tetraploid progenitors. Subsequent gene flow from tetraploid progenitors to hexaploids, as suggested by Caldwell et al. (2004), may have boosted genetic diversity within bread wheat and blurred the genetic evidence for the origin of *aestivum*.

Few studies have been carried out on nucleotide diversity in wheat because the presence of 2 or 3 closely related homologous copies in the genome prevents the direct sequencing of polymerase chain reaction (PCR) products. Nucleotide sequence variation is much less prone to homoplasy than microsatellite polymorphism. It provides a powerful mean of unraveling the evolutionary history of crop plants and reconstructing genealogies in populations. Microsatellite analyses, as described by Thuillet et al. (2005), may underestimate the consequences of bottlenecks for nucleotide diversity because the high mutation rate of microsatellites might have allowed some recovery of diversity since domestication (Vigouroux et al. 2002). Most crops were domesticated around 10,000 years ago and therefore cannot be considered to be at the mutation-drift equilibrium. Consequently, studies of domestication require demographic scenarios for reconstructing gene genealogy. Coalescent theory (Hudson 1990) allows efficiently simulating sequence samples under different scenarios. Statistical tests can then be used to identify the scenario most likely to account for the observed polymorphism patterns of the studied samples (Nordborg 2003). DNA sequences and a coalescent framework have been used to investigate population bottlenecks in humans (Wakeley and Hey 1997) and in maize (Eyre-Walker et al. 1998; Tenaillon et al. 2004; Wright et al. 2005). Comparisons of the loss of genomewide diversity between wild and cultivated species for large sets of genes can be used to calibrate a plausible scenario for domestication bottleneck. Nonselected genes should have levels of nucleotide diversity consistent with a genomewide demographic bottleneck, whereas genes selected during or after domestication would be expected to show a locally more severe decrease in nucleotide diversity (Wright et al. 2005). This contrast can be used to test whether the patterns of diversity at a given candidate locus in a crop and its wild progenitor can be explained by a demographic event alone or by selection during domestication (Wright et al. 2005).

The aim of this study was to characterize genetic diversity in domesticated (*dicoccum*) and cultivated wheats (*durum* and *aestivum*) and their wild tetraploid relative (*dicoccoides*), to try to unravel the evolutionary history of cultivated durum and bread wheats, and to quantify the impact of domestication bottlenecks on genetic diversity. We addressed these issues by investigating the nucleotide diversity of 21 genes in a sample of 101 individuals from

Table 1									
Sequenced	Genes,	Their	Chromosome	Location,	Putative	Function,	and	Sequenced	Length

Gene Name	Location (chromosome, bin ^a)	Putative Gene Function	Sequence Length (bp)
11B	1BS9	Unknown	692
91A	3AL5	Unknown	1,252
AapA	2AL1	Amino acid permease	1,019
AlperA	6AL	Xanthine/uracil/vitamin C permease	1,169
Bp2A	3AL3	ATP biosynthesis	1,433
Bp3B	3BL10	ATP biosynthesis	511
Bp5A	3AL3	ATP biosynthesis	635
ChsA	5AS7	Chalcone synthase	436
GdhA	5AL10	Unknown	1,234
Gsp1A	5AS7	Grain softness protein	939
Gsp1B	5BS	Grain softness protein	473
HgA	5AS7	Hypothetical gene	847
HiplA	5AS7	Hedgehog-interacting protein	615
MdhA	3AL3	Malate dehydrogenase	845
Mdh4B	3BL10	Malate dehydrogenase	1,491
Mp7A	2AS	SNF2 family N-terminal domain	878
<i>MybA</i>	3AL3	Transcription factor	1,427
MybB	3BL10	Transcription factor	3,372
NrpA	5AS7	Nodulin-related protein (?)	963
PsyA	5A	Phytoene synthase (carotenoid biosynthesis)	598
ZdsB	2B	Lycopene synthase (carotenoid biosynthesis)	891

^a Location on the chromosome bin map build, as determined with a set of wheat aneuploids and deletion stocks (Qi et al. 2004).

4 taxa corresponding to representative stages in recent wheat evolution (wild, domesticated, cultivated durum, and bread wheats). We used these data to assess the genealogical relationships between the 4 taxa, to provide insight into the origin of cultivated wheats. We then compared the genetic diversity of the wild population with that of its cultivated relatives and used coalescent simulations to quantify bottlenecks associated with wheat domestication and subsequent selection. Finally, we tested whether some of the genes in our sample were selected during domestication.

Materials and Methods

Plant Materials

We used 4 wheat taxa for DNA sequence analysis: the wild dicoccoides, the domesticated dicoccum, and 2 wheats cultivated today: durum and bread wheat-durum and aestivum, respectively. For each taxon, we used a core set of individuals representing the highest available levels of allelic diversity. These individuals were chosen to maximize the number of alleles observed at 30 microsatellite loci (David et al. 2003). We sequenced 28 dicoccoides, 12 dicoccum, 20 durum, and 41 aestivum individuals. The accession numbers and geographic origins of the samples are shown in supplementary table S1 (Supplementary Material online). The sample sizes of different loci differed because not all loci were successfully amplified or sequenced in all individuals. Finally, we studied the genetic structure of the wild dicoccoides population, using a previously reported data set corresponding to the 52 accessions genotyped with 15 microsatellites (Thuillet et al. 2005).

Design of Genome-Specific Primers

The allopolyploid origin of wheat from 2 (*T. turgid-um*) or 3 (*T. aestivum*) ancestral genomes prevents direct sequencing. The sequencing of genes in polyploid wheat

requires either cloning or the development of genomespecific primers to ensure that only the targeted copy is amplified. We amplified gene fragments with locus- and genome-specific primers, designed as previously described (Ravel et al. 2006), to prevent the amplification of paralogous and homeologous loci. We then tested the genome specificity of amplification systematically on a set of 7 genotypes: 2 AA diploid accessions (Triticum monococcum and T. urartu), 2 BB-like diploid accessions (Ae. speltoides selfed progeny), 2 DD diploid accessions (Ae. tauschii), and 1 tetraploid AABB accession (Triticum turgidum ssp. durum var. Langdon). If a single fragment of the expected size was amplified only in individuals with the targeted genome, the complete sample was amplified. Sequences were submitted to GenBank (GenBank accession numbers are listed in supplementary table S2 [Supplementary Material online]).

Loci Sampled

We sequenced 21 gene fragments from 101 individuals: 15 from the A genome—91A, AapA, AlperA, Bp2A, Bp5A, ChsA, GdhA, Gsp1A, HgA, HiplA, MdhA, Mp7A, MybA, NrpA, and PsyA—and 6 from the B genome—11B, BP3B, Gsp1B, MdhB, MybB, and ZdsB. Six of these genes are located in the vicinity of the hardness locus: ChsA, HgA, HiplA, Gsp1A, NrpA, and Gsp1B (Chantret et al. 2005). The size of the amplified fragment, the chromosome location and the putative function of the corresponding gene are reported in table 1. The PCR and sequencing conditions used are described in supplementary table S3 (Supplementary Material online).

Statistical Analysis

Sequences were aligned manually with the Staden Package (Staden et al. 2001). Because of recombination events among loci, it is difficult to assess genealogical relationships among accessions. To get a rough idea of main relationships, we concatenated all loci except *ChsA* (sequence data lacking for *aestivum*) and performed maximum likelihood (ML) reconstruction (model general time reversible + gamma distribution) using the PHYML software (Guindon and Gascuel 2003). Concatenation resulted in many missing data. We therefore used one of the most parsimonious trees (using DNAPARS procedure of the PHYLIP package version 3.6, Felsenstein 2005) as the starting point for ML search, which is more robust to missing data than using distance trees. Five sequences from *T. timopheevii*, an allotetraploid sister species, were used as outgroup.

We used DnaSp version 4.10 (Rozas et al. 2003) to calculate the number of polymorphic sites (S), the number of haplotypes (h), and the nucleotide diversity per site (π) (Tajima 1983) calculated for the whole sequence (π_{total}) and for noncoding and synonymous sites (π_{silent}). Single-locus and multilocus Tajima's D test (Tajima 1989) was performed in each group using J. Hev's HKA software (http: //lifesci.rutgers.edu/~hevlab/ProgramsandData/Programs/ HKA/HKA Documentation.htm). We investigated the consequences of domestication for diversity in the wheat genome, using the current wild group dicoccoides as a proxy for the initial population before domestication. A recent study has suggested that *dicoccoides* may form 2 main populations (Ozkan et al. 2005). We tested for the presence of these 2 main populations 1) by analysis with STRUCTURE software (Pritchard et al. 2000) of the microsatellites data set for the collection of 52 accessions of dicoccoides from Thuillet et al. (2005) and 2) by classical Fst statistic analysis between the 2 groups detected by Ozkan et al. (2005) using the Genetix software (Belkhir et al.; http:// www.genetix.univ-montp2.fr/genetix/intro.htm). We investigated the distribution of both neutral and nucleotide diversity in *dicoccoides* by carrying out Mantel's correlations of genetic and geographic distances with microsatellite and sequence data using the GenAlEx 6.0 software (Peakall and Smouse 2006).

Demographic Model

We used a simple model of reduction in effective population size (fig. 1), in which a single ancestral population (the wild population) experienced an instantaneous change in effective population size, t generations ago. The bottleneck intensity α was defined as the ratio of the wild population size (N_a) to cultivated population size (N_p) . Higher values of α correspond to more severe bottlenecks. We kept the demographic scenario simple by not allowing for an increase in population size after the bottleneck. This approximation has been shown to have little effect on levels of nucleotide diversity as shown by simulations with this type of bottleneck model in maize (Eyre-Walker et al. 1998). Previous studies have shown that diversity after bottleneck scales to the ratio of the size of the bottleneck population (N_b) and the duration of the bottleneck (d), such that the 2 parameters cannot be estimated separately (Eyre-Walker et al. 1998; Tenaillon et al. 2004; Wright et al. 2005). Here, d = t, the number of generations since domestication, and



FIG. 1.—Schematic diagram of the coalescent model used in simulations. The ancestral population experienced an instantaneous change in effective population size (N_a) , *t* generations ago. The bottleneck intensity α is defined as the ratio of ancestral population size (N_a) to cultivated population size (N_p) .

the constant parameter is the product $\alpha \times t$. Assuming a shorter duration of the bottleneck will increase α . The choice of this scenario can also be justified due to the relatively short time (on an evolutionary scale) for the recovery of nucleotide polymorphism after domestication and the continuous selection experienced by wheat populations since domestication.

For the estimation of bottleneck intensities, we used the *dicoccoides* data to calibrate the simulation parameters for the ancestral population, and we used data for *dicoccum* (cultivated emmer), *durum* (durum wheat), or *aestivum* (bread wheat) as the observed data in cultivated groups for a goodness-of-fit analysis (see below). For each locus, the model had 5 parameters (τ , N_a , N_p , θ_{wild} , and 4Nc):

- τ , the time since the bottleneck was expressed in units of time scaled relative to effective size as $\tau = t/2N_p$, where N_p is the effective population size after the bottleneck. As domestication is thought to have occurred 12,000 years ago (Harlan 1992), t = 12,000. The N_p is equivalent to the ancestral population size (N_a about 30,000; Thuillet et al. 2005) divided by the bottleneck intensity: $N_p = N_a/\alpha$. We therefore used $\tau = \alpha t/2N_a$ in simulations, that is, $\tau = 0.2\alpha$.
- Assuming that *dicoccoides* is the progenitor of the A and B genomes of wheats, we used $\theta_{dicoccoides}$ as a proxy of the initial θ_{wild} . The population mutation rate $\theta_{dicoccoides}$ was estimated by Tajima's π statistic (Tajima 1983), based on sequences from wild *dicoccoides*.
- For each locus, the population recombination parameter (4Nc) was estimated from *dicoccoides* data, using Hudson's 2001 method by LDhat program (http://www.stats.ox.ac.uk/~mcvean/LDhat/LDhat1.0/

LDhat1.0.html). This parameter was included in simulations when it could be estimated. Otherwise, we assumed that no recombination occurred.

Goodness-of-Fit between Simulations and Observed Data

Coalescent simulations were performed and compared with our data to model the impact of a bottleneck on sequence diversity. Coalescent simulations were run with the "ms" program (Hudson 2002). For each locus and each cultivated group, 50 values of α were explored on a grid ranging from 1 (i.e., no decrease in effective population size) to 25.5. For each locus and group considered, 10,000 simulations were carried out. Each coalescent simulation was summarized by a π_{simul} and a S_{simul} value. For each scenario, the approximate likelihood of the data at locus *i* within group *j*, $L_{ij}(\alpha)$, was calculated as the number of simulations in which both π_{simul} and S_{simul} were within 20% of the observed values of π and S for the data (Weiss and von Haeseler 1998). The intensity of the bottleneck at locus *i* within group *j* was estimated as the value α maximizing $L_{ij}(\alpha)$.

We estimated the average bottleneck intensity for each of the 3 cultivated groups by calculating a multilocus likelihood $Lm_i(\alpha)$ as the product over all loci of $L_{ii}(\alpha)$. This approach implicitly assumes that the loci are independent. However, 5 of the loci considered here are located in the same chromosome region, 5AS7 (ChsA, HgA, GsplA, NrpA, and HiplA; Sourdille et al. 1996). We estimated the linkage disequilibrium (LD) between the polymorphic sites for these loci, using TASSEL software (http://www. maizegenetics.net/index.php?page=bioinformatics/tassel/ index.html). LD was significant only at the intragene level, and no LD was detected between the different loci of the 5A region ($r^2 < 0.2$ within 1 kb of the "hardness locus," data not shown), so all genes can be assumed to be independent in our likelihood calculation. The intensity of the bottleneck within each group j was calculated as the value of α maximizing $Lm_i(\alpha)$. A 95% confidence interval (CI) was constructed around the estimate of α by identifying the value of α at which the log-likelihood value was 2 log-likelihood units lower than the ML.

Using the Demographic Model to Test for Selection

Selection at some loci would result in the distribution of polymorphisms being skewed at these loci, which might account for the observed variability in bottleneck intensity α among loci. The loci with the most severe bottleneck estimates were considered to be candidate loci for selection during domestication. We investigated whether some loci were outliers in the empirical distribution of most of the loci, by calculating the P value associated with their observed π value. We used the mean value and upper CI limit of α determined by the demographic model to perform additional simulations at each locus. The *P* value of $\pi_{observed}$ was calculated for these distributions. If significant (P <0.05), the locus was discarded and the analysis was repeated with n - 1 loci. This procedure was repeated until no significant effect was detected and for all loci without polymorphism in the cultivated group.

Results

Relationships Between Taxa

Because of low diversity levels, single-locus analyses are not powerful enough to detect clear relationships among the different forms of wheat. We thus performed multilocus analyses combining all genes (fig. 2). Like other tree representation based on a combination of marker information widespread over the genome, this tree should be interpreted with caution because of recombination events between loci. It can give a general picture of accessions relationships but detailed analyses can be misleading. The general topology shows that all cultivated forms are subsets of the wild dicoccoides group, consistent with a single domestication event. Three dicoccoides accessions fall within cultivated accessions, but we do not have clear explanation for this finding. Long branches are due to the higher diversity in the wild group. The domesticated dicoccum forms are dispersed within this cultivated group. Durum wheat individuals cluster together in a subgroup. They are included in the dicoccum lineage rather than forming a different lineage. Bread wheat presents a singular pattern, with 2 different groups, 1 lying on an external branch (I) and the other (II) spanning a large proportion of cultivated wheat diversity. To test further this pattern, we performed a STRUC-TURE analysis with admixture on the bread wheat data set. We also found 2 clearly distinct groups mainly corresponding to those observed on the tree (see supplementary fig. S4, Supplementary Material online). We found no clear relationship between the genealogy of *aestivum* accessions and their geographic origin (see supplementary table S1, Supplementary Material online).

Genomic Characterization of the Wild *dicoccoides* Group

On the 21,720 bp, corresponding to the 21 loci, in dicoccoides, we identified a total of 212 single nucleotide polymorphisms (SNPs). The nucleotide diversity π ranges from 0.0006 (AapA) to 0.0116 (HgA), with a mean value of 0.0027 (table 2). For both π and θ_{W} , diversity is greatest for the genes HgA ($\pi = 0.0116$ and $\theta_W = 0.0141$) and ChsA $(\pi = 0.0113 \text{ and } \theta_{W} = 0.0112)$. The lack of genetic variability made it difficult to estimate the population recombination parameter 4Nc. We were unable to estimate this parameter for 7 loci (table 2), and it was therefore set to zero when running coalescent simulations (see below). For the other 13 genes, the population recombination rate 4Nc ranges from 0.002 (Mdh4B) to 0.067 (ChsA), with a mean value of 0.015 per nucleotide. When considering each locus individually, we detected no significant departure from the neutral equilibrium model in Tajima's test (Tajima 1989) but values of Tajima's D statistic tend to be overall slightly negative and the multilocus test is highly significant (mean D = -0.76762, P < 0.001).

Ozkan et al. (2005) found 2 main geographic groups in *dicoccoides* and suggested that domestication occurred likely in the Turkish area. Surprisingly, using the STRUC-TURE software (Pritchard et al. 2000), we did not find the same 2 groups. Using sequence data, we found only one group, and using microsatellite data, we only detected a small group of Palestinian and Israeli accessions (but the likelihood of the data assuming 2 populations was only slightly higher than when assuming a single one, supplementary fig. S4 [Supplementary Material online]). In addition, we found no significant correlation between genetic and geographical distances in our samples of *dicoccoides*, either for microsatellite or for sequence data. Significant genetic isolation by distance was detected only between



Fig. 2.—Maximum likelihood phylogenetic tree (GTR + Γ) built with sequences for the 20 loci concatenated, with 5 sequences from *T. timopheevii* used as outgroup. Individuals are denoted as shown in supplementary table S1 (Supplementary Material online). The 2 groups of *aestivum* are identified as I and II.

 Table 2

 Sequence Statistics for the Loci Studied in Wild Emmer

Gene	п	L (bp)	S	$\pi imes10^{-3}$	$\theta_{\rm W}\times10^{-3}$	4Nc	Tajima's D	π_a/π_s
11B	10	692	10	3.28	5.11	0.0260	-1.59 ns	0.504
91A	10	1,252	7	2.18	1.98	0.0399	0.44 ns	0.061
AapA	9	1,019	2	0.6	0.72	_	-0.58 ns	
AlperA	10	1,169	7	2.28	2.12	0.0026	0.33 ns	0
Bp2A	27	1,433	18	1.67	3.26	0.0028	-1.72 ns	0.165
Bp3B	10	511	1	1.04	0.69	_	1.3 ns	
Bp5A	26	635	5	0.83	2.06	_	-1.7 ns	0
ĊhsA	6	436	11	11.27	11.15	0.0673	0.06 ns	0.39
GdhA	26	1,234	11	2.4	2.45	0.0040	-0.08 ns	0
Gsp1A	25	939	26	4.2	7.33	0.0043	-1.58 ns	0.12
Gsp1B	28	473	7	2.23	3.83	_	-1.25 ns	0.49
HgA	14	847	38	11.63	14.12	0.0047	-1.03 ns	0
HiplA	26	615	6	1.8	2.56	0.0049	-0.87 ns	0
Mdh4B	8	1,491	6	1.72	1.55	0.0020	-0.22 ns	
MdhA	27	845	7	1.16	2.15	_	-1.4 ns	
Mp7A	24	878	4	0.97	1.27	_	-0.65 ns	
<i>MybA</i>	10	1,427	3	0.61	0.74	0.0216	-0.66 ns	1
МybB	10	3,372	16	1.37	1.68	0.0203	-0.86 ns	0
NrpA	17	963	15	2.58	4.61	0.0052	-1.69 ns	0.69
PsyA	28	598	7	1.84	3.02	0.0042	-1.18 ns	0.32
ZdsB	28	891	5	0.82	1.44		-1.19 ns	

NOTE.—*n*, number of *dicoccoides* individuals sampled; *L* (bp), length of aligned sequence, excluding gaps; *S*, total number of segregating sites; $\pi \times 10^{-3}$, average number of pairwise differences calculated on all sites; $\theta_W \times 10^{-3}$, per-site estimates of diversity by Watterson's theta; 4Nc, population recombination parameter 4Nc (no recombination could be detected for 7 loci [—]); Tajima's *D*, Tajima's *D* statistic for all sites; π_a/π_s , ratio of nonsynonymous diversity (π_a) to synonymous diversity (π_s) (when π_s was zero, it was not possible to calculate the π_a/π_s ratio [—]); ns, not significant.

populations located less than 50 km apart (supplementary fig. S4, Supplementary Material online). However, using the microsatellite data set, we found low but significant *F*st values assigning our accessions to the 2 geographic groups identified by Ozkan et al. (2005) (*F*st = 0.026, *P* value = 0.012 after 1,000 random permutations). Nucleotide diversity is similar in both populations ($\pi = 0.0024$ for the Central East group and $\pi = 0.0027$ for the Western one) and close to the mean diversity obtained over the whole populations ($\pi = 0.0027$). Overall, these results suggest that population structure at a large geographic scale does occur in *dicoccoides* but such a structure is weak and can hardly be detected with 21 genes or 15 microsatellites markers.

From Wild to Cultivated Wheats: Important Losses of Nucleotide Diversity

The transition from wild to cultivated forms was marked by a large decrease in nucleotide diversity. We detected a mean of 10.1 polymorphic sites per locus in the wild dicoccoides, whereas only 3 polymorphic sites per locus were found in dicoccum and aestivum and 1.7 such sites were found in durum (supplementary table S5, Supplementary Material online). Nucleotide diversity (π) in the cultivated groups ranged from 0 (3 of 21 genes in dicoccum, 7 in durum, and 6 in aestivum) to 0.003 (Gsp1B in dicoccum). The mean value of π was 0.0008 for *dicoccum* and *aesti*vum, and this value was halved in durum. The rate of nucleotide diversity loss was similar when comparing silent sites (measured on noncoding and synonymous sites) and all sites (fig. 3). The domesticated dicoccum has 70% less diversity than the wild dicoccoides, whereas durum is 84% less diverse than the wild taxon. In aestivum, nucleotide diversity is 69% lower than that in wild *dicoc-coides*. Tajima's *D* statistic tended to be negative in *dicoc-cum* and *durum* (supplementary table S5, Supplementary Material online). Higher values of Tajima's *D* statistic were obtained in the *aestivum* group (*Gsp1A* and *MybB*).

Coalescent theory–based analyses of nucleotide polymorphism were consistent with strong bottlenecks in each cultivated group (fig. 4). The ML estimate of bottleneck intensity α for the domestication bottleneck (transition *dicoccoides* to *dicoccum*) was $\alpha_{dicoccum} = 3.15$ (CI = 2.07–4.53). The ML estimate of α for the transition from *dicoccoides* to durum wheat was 5.83 (CI = 4.35–7.94), probably due to



FIG. 3.—Nucleotide diversity (π) expressed in 10⁻³ in the wild (*dicoccoides*) and cultivated groups (*dicoccum*, *durum*, and *aestivum*). Total nucleotide diversity π_{total} was estimated for the whole sequence, whereas π_{silent} corresponds to nucleotide diversity for noncoding and synonymous sites only.



FIG. 4.—Likelihood (yaxis, logarithmic scale) profiles as a function of bottleneck intensity, α (xaxis) for each cultivated group, based on the 21 loci ($\alpha = Ne_{dicoccoides}/Ne_{cultivated}$).

a further loss of diversity after domestication. Bread wheat (aestivum) displayed a lower loss of diversity after domestication ($\alpha_{aestivum} = 4.20$, CI = 3.10–5.74). The CI of the bottleneck intensities (α) experienced by the *aestivum* and dicoccum groups overlapped considerably. These bottleneck intensities reflect the diversity reduction from wild to domesticated populations when *dicoccoides* is taken as a whole. If only Central-Eastern dicoccoides populations were the founders of the domesticated group, true α values for the domestication process per se might be smaller. Therefore, we redid coalescent simulations using the parameters of the Central-Eastern group to characterize the wild initial population. Because, nucleotide diversities are similar to those computed on the whole set of accessions, bottleneck intensities are almost unchanged $(\alpha_{dicoccum} = 2.61, \alpha_{durum} = 5.45, \text{ and } \alpha_{aestivum} = 4.53)$ and CI overlap (see supplementary fig. S4 [Supplementary Material online] for details).

Variation in Bottleneck Intensity Among Loci

The loss of nucleotide diversity π in cultivated groups compared with the wild *dicoccoides* varied widely among loci (fig. 5). Some genes displayed little or no loss of diversity in the *dicoccum* group (*Gsp1B*, *MybB*, *NrpA*, and *ZdsB*) and in bread wheat (*HiplA* and *MybB*). However, most genes showed a sharp decrease of genetic variability compared with the wild *dicoccoides*. In particular, the *HgA* locus (circled points in fig. 5) displayed a drastic loss of diversity from the wild to the 3 cultivated groups. Some genes were monomorphic in the cultivated groups (3/21 genes in *dicoccum*, 6/20 in *aestivum*, and 8/21 in *durum*). We detected only one instance of departure from the null model of neutral demographic bottleneck. In the *aestivum*



FIG. 5.—Joint patterns of nucleotide diversity observed in wild *dicoccoides* and cultivated relatives at 21 gene fragments, with each gene plotted 3 times. The $\pi_{\text{wild}}/\pi_{\text{cultivated}}$ ranged from 0 to more than 1, depending on the gene considered. The line (slope = 1) indicates equivalent levels of diversity in *dicoccoides* and the cultivated taxa. The encircled points represent diversity at the *HgA* locus in the 3 cultivated groups.

group, the pattern of HgA polymorphism, as summarized by π , cannot be explained by the estimated mean bottleneck intensity of $\alpha_{aestivum} = 4.20$ (*P* value = 0.0054). Even when the upper limit of the CI of $\alpha_{aestivum}$ (5.74) was used for coalescent simulation, the pattern of polymorphism at HgA remained atypical (*P* value = 0.0162). In such multiple testing, with an individual threshold value of 5% for significance, 1 gene of the 20 tested in *aestivum* would be expected to give false-positive results. This weakens the evidence for a possible selective event on HgA, which should be viewed with caution.

Discussion

Relationships between Wild Emmer and Cultivated Wheats

Wild emmer has been identified as the wild progenitor of cultivated wheat. Current populations of *dicoccoides* have been reported to fall into 2 genetically different groups (Ozkan et al. 2005). In our *dicoccoides* sample, we detected no population structure without a priori assumption and no significant correlation between genetic and geographic distances on a large scale. However, we detected low, but significant, *F*st values between the 2 groups previously identified (a Western race and a Central-Eastern race). Our sample and data set may have been too small to detect weak genetic differentiation between *dicoccoides* groups. As our *dicoccoides* sample contains accessions from the whole species distribution, we could assume that this sample covers a large proportion of the diversity available in the wild species.

The tree reconstructed from the concatenated 20 gene fragments revealed the distribution of nucleotide diversity within the 4 groups (wild, domesticated, durum, and bread wheat). The diversity in the cultivated group is clearly a subset of the diversity of the wild group, as would be expected for a domestication event (Buckler et al. 2001). The monophyly of all the cultivated individuals in the tree is consistent with a single domestication event for emmer wheat (Zohary 1999). Our results are not consistent with recent

Table 3					
Nucleotide	Diversity i	in Wild	and	Domesticated	Relatives

Mating System	Diversity in Wild (10^{-3})	Diversity in Cultivated (10^{-3})	Loci	<i>L</i> π (%)	References
	Zea mays ssp. parviglumis	Zea mays ssp. mays			
Outbreeding	$\pi_{\text{total}} = 9.7$	$\pi_{\text{total}} = 6.4$	774	35	Wright et al. (2005)
	$\pi_{\text{silent}} = 21.1$	$\pi_{\text{silent}} = 13.1$	12	38	Tenaillon et al. (2004)
	Medicago sativa ssp. sativa	M. s. ssp. sativa	2		Muller et al. (2006)
Outbreeding	$\pi_{\text{total}} = 20.2$	$\pi_{\text{total}} = 13.5$		31	
	$\pi_{\text{silent}} = 29$	$\pi_{\text{silent}} = 20$		31	
	Helianthus annuus	H. annuus	9		Liu and Burke (2006)
Outbreeding	$\pi_{\text{total}} = 12.8$	$\pi_{\text{total}} = 5.6$		55	
	$\pi_{\text{silent}} = 23.4$	$\pi_{\text{silent}} = 9.6$		59	
Mixed	Pennisetum glaucum	P. glaucum	1		Gaut and Clegg (1993)
	$\theta_{\text{silent}} = 3.6$	$\theta_{\rm silent} = 2.4$		33	
	Glycine soja	Glycine max	102		Hyten et al. (2006)
Inbreeding	$\pi_{\text{total}} = 2.17$	$\pi_{\text{total}} = 1.43$		34	
	$\pi_{\rm silent} = 2.76$	$\pi_{\rm silent} = 1.77$		36	
	Hordeum spontaneum	Hordeum vulgare			
Inbreeding	$\pi_{\text{silent}} = 16.7$	$\pi_{\rm silent} = 7.1$	5	57	Caldwell et al. (2006)
	$\pi_{\text{total}} = 8.3$	$\pi_{\text{total}} = 3.1$	7	62	Kilian et al. (2006)
	Triticum turgidum ssp. dicoccoides	Triticum turgidum ssp. dicoccum	21		This study
Inbreeding	$\pi_{\text{silent}} = 3.6$	$\pi_{\text{silent}} = 1.2$		65	
	$\pi_{\text{total}} = 2.7$	$\pi_{\text{total}} = 0.8$		70	

Note.—The loss of diversity during domestication was calculated as $L\pi = 1 - \pi_{dom}/\pi_{wild}$ and nucleotide diversity calculated on silent sites (π_{silent}) and on all sites (π_{total}) are reported (when possible).

suggestions of a possible diphyletic origin for domesticated tetraploid wheats as suggested by Ozkan et al. (2005). The *dicoccum* sequences are widely distributed throughout wheat lineages, spanning the whole range of diversity found in the cultivated group. *Durum* individuals fall into a single clade including some *dicoccum* individuals: these observations are consistent with *dicoccum* being a progenitor of durum wheat. It is not possible to identify precisely which tetraploid donated its A and B genomes to *aestivum*, but *durum* is unlikely the donor subspecies because its nucleotide diversity does not include that of *aestivum*.

Genetic Diversity in dicoccoides

The mean nucleotide diversity observed for these 21 genes ($\pi_{total} = 0.0027$ and $\pi_{silent} = 0.0036$) suggests that dicoccoides is not a highly polymorphic species. All else being equal, self-fertilizing species are expected to have a lower diversity level than outcrossing species. Selfing reduces effective population size N_e by reducing gamete sampling, and because of low effective recombination rates, hitchhiking effects further reduce diversity (Charlesworth and Wright 2001). Inbreeding and asexual species often have a life history involving frequent local colonization and extinction events, potentially reducing diversity even further (Kimura and Ohta 1971; Charlesworth D and Charlesworth B 1995). Triticum dicoccoides displays lower levels of variation than teosinte ($\pi_{total} = 0.0097$, table 3). Another highly inbreeding species, Glycine soja, from which soybean was domesticated, also has low levels of diversity ($\pi_{total} = 0.0022$, Hyten et al. 2006). But *dicoccoides* also has a lower level of diversity than the mean observed in a survey of selfing species (mean $\pi_{total} = 0.006$, Glémin et al. 2006). There may be several reasons for this. First, dicoccoides arose through a relatively recent allopolyploidy event that may have resulted in a large decrease in diversity in the new species with respect to its diploid ancestors. As the nucleotide mutation rate is low, it is likely to take a long time for diversity to be restored through mutation (Lande and Barrowclough 1987). Thus, the mutation-drift equilibrium may not yet have been reached in *dicoccoides*. The small effective population size of the current population of *dicoccoides* may also account for the low level of diversity. Using microsatellite markers and assuming mutationdrift equilibrium, Thuillet et al. (2005) estimated N_e at 32,500 for *dicoccoides*. The spread of agriculture might have restricted the range of *dicoccoides*, potentially accounting for this low effective population size.

Consequences of Domestication History for DNA Sequences: A Drastic Loss of Diversity

Nucleotide diversity levels were found to be much lower in the 3 cultivated forms than in the wild pool. Assuming that our sample of *dicoccoides* accurately reflects the diversity of the wild progenitor of cultivated wheat 12,000 years ago, initial diversity was reduced by 69% in *aestivum* and 84% in *durum*. Considering the Central East group alone, the diversity reduction associated with domestication is a bit lower ($L\pi_{total} = 67\%$ from Central East population against 70% from the whole wild sample). The increase in Tajima's *D* from *dicoccoides* (D = -0.77, P < 0.001) to domesticated wheats (D = -0.55, P =0.015; D = -0.45, not significant; and D = 0.48, P =0.041 for *dicoccum*, *durum*, and *aestivum*, respectively) is also a signature of a recent bottleneck (Tajima 1989), as observed in maize (Wright et al. 2005).

Major losses of neutral diversity have already been demonstrated in the history of *T. turgidum* ssp. by microsatellite analysis (Thuillet et al. 2005). Our coalescent simulations suggest the average domestication bottleneck intensity (from *dicoccoides* to *dicoccum*) of about 3.15,

equivalent to an effective population size of 10,317 in *dicoccum*, assuming that the N_e of *dicoccoides* is 32,500. The nucleotide diversity in bread wheat could be accounted for a bottleneck intensity of 4.2 (corresponding to an N_e of 7,738). Durum wheat experienced the most severe bottleneck ($N_e = 5,575$), with a population size about one sixth that of wild *dicoccoides*. Using microsatellites, the N_e of domesticated emmer, *dicoccum*, was estimated at 12,000. These 2 estimates of the intensity of the domestication bottleneck in *dicoccum* are qualitatively similar, but the loss of diversity is somewhat greater when estimated with sequence data than with microsatellites, as also reported in a recent study of sunflower domestication (Liu and Burke 2006).

The loss of nucleotide diversity (total and silent) we found during domestication is one of the largest reported so far for a crop species (table 3). Most crops have nucleotide diversities about 30% lower than that of their wild progenitor. However, it is worth noting that wheat and barley lost high and similar amount of silent diversity, 65% and between 57% and 73%, respectively (Caldwell et al. 2006; Kilian et al. 2006).

After domestication, subspecies *durum* and *aestivum* were subject to additional selective events during the evolution of landraces and modern breeding. In durum wheat, 84% of the nucleotide diversity originally present in *dicoccoides* has been lost, with only 20 of the 212 SNPs identified in the wild *dicoccoides* segregating in elite varieties.

Previous studies have reported the existence of at least 2 genetically different progenitors of the D genome of aestivum, suggesting independent polyploidization events (Dvorak et al. 1998; Giles and Brown 2006). Nucleotide diversity has been reported to be 30 times higher in Ae. tauschii than in the D genome of T. aestivum (Caldwell et al. 2004). Thus, if only a few Ae. tauschii individuals were involved in the creation of *aestivum*, then only a few tetraploid progenitors are likely to have been involved in the founding of amphiploids. Two groups of aestivum (marked as I and II on fig. 2) were identified in the phylogenetic tree reconstructed from nucleotide diversity in the A and B genomes and were confirmed by the STRUCTURE analysis. Positive Tajima's D (D = 0.48357, P = 0.041) also indicates possible population subdivision. These findings support a diphyletic (at least) origin for bread wheat involving genetically different tetraploid progenitors (AABB genome). As *aestivum* is believed to have arisen from rare intergeneric crosses between cultivated tetraploid wheat and the wild diploid Ae. tauschii. Thus, it is surprising that di*coccum* does not include significantly more diversity than aestivum. Recurrent gene flow between the tetraploid and hexaploid forms after the emergence of hexaploid forms would have been required to restore the level of diversity of the A and B genomes of *aestivum* after polyploidization. Indirect measurements of sequence polymorphism based on restriction fragment length polymorphism have already suggested the existence of gene flow from parental species to polyploids, especially from dicoccoides (Dvorak et al. 2006). As dicoccum sequences are widely distributed throughout the tree and span the whole range of diversity found in aestivum, we also suggest that gene flow occurred between neighboring dicoccum populations.

Using the Demographic Model to Detect Selection

In theory, diversity surveys for identifying selected genes can be applied to any domesticated animal or plant. However, the power of such approaches depends on the relative levels and patterns of diversity for neutral and selected genes in the wild taxon. If neutral genes retain very little diversity after domestication, it is difficult to discriminate neutral from selected genes (Yamasaki et al. 2005). Bottleneck intensity cannot be estimated for nonpolymorphic genes in the cultivated population. However, such genes may be good candidates for selection.

Wright et al. (2005) estimated that 2-4% of maize genes were subject to selection during maize domestication. We found evidence for a similar proportion in wheat domestication, with only 1 of the 21 loci analyzed presenting a pattern of diversity loss suggestive of selection. Although the evidence for possible selection acting on HgA should be interpreted with caution, this locus presents a striking pattern of polymorphism (fig. 5). It has been annotated as a hypothetical gene located in the "hardness" locus, about 30 kb from Gsp1A (Chantret et al. 2005). Gsp1A is thought to be involved in controlling grain softness (Morris 2002), an important trait for end-use quality in wheat. No polymorphism was observed for Gsp1A in the durum group and for Gsp1B in the aestivum group, whereas these 2 genes harbor 26 and 7 polymorphic sites, respectively, in the wild dicoccoides. The lack of diversity in 1 of the 2 copies in both cultivated wheats suggests that this gene may have been the target of selection during domestication. The HgA-linked locus may have been subject to hitchhiking during selection, but further investigations of this candidate region are required to confirm this hypothesis. The authors would like to draw the reader's attention to a recent study of (Luo et al. 2007) where the wild emmer population structure is analyzed on the basis of the restriction fragment length polymorphism at 131 loci.

Supplementary Material

Supplementary tables S1, S2, S3, and S5 and figure S4 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Note Added in Proof

The authors would like to draw the reader's attention to a recent study of Luo et al. where the wild emmer population structure is analyzed on the basis of the restriction fragment length polymorphism at 131 loci. (Luo M-C, Yang Z-L, You FM, Kawahara T, Waines JG, Dvorak J. 2007. The structure of wild and domesticated emmer wheat populations, gene flow between them, and the site of emmer domestication. Theor. Appl. Genet. 114:947–959).

Acknowledgments

The authors thank Y. Vigouroux, N. Chantret, S. De Mita, and A. Tsitrone for discussion as well as F. Salamini and 2 anonymous reviewers for helpful comments on an earlier version of the manuscript. This work was supported by a grant from Tritipol program founded by Institut National de la Recherche Agronomique and the Bureau des Ressources Génétiques.

Literature Cited

- Bar-Yosef O. 1998. The Natufian culture in the Levant, threshold to the origins of agriculture. Evol Anthropol. 6:159–177.
- Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F. 2004. GENETIX 4.05, logiciel sous Windows™ pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5171. Université de Montpellier II, Montpellier.
- Buckler E, Thornsberry JM, Kresovich S. 2001. Molecular diversity, structure and domestication of grasses. Genet Res. 77:213–218.
- Caldwell KS, Dvorak J, Lagudah ES, Akhunov E, Luo MC, Wolters P, Powell W. 2004. Sequence polymorphism in polyploid wheat and their d-genome diploid ancestor. Genetics. 167:941–947.
- Caldwell KS, Russell J, Langridge P, Powell W. 2006. Extreme population-dependent linkage disequilibrium detected in an inbreeding plant species, *Hordeum vulgare*. Genetics. 172: 557–567.
- Chantret N, Salse J, Sabot F, et al. (19 co-authors). 2005. Molecular basis of evolutionary events that shaped the *hardness* locus in diploid and polyploid wheat species (*Triticum* and *Aegilops*). Plant Cell. 17:1033–1045.
- Charlesworth D, Charlesworth B. 1995. Quantitative genetics in plants—the effect of the breeding system on genetic variability. Evolution. 49:911–920.
- Charlesworth D, Wright SI. 2001. Breeding systems and genome evolution. Curr Opin Genet Dev. 11:685–690.
- Daud HM, Gustafson JP. 1996. Molecular evidence for Triticum speltoides as a B-genome progenitor of wheat (*Triticum aestivum*). Genome. 39:543–548.
- David JL, Bataillon T, Poirier S, Roumet P, Santoni S, Thuillet A-C. 2003. Impact of demographic and selective events on the current genetic diversity of the *Triticum turgidum* complex. In: Pogna NE, Romano M, Pogna EA, Galterio G, editors. Tenth International Wheat Genetics Symposium, S.I.M.I., Paestum, Italy. p. 7–10.
- Dvorak J, Akhunov ED. 2005. Tempos of gene locus deletions and duplications and their relationship to recombination rate during diploid and polyploid evolution in the *Aegilops*-*Triticum* alliance. Genetics. 171:323–332.
- Dvorak J, Akhunov ED, Akhunov AR, Deal KR, Luo MC. 2006. Molecular characterization of a diagnostic DNA marker for domesticated tetraploid wheat provides evidence for gene flow from wild tetraploid wheat to hexaploid wheat. Mol Biol Evol. 23:1386–1396.
- Dvorak J, Diterlizzi P, Zhang HB, Resta P. 1993. The evolution of polyploid wheats—identification of the A-genome donor species. Genome. 36:21–31.
- Dvorak J, Luo M-C, Yang Z-L, Zhang H-B. 1998. The structure of the *Aegilops tauschii* genepool and the evolution of hexaploid wheat. Theor Appl Genet. 97:657–670.
- hexaploid wheat. Theor Appl Genet. 97:657–670. Dvorak J, Zhang HB. 1990. Variation in repeated nucleotide sequences sheds light on the phylogeny of the wheat B and G genomes. Proc Natl Acad Sci USA. 87:9640–9644.
- Eyre-Walker A, Gaut RL, Hilton H, Feldman DL, Gaut BS. 1998. Investigation of the bottleneck leading to the domestication of maize. Proc Natl Acad Sci USA. 95:4441– 4446.
- Felsenstein J. 2005. PHYLIP (phylogeny inference package) version 3.6. Seattle (WA): Department of Genome Sciences, University of Washington.

- Gaut BS, Clegg MT. 1993. Nucleotide polymorphism in the Adh1 locus of pearl-millet (Pennisetum-Glaucum) (Poaceae). Genetics. 135:1091–1097.
- Giles RJ, Brown TA. 2006. GluDy allele variations in *Aegilops tauschii* and *Triticum aestivum*: implications for the origins of hexaploid wheats. Theor Appl Genet. 112:1563–1572.
- Glémin S, Bazin E, Charlesworth D. 2006. Impact of mating systems on patterns of sequence polymorphism in flowering plants. Proc R Soc Lond B Biol Sci. 273:3011–3019.
- Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol. 52:696–704.
- Harlan JR. 1992. Crops and man. Foundations for Modern Crop Science Series. Madison (WI): American Society of Agronomy and Crop Science Society of America.
- Heun M, SchaferPregl R, Klawan D, Castagna R, Accerbi M, Borghi B, Salamini F. 1997. Site of einkorn wheat domestication identified by DNA fingerprinting. Science. 278:1312–1314.
- Huang S, Sirikhachornkit A, Su XJ, Faris J, Gill B, Haselkorn R, Gornicki P. 2002. Genes encoding plastid acetyl-CoA carboxylase and 3-phosphoglycerate kinase of the *Triticum/ Aegilops* complex and the evolutionary history of polyploid wheat. Proc Natl Acad Sci USA. 99:8133–8138.
- Hudson RR. 1990. Gene genealogies and the coalescent process. In: Futuyma DJ, Antonovics J, editors. Oxford surveys in evolutionary biology. New York: Oxford University Press. p. 1–44.
- Hudson RR. 2001. Two-locus sampling distributions and their application. Genetics. 159:1805–1817.
- Hudson RR. 2002. Generating samples under a Wright-Fisher neutral model of genetic variation. Bioinformatics. 18: 337–338.
- Hyten DL, Song Q, Zhu Y, Choi I-Y, Nelson RL, Costa JM, Specht JE, Shoemaker RC, Cregan PB. 2006. Impacts of genetic bottlenecks on soybean genome diversity. Proc Natl Acad Sci USA. 103:16666–16671.
- Isidore E, Scherrer B, Chalhoub B, Feuillet C, Keller B. 2005. Ancient haplotypes resulting from extensive molecular rearrangements in the wheat A genome have been maintained in species of three different ploidy levels. Genome Res. 15:526–536.
- Khlestkina EK, Salina EA. 2001. Genome-specific markers of tetraploid wheats and their putative diploid progenitor species. Plant Breed. 120:227–232.
- Kihara H. 1944. Discovery of the DD-analyser, one of the ancestors of *Triticum vulgare*. Agric Hortic (Tokyo). 19: 13–14.
- Kilian B, Ozkan H, Kohl J, von Haeseler A, Barale F, Deusch O, Brandolini A, Yucel C, Martin W, Salamini F. 2006. Haplotype structure at seven barley genes: relevance to gene pool bottlenecks, phylogeny of ear type and site of barley domestication. Mol Genet Genomics. 276:230–241.
- Kimura M, Ohta T. 1971. Theoretical topics in population genetics. Princeton (NJ): Princeton University Press.
- Lande R, Barrowclough GF. 1987. Effective population size, genetic variation, and their use in population management. Viable populations for conservation. Cambridge: Cambridge University Press. p. 87–123.
- Lev-Yadun S, Gopher A, Abbo S. 2000. Archaeology—the cradle of agriculture. Science. 288:1602–1603.
- Liu A, Burke JM. 2006. Patterns of nucleotide diversity in wild and cultivated sunflower. Genetics. 173:321–330.
- Luo M-C, Yang Z-L, You FM, Kawahara T, Waines JG, Dvorak J. 2007. The structure of wild and domesticated emmer wheat populations, gene flow between them, and the site of emmer domestication. Theor Appl Genet. 114:947–959.

- McFadden ES, Sears ER. 1946. The origin of *Triticum spelta* and its free-threshing hexaploid relatives. J Hered. 37: 81–89.
- Mori N, Liu YG, Tsunewaki K. 1995. Wheat Phylogeny Determined by Rflp Analysis of Nuclear-DNA. 2. Wild Tetraploid Wheats. Theor Appl Genet. 90:129–134.
- Mori N. 2003. Wheat domestication: when, where and how? Plant Cell Physiol. 44:S2–S2.
- Morris CF. 2002. Puroindolines: the molecular genetic basis of wheat grain hardness. Plant Mol Biol. 48:633–647.
- Muller M-H, Poncet C, Prosperi JM, Santoni S, Ronfort J. 2006. Domestication history in the *Medicago sativa* species complex: inferences from nuclear sequence polymorphism. Mol Ecol. 15:1589–1602.
- Nesbitt M, Samuel D. 1996. From staple crop extinction? The archaeology and history of the hulled wheats. In: Padulosi S, Hammer K, Heller J, editors. Hulled wheat. Promoting the conservation and use of underutilized and neglected crops. Rome (Italy): International Plant Genetic Resources Institute. p. 41–100.
- Nesbitt M, Samuel D. 1998. Wheat domestication: archaeobotanical evidence. Science. 279:1433–1433.
- Nordborg M. 2003. Coalescent theory. In: Balding D, Bishop M, Cannings C, editors. Handbook of statistical genetics. Chichester (UK): Wiley. p. 602–635.
- Ozkan H, Brandolini A, Pozzi C, Effgen S, Wunder J, Salamini F. 2005. A reconsideration of the domestication geography of tetraploid wheats. Theor Appl Genet. 110: 1052–1060.
- Ozkan H, Brandolini A, Schafer-Pregl R, Salamini F. 2002. AFLP analysis of a collection of tetraploid wheats indicates the origin of emmer and hard wheat domestication in southeast Turkey. Mol Biol Evol. 19:1797–1801.
- Peakall R, Smouse PE. 2006. Genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes. 6:288–295.
- Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics. 155:945–959.
- Qi LL, Echalier B, Chao S, et al. (47 co-authors). 2004. A chromosome bin map of 16,000 expressed sequence tag loci and distribution of genes among the three genomes of polyploid wheat. Genetics. 168:701–712.
- Ravel C, Praud S, Murigneux A, et al. (12 co-authors). 2006. Single-nucleotide polymorphism frequency in a set of selected lines of bread wheat (*Triticum aestivum* L). Genome. 49: 1131–1139.
- Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics. 19:2496–2497.
- Salamini F, Ozkan H, Brandolini A, Schafer-Pregl R, Martin W. 2002. Genetics and geography of wild cereal domestication in the near east. Nat Genet. 3:429–441.
- Sasanuma T, Chabane K, Endo TR, Valkoun J. 2002. Genetic diversity of wheat wild relatives in the near east detected by AFLP. Euphytica. 127:81–93.

- Septiningsih EM, Prasetiyono J, Lubis E, Tai TH, Tjubaryat T, Moeljopawiro S, McCouch SR. 2003. Identification of quantitative trait loci for yield and yield components in an advanced backcross population derived from the *Oryza sativa* variety IR64 and the wild relative O-rufipogon. Theor Appl Genet. 107:1419–1432.
- Sourdille P, Perretant MR, Charmet G, Leroy P, Gautier MF, Joudrier P, Nelson JC, Sorrells ME, Bernard M. 1996. Linkage between RSLP markers and genes affecting kernel hardness in wheat. Theor Appl Genet. 93:580–586.
- Staden R, Judge DP, Bonfield JK. 2001. Sequence assembly and finishing methods. Methods Biochem Anal. 43:303–322.
- Tajima F. 1983. Evolutionary relationship of DNA sequences in finite populations. Genetics. 105:437–460.
- Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 123: 585–595.
- Tanksley SD, McCouch SR. 1997. Seed banks and molecular maps: unlocking genetic potential from the wild. Science. 277:1063–1066.
- Tanno K, Willcox G. 2006. How fast was wild wheat domesticated? Science. 311:1886–1886.
- Tenaillon MI, U'Ren J, Tenaillon O, Gaut BS. 2004. Selection versus demography: a multilocus investigation of the domestication process in maize. Mol Biol Evol. 21:1214–1225.
- Thuillet A-C, Bataillon T, Poirier S, Santoni S, David JL. 2005. Estimation of long-term effective population sizes through the history of durum wheat using microsatellite data. Genetics. 169:1589–1599.
- Vigouroux Y, Jaqueth JS, Matsuoka Y, Smith OS, Beavis WD, Smith JSC, Doebley J. 2002. Rate and pattern of mutation at microsatellite loci in maize. Mol Biol Evol. 19:1251–1260.
- Wakeley J, Hey J. 1997. Estimating ancestral population parameters. Genetics. 145:847–855.
- Weiss G, von Haeseler A. 1998. Inference of population history using a likelihood approach. Genetics. 149:1539–1546.
- Wright SI, Bi IV, Schroeder SG, Yamasaki M, Doebley JF, McMullen MD, Gaut BS. 2005. The effects of artificial selection on the maize genome. Science. 308:1310–1314.
- Yamasaki M, Tenaillon MI, Bi IV, Schroeder SG, Sanchez-Villeda H, Doebley JF, Gaut BS, McMullen MD. 2005. A large-scale screen for artificial selection in maize identifies candidate agronomic loci for domestication and crop improvement. Plant Cell. 17:2859–2872.
- Zohary D. 1999. Monophyletic vs. polyphyletic origin of the crops on which agriculture was founded in the near east. Genet Resour Crop Evol. 46:133–142.
- Zohary D, Hopf M. 2000. Domestication of plants in the Old World: the origin and spread of cultivated plants in West Asia Europe and the Nile Valley. New York: Oxford University Press.

William Martin, Associate Editor

Accepted March 27, 2007
2.3. Discussion

2.3.1. Importance de la connaissance de la population sauvage

Afin d'étudier le processus de domestication d'une espèce cultivée, un point primordial consiste à identifier l'espèce sauvage apparentée, à l'origine de la domestication. Chez le riz *Oryza sativa*, les espèces sauvages asiatiques *O. rufipogon* et *O. nivara* sont toutes deux considérées comme les progéniteurs de l'espèce cultivée. Ne pouvant pas identifier laquelle de ces espèces est l'espèce sauvage à l'origine de la domestication, Zhu et al. (2007) ont représenté le compartiment sauvage par un échantillon mixte de *O. rufipogon* et *O. nivara*. Un tel complexe d'espèces s'hybridant illustre la difficulté de dissocier les évènements de domestication et d'introgression.

Une fois l'apparentée sauvage la plus proche de l'espèce cultivée identifiée, on pose généralement l'hypothèse que celle-ci peut servir de référence de la population ancestrale. La caractérisation des paramètres du goulot d'étranglement de domestication passe par l'étude comparative des patrons de diversité au sein des compartiments sauvage et cultivé. En utilisant l'espèce sauvage actuelle pour estimer la diversité de la population ancestrale, on fait deux hypothèses :

- la population sauvage est à l'équilibre mutation-dérive. La diversité θ peut alors être correctement approximée par les estimateurs π ou S pour paramétrer les simulations de coalescence générant des échantillons sous un modèle neutre Wright-Fisher.
- la population sauvage n'a pas connu de changements démographiques depuis la domestication, son niveau de diversité est resté constant.

De plus, le patron de diversité intra spécifique de l'espèce sauvage peut être affecté par une subdivision de la population, entraînant une répartition hétérogène de la diversité entre les sous populations (Figure 2.5.).



Figure 2.5. : Représentation schématique d'un épisode de domestication unique à partir d'une population sauvage structurée.

L'espèce sauvage est représentée en gris, les traits pointillés délimitent les sous-populations. Les figures géométriques illustrent la répartition de la diversité dans la métapopulation. L'échantillonnage des gènes lors de la formation de l'espèce cultivée (en noir) se fait au sein d'une seule sous population sauvage (flèche).

Dans le cas d'un évènement unique de domestication, une seule des sous-populations sauvages, et par conséquent un pool de gènes restreint par rapport à la diversité de l'espèce, va participer à la formation de l'espèce cultivée (Figure 2.5.). Comment quantifier le goulot d'étranglement lié à la domestication : en estimant la diversité ancestrale d'après l'espèce ou d'après la sous-population ?

Moeller et al. (2007) ont montré dans une étude récente sur la téosinte (espèce sauvage apparentée au maïs) que la méthode d'échantillonnage au sein d'une espèce fortement structurée géographiquement pouvait affecter l'estimation de la diversité. Si un seul individu est échantillonné par sous-population (ce qui est généralement le cas quand on cherche à représenter la diversité de toute l'espèce Wright and Gaut 2005), le spectre de fréquence des mutations sera biaisé vers un excès de variants rares (D_{Tajima} négatif). Le niveau de diversité (estimé par π et θ_W) variant d'une population à l'autre, deux cas ont été observés :

- *i*. la diversité intra population est similaire à la diversité de l'espèce. Dans ce cas un échantillonnage à l'échelle de l'espèce permet de faire des inférences sur l'intensité du goulot d'étranglement de domestication.
- *ii.* la diversité de la sous-population est inférieure à la diversité de l'espèce. Un échantillonnage à l'échelle de l'espèce entraîne alors la surestimation de la diversité du pool sauvage ancestral et par conséquent l'intensité du goulot d'étranglement. Dans ce

cas, on peut considérer deux intensités de goulot d'étranglement ayant des significations différentes. Un goulot d'étranglement global estimé à partir de l'ensemble de la diversité de départ est révélateur de la perte de diversité entraînée par la domestication entre l'espèce sauvage et cultivée. Mais si la question à laquelle on s'intéresse est de préciser la dynamique de la domestication, il convient alors d'estimer un goulot d'étranglement local, entre la sous-population sauvage à l'origine de la domestication et la forme cultivée.

Dans un cas de forte structure où la sous-population sauvage à l'origine de l'espèce domestiquée a un niveau de diversité inférieur à celui de l'espèce sauvage, il faut identifier la structure géographique afin d'intégrer la subdivision et les flux géniques au modèle de coalescence, comme l'on fait chez le riz Caicedo et al. (2007).

Une structure géographique a été identifiée au sein de la sous espèce T. t. dicoccoides avec deux principales populations (Ozkan et al. 2005), encore subdivisées (Luo et al. 2007). Dans notre étude (Haudry et al. 2007), l'échantillon sauvage de T. t. dicoccoides a été choisi afin d'en maximiser la diversité. Bien qu'aucune structure n'ait été détectée dans cet échantillon, le D_{Tajima} calculé sur l'ensemble des locus est significativement négatif ($D_{\text{Tajima}} = -0.76762$, P <0.001), pouvant être la marque de l'échantillonnage. Afin de comparer la diversité de chaque population à celle mesurée sur l'échantillon global, les accessions de notre échantillon ont été réparties dans deux groupes géographiques définis d'après Ozkan et al. (2005). La diversité nucléotidique mesurée était similaire dans le groupe Est ($\pi = 0.0024$) et le groupe Ouest (π =0.0027) ainsi que dans l'échantillon global ($\pi = 0.0027$). Ces résultats suggèrent que la structuration génétique de T. t. dicoccoides est faible, et qu'un échantillon représentant la diversité de toute cette sous-espèce est approprié pour faire des inférences sur le goulot d'étranglement de domestication du blé. Pour confirmation, les intensités de goulots d'étranglement ont été ré-estimées à partir de l'échantillon Est de T. t. dicoccoides, population qui serait à l'origine de la domestication (Ozkan et al. 2005). Ces estimations sont semblables à celles réalisées à partir de l'échantillon global (les intervalles de confiance sont précisées entre crochets):

Intensités des	dicoccoides -	dicoccoides	dicoccoides -
goulots	dicoccum	- durum	aestivum
échantillon global	3.15	5.83	4.20
	[2.07 - 4.53]	[4.35 - 7.94]	[3.10 - 5.74]
échantillon Est	2.61	5.45	4.53
	[1.06 - 4.60]	[3.54 - 8.65]	[3.06 - 6.63]

2.3.2. Un modèle alternatif de détection de sélection

L'une des fortes motivations des études du processus de domestication est d'identifier des gènes directement impliqués dans la transition de l'état sauvage à l'état domestiqué. Deux types d'approches peuvent être retenus dans ce but : partir du phénotype d'intérêt et rechercher ses bases génétiques ou inversement partir d'un génotype puis identifier le phénotype pour lequel il code. Pour l'heure, la plupart des études ont suivi l'approche conventionnelle, allant du phénotype au génotype (Figure 2.6. gauche), identifiant les régions génomiques impliquées dans le caractère par des méthodes de génétique quantitative (recherche de QTL, quantitative trait loci) ou de génétique d'association « DL mapping » (revue Rafalski and Morgante 2004). Une approche alternative a plus récemment vu le jour : elle consiste à utiliser les méthodes de génétique des populations pour rechercher un signal génomique d'adaptation (Figure 2.6. droite). C'est cette démarche appliquée à l'étude des plantes cultivées (utilisée dans notre étude de l'impact de la domestication du blé sur la diversité des génomes) que je vais développer ici, les apports de cette méthode dans la détection de gènes sous sélection, sa puissance et ses limites.



Figure 2.6.: Représentation schématique des deux types d'approches pour relier phénotype et génotype

A gauche, la démarche « de haut en bas » qui part d'un phénotype et cherche à identifier le gène impliqué. A droite, la démarche « du bas vers le haut » se basant sur une signature de sélection au niveau moléculaire et va vers l'identification du caractère en jeu.

La logique de cette approche est simple : si un gène a été la cible de sélection directionnelle puisqu'il influençait un caractère agronomique, il doit montrer une réduction de diversité nucléotidique ainsi qu'une augmentation du déséquilibre de liaison au sein du compartiment cultivé par rapport à la population sauvage (paragraphe 1.2.3.). En plus d'engendrer une perte de diversité, la sélection positive affecte la distribution des fréquences des polymorphismes au niveau du gène cible même et des régions génétiquement liées (Maynard Smith and Haigh 1974). Ce type de signature de sélection peut être détectée au niveau nucléotidique par plusieurs tests de neutralité, par exemple, le D_{Tajima} (Tajima 1983). Dans un scénario de sélection positive, le balayage sélectif va entraîner soit un déficit (balayage sélectif complet, sans recombinaison) soit un excès (balayage sélectif incomplet, au moins un évènement de recombinaison) de mutations présentes en fréquence intermédiaire (se traduisant respectivement par un D négatif et un D positif, Encadré 3). Or la sélection n'est pas la seule force évolutive qui peut biaiser la distribution de la diversité : un goulot d'étranglement récent entraîne la perte de mutations en faible fréquence, par dérive. Puisqu'un goulot d'étranglement peut mimer les effets de la sélection sur le patron de polymorphisme de séquence au niveau d'un locus, il convient de définir préalablement un modèle démographique le plus en accord possible avec les principales étapes de l'histoire évolutive de la population. Il est en effet impossible d'identifier l'ensemble des évènements qui ont abouti à la transition sauvage-cultivée en tenant compte de l'ensemble des facteurs temporels et géographiques brouillés eux aussi par un facteur anthropique très efficace dans sa capacité migratoire. Grâce à l'utilisation d'un large nombre de locus, et sous l'hypothèse que la plupart sont neutres (du moins vis-à-vis de la domestication), le cadre démographique moyen du goulot d'étranglement de domestication peut être défini de manière empirique. Il devient alors possible de rechercher les effets de la sélection dans les déviations observées à ce modèle démographique.

On a vu que la sélection sur les caractères liés à la domestication peut être modélisée comme un goulot d'étranglement de plus forte intensité (chap. 1.2.3.). Afin de définir des locus candidats, on peut rechercher des locus présentant une intensité de goulot d'étranglement significativement plus fort que le goulot de domestication (Haudry et al. 2007) ou, si le nombre de locus analysés est suffisant, utiliser un modèle emboîté au modèle neutre démographique, avec deux paramètres d'intensité de goulot d'étranglement : une de domestication et une de sélection (Wright et al. 2005). Les gènes ne présentant qu'un seul haplotype dans le compartiment cultivé peuvent également être considérés comme « hors-types » (Thornton and Jensen 2007). Cependant, l'absence de diversité peut être uniquement

le résultat de la dérive, notamment si le niveau de diversité de la population initiale est bas ou le goulot d'étranglement démographique très sévère.

Cette méthode de détection de gènes sous sélection peut en théorie s'appliquer à n'importe quelle espèce domestiquée (animale ou végétale) mais peut être étendue à une espèce ayant subi un goulot d'étranglement récent, pour laquelle une population représentant la diversité de la population ancestrale est identifiée. Si ce type d'approche ne sert pas à identifier les gènes impliqués dans l'expression d'un phénotype particulier (préférer dans ce cas l'approche « de haut en bas»), il a permis l'identification de 30 gènes candidats chez le maïs (parmi 774 analysés), co-localisant avec des régions QTL (Wright et al. 2005). Sa puissance de détection va varier selon 4 principaux critères :

- le niveau de diversité. Si les gènes « neutres » contiennent peu de diversité avant et, a fortiori, après la domestication, il est difficile de discriminer les gènes sous sélection des gènes neutres (Yamasaki et al. 2005). La situation est extrême lorsque les gènes « neutres » ont perdu tout polymorphisme par dérive lors de la domestication, ils n'apportent alors pas d'informations pour l'estimation du scénario démographique et peuvent être considérés comme candidats à la sélection, à tort. Plus l'espèce sauvage présente un important niveau de diversité, plus facile sera la discrimination entre sélection et démographie. Puisque le passage à l'autogamie s'accompagne d'une réduction de *Ne* (chap. 1.3.), les espèces autogames sont généralement attendues moins polymorphes que les espèces allogames. L'estimation des paramètres du goulot d'étranglement de domestication et la détection de gènes sous sélection pourraient par conséquent être moins puissante pour les espèces autogames. Chez une espèce avec aussi peu de diversité que le blé, il faudrait un très grand nombre de gènes pour réduire la variance qu'il y a dans l'estimation de l'intensité de goulot d'étranglement.
- la fréquence de l'allèle favorable p dans la population ancestrale. L'allèle soumis à sélection lors de la domestication pouvant être neutre vis-à-vis de la sélection naturelle, sa fréquence n'était pas forcément faible dans la population sauvage ancestrale. Or la fréquence d'un allèle neutre dans une population est fonction du temps depuis son apparition et de la taille efficace (Kimura and Ohta 1971). Innan et Kim (2004) ont montré que la signature de sélection diminue quand p augmente : l'allèle ségrégeant depuis longtemps dans la population et tendant vers l'équilibre de liaison. Plus p est grand, moins la sélection lors de la domestication n'entraînera d'autres sites que le site cible par balayage sélectif, réduisant le signal génomique. Dans le cas de la domestication, l'allèle domestiqué peut aussi être initialement délétère et donc en faible fréquence dans la

population sauvage. C'est par exemple le cas de l'allèle rachis cassant qui est sans doute délétère en population naturelle. Dans ce cas, on s'attend à une signature plus forte.

- le nombre de gènes utilisés. La fonction de vraisemblance multilocus permettant l'estimation de l'intensité de goulot d'étranglement aura une variance d'autant plus faible qu'un grand nombre de locus sont utilisés pour la calculer, et que le ratio de gènes neutres/sélectionnés est grand, par exemple 24 chez le maïs (96% de gènes neutres pour 4% de gènes sous sélection, Wright et al. 2005). Il est préférable de concentrer les efforts de séquençage sur le nombre de locus, de préférence assez long (2500pb plutôt que 500pb, Thornton and Jensen 2007), tandis qu'une dizaine d'accessions est suffisante pour estimer la diversité au sein d'une population (Felsenstein 2006). Sur une espèce avec aussi peu de diversité que le blé, un jeu d'un millier de gènes serait plus approprié d'une part pour estimer de manière plus précise le scénario démographique, d'autre part pour faire une large détection de gènes potentiellement impliqués dans la domestication.
- les statistiques résumant la diversité. L'efficacité de la détection de locus « hors-type » après intégration du modèle démographique serait meilleure pour des statistiques résumant directement la diversité tels que π ou S plutôt que celles mesurant le biais dans la distribution des fréquences des polymorphismes, telles que le D_{Tajima} (Teshima, Coop, and Przeworski 2006; Thornton and Jensen 2007). Dans notre étude, nous avons estimé la vraisemblance des scénarios explorés comme la proportion de simulations qui s'ajustaient le mieux à la fois aux valeurs de π et de S observés. S'il reste suffisamment de sites polymorphes dans les compartiments cultivés, un estimateur du déséquilibre de liaison peut également être utilisé pour ajuster le modèle (Tenaillon et al. 2004).

Il est important de noter que les gènes candidats ainsi identifiés peuvent ne pas être les cibles directes de la sélection, mais avoir été entraînés par un balayage sélectif. Une vérification fonctionnelle reste nécessaire pour valider la relation phénotype-génotype, de la même façon que dans une approche du type « top-down ». Si le déséquilibre de liaison entre gènes entraîne la détection de faux positifs (gènes neutres entraînés), le déséquilibre de liaison intragénique responsable de l'entraînement autour du site sous sélection est nécessaire à la détection du gène sélectionné. En effet, si la fixation du site sous sélection n'entraîne pas celle des sites polymorphes neutres voisins, la perte de diversité au niveau du gène reflètera uniquement le goulot d'étranglement démographique. La durée de la domestication avec des flux géniques réguliers entre les formes sauvages et cultivées en fonction de leur régime de reproduction affecte aussi la puissance de détection des gènes soumis à sélection lors de la

domestication. Plus la durée est importante, d'autant plus chez une espèce allogame, plus des flux de gènes entre sauvage et domestiqué augmentent le niveau de diversité des locus neutres au sein de la forme sauvage (Papa et al. 2005). Les flux de gènes lors de la domestication réduisent l'intensité du goulot d'étranglement démographique, les gènes soumis à sélection sont alors plus facilement détectés (Yamasaki et al. 2005).

L'épisode de domestication du blé semble s'être étalé sur plus d'un millénaire (Tanno and Willcox 2006), alors que le passage de grains vêtus à nus aurait été plus violent (Nesbitt and Samuel 1996). Les gènes soumis à sélection lors de la domestication seraient par conséquent plus facilement détectables que ceux impliqués dans le passage au blé dur qui a été plus rapide (Figure 2.1.3.).

Chapitre 3.

Comment le système de reproduction et la recombinaison affectent l'évolution moléculaire chez les *Triticeae* ?

Chapitre 3.

Comment le système de reproduction et la recombinaison affectent l'évolution moléculaire chez les *Triticeae* ?

3.1. Problématique et synthèse des résultats

Nous avons vu en introduction que la théorie prédit que le système de reproduction et le taux de recombinaison local affectent à la fois le patron de diversité, l'efficacité de la sélection et la composition en bases (chap. 1.3.). Un des objectifs de ma thèse consiste à tester ces prédictions sur les génomes des *Triticeae*. A partir de données de séquences codantes orthologues sur cinq espèces, nous avons analysé les effets du régime de reproduction et des variations locales du taux de recombinaison sur (*i*) l'efficacité de la sélection, et (*ii*) l'évolution de la composition en bases.

L'analyse a porté sur les patrons de substitutions de 52 fragments de gènes chez cinq représentants diploïdes des *Triticeae*, pour lesquels les relations phylogénétiques et le régime de reproduction étaient connus (Figure 3.1.) :

- deux espèces fortement autogames proches apparentées *Triticum monococcum* et *Triticum urartu* (génomes A^mA^m et A^uA^u, respectivement)
- deux espèces allogames Secale cereale, et Aegilops speltoides (génomes RR et SS, respectivement) toutes deux auto-incompatibles,
- une espèce autogame servant de groupe externe *Hordeum vulgare* (génome HH),
 qui servira de référence pour inférer les états ancestraux.



— H. vulgare

L'impact du régime de reproduction a été étudié en contrastant les paramètres évolutifs des branches terminales menant aux espèces autogames (en pointillé sur la Figure 3.1.) et à ceux des branches menant aux allogames (en gras sur la Figure 3.1.). Parallèlement à l'effet global du système de reproduction, nous avons comparé des régions génomiques *a priori* fortement et faiblement recombinantes pour comprendre l'impact des variations locales du taux de recombinaison au sein des génomes. Sous l'hypothèse que le gradient de recombinaison observé le long des chromosomes de blé est valide sur l'ensemble des *Triticeae*, le taux de recombinaison est attendu beaucoup plus important dans les régions sub-télomériques que dans les régions centromériques des chromosomes (Figure 1.9.,Lukaszewski 1992; Lukaszewski 1993; Akhunov et al. 2003b). Le patron de recombinaison semble également différer entre le bras court et le bras long du chromosome, avec plus de recombinaison dans le bras court (Lukaszewski 1993).

Parmi les 52 locus que nous avons analysés, 46 ont été choisis afin de couvrir l'ensemble du chromosome 3. Le chromosome 3B du blé a l'avantage de présenter une bonne synténie avec le chromosome 1 du riz (Figure 3.2.A., Munkvold et al. 2004). Par ailleurs, comme l'ensemble du génome des *Triticeae* présente en général une forte synténie, nous faisons l'hypothèse qu'il existe un niveau élevé de conservation de l'ordre et de la position des gènes entre les chromosomes du groupe d'homéologie 3 des espèces étudiées, *i.e.*, 3H, 3R, 3S, 3A^m et 3A^u respectivement pour l'orge, le seigle, *Ae. speltoïdes*, *T. monococcum* et *T. urartu* ainsi qu'avec le chromosome 3B du blé. Grâce à la localisation des gènes soit sur le chromosome 1 du riz, soit sur les fragments chromosomiques (« bins de délétions ») du chromosome 3B du blé tendre, les locus échantillonnés ont été approximativement positionnés le long du chromosome 3 et se répartissent de façon plus ou moins homogène le long du chromosome (Figure 3.2.B).

Nous avons utilisé cette localisation pour repartir les gènes en classes de recombinaison : des groupes à fort taux de recombinaison (*FR*) pour les gènes situés dans la région distale et/ou sur le bras court) et des groupes à plus faible taux de recombinaison (*fr*) pour les gènes situés dans les régions centromériques et /ou sur le bras long). Par la suite, nous appellerons région télomérique la zone chromosomique représentant environ le dernier quart physique du chromosome, inférée à partir de la position des gènes sur le chromosome 1 du riz.



Figure 3.2. : Représentation schématique du chromosome 3B, des bins de délétions et de la localisation des gènes étudiés

La figure A représente la synténie entre le chromosome 1 du riz (à gauche) et le chromome 3B du blé tendre (à droite), à l'échelle des bins de délétions. Chaque bin est représenté par une couleur différente sur le chromosome 3B, la couleur indiquant la région d'homologie sur le chromosome du riz. Figure extraite de Munkvold et al (2004).

La figure B représente la répartition des 38 locus échantillonnés dans notre étude (*) cartographiés dans les bins de délétions (différentes couleurs). Les huit autres locus ont été positionnés sur le chromosome d'après leur localisation sur le riz. Figure adaptée de E.Paux, (com. pers.).

Les principaux niveaux de comparaison des estimations de l'efficacité de la sélection et de l'évolution de la composition en bases qui composent notre analyse sont détaillés dans la Figure 3.3.



Figure 3.3. : Représentation schématique des comparaisons réalisées dans cette étude. Par exemple, les GC^* estimés sur les branches autogames et sur les branches allogames pour l'ensemble des gènes ont été comparés pour évaluer l'impact du système de reproduction sur la composition en bases (comparaison schématisée par la flèche rouge).

FR : régions fortement recombinantes ; fr : régions faiblement recombinantes.

Efficacité de la sélection

La réduction de taille efficace *Ne* attendue dans les **populations autogames** devrait entraîner la diminution de l'efficacité de la sélection, laquelle est une fonction du produit *Nes* (*s* étant le coefficient de sélection). De la même façon, les régions faiblement recombinantes, plus soumises aux effets d'entraînement, devraient montrer une efficacité plus faible de la sélection que les régions fortement recombinantes. Chez les espèces étudiées ici, dans la mesure où la **distribution de la recombinaison** n'est pas homogène, nous pouvons tester si l'efficacité de la sélection dépend de la localisation physique des locus étudiés, en comparant les taux d'accumulation de mutations non synonymes délétères dans les régions faiblement recombinantes par rapport aux régions fortement recombinantes (Chap. 1.3.)

Sous l'hypothèse du modèle presque neutre de l'évolution, la plupart des mutations sont supposées faiblement délétères (Ohta 1992). La diminution de N_e chez les autogames devrait entraîner un relâchement de l'efficacité de la sélection et donc une augmentation de la **proportion de mutations non synonymes** faiblement délétères pouvant se fixer ($Nes \le 1$). Dans le cas de gènes soumis à de la sélection purificatrice, les prédictions théoriques se traduisent par des ratios de divergence non synonyme sur synonyme d_N/d_S (noté également ω) attendus plus forts dans les génomes ou régions faiblement recombinants (Tableau 3.1.). Ces prédictions s'inversent dans le cas de sites soumis à de la sélection positive, pour laquelle les sites non synonymes sont considérés comme avantageux (Tableau 3.1.).

L'efficacité de la sélection peut également se regarder au niveau du **biais sur l'usage du code** qui s'exerce sur les gènes. Parmi les codons redondants pour l'expression d'un acide aminé, certains sont plus fréquemment utilisés que d'autres, et ce d'autant plus dans les gènes fortement exprimés (on parle de codons préférés ou optimaux), possiblement parce qu'ils sont plus efficaces pour la traduction (Akashi, Kliman, and Eyre-Walker 1998). Dans les génomes ou régions génomiques pour lesquelles une réduction de l'efficacité de la sélection est attendue, on s'attend donc à une diminution de fréquence des codons préférés (*Fop*). La fréquence d'équilibre en codons préférés vers laquelle tendent les gènes (*Fop**) peut être estimée d'après la dynamique de transition des codons sur la phylogénie :

$$Fop^* = P/(P+U)$$

où P correspond au taux de substitution de codons non préférés vers codons préférés et U au taux de substitution réciproque.

Nous avons donc cherché à mettre en évidence des changements de niveau d'efficacité de la sélection sur la phylogénie présentée en Figure 3.1., en contrastant les d_N/d_S et les Fop^* des gènes :

- (i) des deux autogames (*T. monococcum* et *T. urartu*) et des deux allogames (*S. cereale* et *Ae. speltoides*), pour tester l'impact du système de reproduction (Tableau 3.1.).
- *(ii)* des régions *fr* et *FR* pour tester l'impact des variations locales du taux de recombinaison (Tableau 3.1.).

Composition en bases

Sous l'hypothèse d'une corrélation positive entre le taux de crossing-over et la BGC, la teneur en GC devait augmenter avec le taux de recombinaison local (paragraphe 1.3.3.). Sur nos données, la composition en bases devrait donc différer entre les régions faiblement et fortement recombinantes. Par ailleurs, l'efficacité de la BGC, est sensible à la fréquence des individus hétérozygotes et dépend donc du taux d'autofécondation de la population (Marais, Charlesworth, and Wright 2004). Nous pouvons tester si des différences d'intensités de BGC s'observent dans notre jeu de données en fonction du régime de reproduction et valider sur un set de gènes orthologues les variations de teneur en GC déjà observées globalement au sein de la famille des *Triticeae* : les espèces autogames de la famille montrent des teneurs en GC plus faibles que les allogames (Akhunov et al. 2003a; Glémin, Bazin, and Charlesworth 2006). De même, si les crossing-overs sont plus nombreux dans la zone télomérique (ou le bras court) que dans la zone centromérique (ou le bras long), l'efficacité de la BGC devrait être contrastée entre ces différentes régions.

La teneur en GC d'équilibre vers laquelle tendent les séquences sous l'influence de la BGC (GC^*) peut être estimée selon le même principe que le Fop^* :

$GC^* = u/(u + v)$

où *u* et *v* correspondent à deux taux de substitution $AT \rightarrow GC$ et $GC \rightarrow AT$, respectivement (Sueoka 1962).

Nous avons donc cherché à mettre en évidence des changements de la dynamique d'enrichissement en GC sur la phylogénie présentée en Figure 3.1., en contrastant les GC^* des gènes :

- (i) des deux autogames (*T. monococcum* et *T. urartu*) et des deux allogames (*S. cereale* et *Ae. speltoides*), pour tester l'impact du système de reproduction (Tableau 3.1.).
- (*ii*) des régions fr et FR pour tester l'impact des variations locales du taux de recombinaison (Tableau 3.1.).

Cependant, l'efficacité de la sélection sur l'usage du code pourrait participer également à l'enrichissement des génomes en GC. En effet, chez le blé et l'orge, les codons préférés identifiés se finissent par une base G ou C (Kawabe and Miyashita 2003; Liu and Xue 2005; Wang and Roossinck 2006). Sous l'hypothèse que le patron de codons préférés soit conservé au sein des *Triticeae*, une diminution de la teneur en GC des gènes peut illustrer une réduction de l'efficacité de la sélection sur l'usage du code, entraînant une confusion possible entre une diminution de la BGC due à la réduction de l'hétérozygotie individuelle et une diminution d'efficacité sur la sélection servés.

Pour résumer, nous avons étudié l'impact du système de reproduction et du taux de recombinaison local sur l'efficacité de la sélection et l'évolution de la composition en bases dans le génome de quatre espèces appartenant à la famille des *Triticeae*. Nous avons cherché à mettre en évidence les attendus théoriques récapitulés dans le Tableau 3.1.

Tableau 3.1. : Résumé des attendus théoriques de l'impact du système de l	reproduction
et la recombinaison sur l'efficacité de la sélection et la teneur en GC.	

Impact du système de reproduction	Impact du taux de recombinaison local
Sur l'efficacité de la sélection	Sur l'efficacité de la sélection
Sélection Purificatrice	Sélection Purificatrice
$1 > \omega$ autogame $> \omega$ allogame	$1 > \omega$ régions faible $r > \omega$ régions fort r
Sélection Positive	Sélection Positive
$1 < \omega$ autogame $< \omega$ allogame	$1 < \omega$ régions faible $r < \omega$ régions fort r
Usage du Code	Usage du Code
<i>Fop</i> * autogame < <i>Fop</i> * allogame	Fop* régions faible $r < Fop*$ régions fort r
Sur la BGC	Sur la BGC

ω:	ratio d_N/d_S ;	GC^* : ter	neur en GC	à l'équilibre	; Fop*:	fréquence	des	codons	optimaux	à
ľé	quilibre ; BG	C : conver	sion géniqu	le biaisée ; r	: taux de r	ecombinai	son l	ocal.	-	

 GC^* régions faible $r < GC^*$ régions fort r

Résultats

 GC^* autogame < GC^* allogame

Aucun effet significatif de relâchement de l'efficacité de la sélection n'a pu être mis en évidence chez les espèces autogames étudiées. Au contraire, les ratios de mutations non synonymes sur synonymes sur l'ensemble des gènes étudiés sont légèrement plus forts chez les espèces allogames ($\omega = 0.150$) que chez les autogames ($\omega = 0.105$) contrairement à ce qui était attendu (Tableau 3.1). Nous n'avons donc pas pu mettre en évidence que les espèces autogames fixent plus de mutations délétères que les allogames sur l'ensemble des gènes, tous considérés en première approche comme soumis à de la sélection purificatrice. Wright et al. (2002) n'avaient pas non plus trouvé de signe d'une accumulation de mutations délétères chez l'espèce autogame *A. thaliana* en la comparant à *A. lyrata*, une espèce allogame apparentée. Cependant, des analyses plus fines indiquent que 3% des codons analysés auraient évolué

sous sélection positive dans les espèces allogames ($\omega = 3.01$), par opposition au cas des espèces autogames pour lesquelles ces codons auraient évolué de façon neutre ($\omega = 1.01$). Il semble donc que la sélection positive soit plus efficace chez *S. cereale* et *Ae. speltoides* que chez *T. monococcum* et *T. urartu*, conformément à l'attendu théorique. Enfin, aucun effet significatif du taux de recombinaison local n'a pu être mis en évidence par la comparaison des d_N/d_S estimés dans les régions fortement et faiblement recombinantes.

Les résultats concernant les dynamiques d'enrichissement en GC en fonction du régime de reproduction et du taux de recombinaison local sont conformes aux prédictions théoriques. Les GC^* estimés sur les branches allogames (0.453) sont plus forts que ceux estimés sur les branches autogames (0.353, p = 0.03). Les prédictions ont également été confirmées à l'échelle intra-génomique. Les gènes localisés dans des régions faiblement recombinantes montrent une dynamique d'enrichissement en GC moins importante que les gènes des régions fortement recombinantes.

Enfin, l'évolution du biais d'usage du code est plus marquée dans les espèces allogames (Fop *= 0.339) que dans les espèces autogames (Fop *= 0.235, p=0.06), de même que dans les régions télomériques (FR, Fop *= 0.384) par rapport aux régions centromériques (fr, Fop *= 0.300, p=0.04).

Les résultats concernant le GC^* et le Fop^* étaient attendus à la fois du fait de la réduction d'efficacité de la sélection sur l'usage du code et de la réduction d'efficacité de la BGC dans ces régions à faible taux de recombinaison. Afin de démêler cette confusion, nous avons pris appui sur une caractéristique qui oppose la BGC et le biais d'usage du code. La BGC est un processus devant favoriser de manière équivalente les bases C et G. En revanche, le biais d'usage du code chez les *Triticeae* tend à favoriser les codons se terminant en C au niveau des codons 4 et 6 fois dégénérés. Dans ce cas, la décomposition de l'enrichissement global en GC (GC^*) en deux composantes distinctes, devrait montrer un enrichissement en C (C^*) plus important que l'enrichissement en G (G^*). Malgré un C^* plus fort chez les espèces allogames ($C^*=0.20$) que chez les espèces autogames ($C^*=0.13$), cette différence n'est pas significative (p=0.2). De plus, le patron d'enrichissement en GC a été observé également en 1^{ère} et 2^{nde} position de codons, ce qui n'est pas attendu sous un modèle de sélection sur l'usage du code. Par conséquent, les différences observées dans la dynamique de la teneur en GC entre génomes de différent système de reproduction et entre régions de différents taux de recombinaison seraient essentiellement dues à la BGC dans notre étude.

Enfin, la BGC pourrait affecter l'efficacité de la sélection. En effet, en répartissant les gènes en deux classes de GC^* (faibles et élevés), nous avons détecté un excès de mutations non synonymes chez les espèces allogames ($\omega = 0.180$) par rapport aux autogames ($\omega = 0.099$) dans les gènes présentant de forts GC^* (p = 0.032). En revanche, les gènes à GC^* faibles ne montrent pas de différence de d_N/d_S entre espèces autogames et allogames (p = 0.664). En affectant non seulement les régions non codantes, mais aussi les régions codantes, la BGC biaisée vers GC pourrait créer un fardeau de mutation chez les espèces allogames en favorisant la fixation de mutations faiblement délétères en G ou C. Récemment, Galtier et Duret (2007) ont proposé que l'accélération des vitesses d'évolution observée dans certaines espèces pourrait ne pas être le fruit de l'adaptation, mais la signature d'une augmentation de l'intensité de la BGC.

Cette hypothèse de fardeau de BGC chez les espèces allogames, si elle se trouve confirmée par des études complémentaires, apporterait un nouvel élément dans la compréhension de l'évolution des systèmes de reproduction.

Le manuscrit reproduit ci-après détaille cette étude et ses résultats.

3.2. Mating system and recombination affect molecular evolution in four Triticeae species

Mating system and recombination affect molecular evolution in four *Triticeae* species

A. HAUDRY^{1,2}, A. CENCI¹, C. GUILHAUMON¹, E. PAUX³, S. POIRIER¹, S. SANTONI¹, J. DAVID¹ and S. GLÉMIN^{2*}

¹ UMR Diversité et Adaptation des Plantes Cultivées, Montpellier SupAgro – INRA – IRD – UMII, 2, Place Pierre Viala, 34060 Montpellier Cedex 1, France

² Institut des Sciences de l'Evolution, Université Montpellier 2, place Eugène Bataillon, Montpellier, France

³ UMR ASP 1095, Université Clermont Ferrand, INRA, F-63100 Clermont Ferrand, France

(Received 26 August 2007 and in revised form 20 October 2007)

Summary

Mating systems and recombination are thought to have a deep impact on the organization and evolution of genomes. Because of the decline in effective population size and the interference between linked loci, the efficacy of selection is expected to be reduced in regions with low recombination rates and in the whole genome of self-fertilizing species. At the molecular level, relaxed selection is expected to result in changes in the rate of protein evolution and the pattern of codon bias. It is increasingly recognized that recombination also affects non-selective processes such as the biased gene conversion towards GC alleles (bGC). Like selection, this kind of meiotic drive in favour of GC over AT alleles is expected to be reduced in weakly recombining regions and genomes. Here, we investigated the effect of mating system and recombination on molecular evolution in four Triticeae species: two outcrossers (Secale cereale and Aegilops speltoides) and two selfers (*Triticum urartu* and *Triticum monococcum*). We found that GC content, possibly driven by bGC, is affected by mating system and recombination as theoretically predicted. Selection efficacy, however, is only weakly affected by mating system and recombination. We investigated the possible reasons for this discrepancy. A surprising one is that, in outcrossing lineages, selection efficacy could be reduced because of high substitution rates in favour of GC alleles. Outcrossers, but not selfers, would thus suffer from a 'GC-induced' genetic load. This result sheds new light on the evolution of mating systems.

1. Introduction

Mating systems are thought to play a major role in genome evolution (Charlesworth & Wright, 2001). They affect the effective population size, N_e , and the efficacy of recombination, which both play a crucial role in molecular evolution by controlling patterns of polymorphism, the efficacy of selection, and specific processes such as biased gene conversion towards GC (Marais *et al.*, 2004). Inbreeding reduces N_e because of non-independent gamete sampling (corresponding to a 50% reduction under complete selfing) (Pollak, 1987) and the efficacy of recombination (Nordborg, 2000), which in turn reduces N_e further through hitchhiking effects due to the recurrent elimination of deleterious alleles, (background selection; Charlesworth *et al.*, 1993), or the spread of advantageous mutations (selective sweeps; Maynard-Smith & Haigh, 1974). Such a reduction in N_e is also expected in regions of low recombination in outcrossing species (Charlesworth *et al.*, 1993). Finally, self-fertilizing species are usually more prone to recurrent bottlenecks (Schoen & Brown, 1991) thanks to their capacity for founding new populations with few seeds or even only one seed (Baker, 1955). In many cases, extinction-recolonization dynamics also reduces local and species-wide N_e (Ingvarsson, 2002).

Because the efficacy of selection mainly depends on the product N_{es} , where s is the selection coefficient,

^{*} Corresponding author. Telephone: (+ 33) 4 67 14 48 18. Fax: (+ 33) 4 67 14 36 10. e-mail: glemin@univ-montp2.fr

highly self-fertilizing species should be less efficient than outcrossers at purging slightly deleterious alleles or fixing new advantageous mutations. At the molecular level, we would expect to observe signatures of relaxed selection in selfers, such as an elevated ratio of non-synonymous over synonymous substitution rates (d_N/d_S ratio) due to the fixation of slightly deleterious mutations, and la ow level of codon bias due to the inefficacy of selection for preferred codons (Akashi, 1995).

While the effect of selfing on polymorphism is now well documented (Glémin et al., 2006; Hamrick & Godt, 1996; Nybom, 2004), its impact on selection efficacy has been only weakly supported. In a metaanalysis on a wide set of plants, Glémin et al. (2006) found a weak signal of relaxed selection both against slightly deleterious alleles and in favour of new advantageous mutations in self-fertilizing species, compared with outcrossing ones. Wright et al. (2002), however, did not find any difference between the selfer Arabidopsis thaliana and its outcrossing close relative Arabidopsis lyrata, either in the rate of protein evolution or in codon bias. Patterns of selection, especially patterns of codon bias, can be obscured by biased gene conversion towards GC (bGC), which mimics selective effects (Marais, 2003). Biased gene conversion is a kind of meiotic drive, in which GC alleles are favoured over AT alleles. Increasing evidence suggests that it might affect genome evolution and that it should be taken into account in genomic studies (Marais, 2003; Meunier & Duret, 2004; Galtier & Duret, 2007). It occurs at heterozygote sites involved in the Holliday junction during recombination, so that it is expected to be rare or absent in selfers and in regions of low recombination (Marais, 2003; Marais et al., 2004). Recently, Wright et al. (2007) compared codon usage and base composition between the outcrossers A. lyrata and Brassica oleracea and the selfer A. thaliana. Because most preferred codons ended in G or C, the higher GC content at synonymous sites in outcrossers can be the result of more efficient selection for codon usage or the result of stronger bGC in outcrossers. Because the shift in base composition is independent of gene expression level, Wright et al. (2007) concluded that base composition is more probably caused by bGC (or change in mutation bias) rather than by a reduction in the efficacy of selection in A. thaliana. Until now, studies using species other than A. thaliana have been very scarce, and the effect of mating systems on protein and base composition evolution remains unclear. Testing the effect of mating systems in other groups of species appears timely.

In this study we investigated the effect of mating systems on molecular evolution in four diploid *Triticeae* species: two outcrossers, (i) rye (*Secale cereale*) (Lundqvist, 1954) and (ii) *Aegilops speltoides* (Dvorak et al., 1998); and two self-fertilizing sister species, (iii) Triticum urartu and (iv) Triticum monococcum (Dvorak et al., 1993; Yamane & Kawahara, 2005). Ae. speltoides, T. urartu and T. monococcum are wild diploid relatives of durum wheat (T. turgidum subsp. durum). The phylogenetic relationships and timeline of the evolution of these lineages were estimated by Huang et al. (2002) (Fig. 1). Under the parsimony hypothesis, selfing should have evolved sometime after the last common ancestor of T. urartu and T. monococcum with Ae. speltoides.

In durum wheat, a representative of polyploid Triticeae, a steep recombination gradient along chromosome arms has been found using C-banding polymorphism among chromosomes: about 80% of recombination occurs in the most distal 20% of the chromosomes (Lukaszewski, 1992; Lukaszewski & Curtis, 1993). The distribution of recombination differs also between physically short and long chromosome arms in wheat, with presumably higher recombination rates in short than in long arms (Lukaszewski & Curtis, 1993). This pattern of recombination is found in both wheat and barley (Kunzel et al., 2000). These two species are quite divergent among Triticeae so that this pattern is probably the norm in this tribe. In Aegilops species, including Ae. speltoides, Dvorak et al. (1998) showed that these recombination gradients affect levels of diversity. RFLP polymorphism is higher in telomeric regions than in centromeric regions, the ratio being much higher in some self-fertilizing (Ae. searsi: 24.0) than in outcrossing species (Ae. speltoides: 1.5). These results suggest that recombination should strongly affect genome evolution, even in selfers. Triticeae species thus offer a good opportunity to test for the effect of both mating systems and recombination on patterns of molecular evolution. It is also worth noting that bGC may occur in Triticeae species. Glémin et al. (2006) found highly significant differences in GC content between self-fertilizing and outcrossing Gramineae species, in both coding and non-coding regions. They suggested that bGC could be higher in Gramineae than in other flowering plants, which is in agreement with the high GC content and heterogeneity in GC found in numerous Gramineae genomes (Barakat et al., 1997; Carels & Bernardi, 2000; Wong et al., 2002).

Using the wheat, rye and barley genomic resources available in the publicdomain, we sequenced 52 genes in the four *Triticeae* species and used *Hordeum spontaneum* as an outgroup (or *H. vulgare* when *H. spontaneum* was not available). Most genes were chosen to belong to the same chromosome in order to cover the recombination gradient between centromeric and telomeric regions. To locate genes, we used the genomic resources of the 3B chromosome of wheat, which is the most conserved chromosome

99



Fig. 1. Phylogenetic relationships of the studied species. Barley diverged between 11 and 12 MYA from the diploid *Triticum* and *Aegilops* species, whereas rye diverged more recently, about 7 MYA. *Aegilops speltoides* diverged between 2.5 and 6 MYA from the two *Triticum species*, which separated more recently (around 1 MYA) (Huang *et al.*, 2002). Dashed lines, self-fertilizing lineages leading to *T. monococcum* and *T. urartu*; bold lines, outcrossing lineages leading to *S. cereale* and to *Ae. speltoides*. GC* are shown above each branch. Ratios of non-synonymous to synonymous substitutions (v) shown below each branch were calculated using CODEML (free-ratio model, Mb-8; see text).

Hordeum sp.

between wheat and rice (chromosome 1; Munkvold *et al.*, 2004; Sorrells *et al.*, 2003), and for which the relationship between genetic and physical maps has been intensively investigated (Paux *et al.*, 2006).

We studied the effect of mating systems in regions of low and high recombination on the efficacy of selection through the rate of protein evolution (d_N/d_S) and the pattern of codon bias. We also investigated GC content and GC evolution between mating systems and chromosomal regions. We thus tested whether d_N/d_S ratios are higher, and codon bias and GC content lower, in self-fertilizing than in outcrossing species. Similarly, we tested the same predictions between regions of high and low recombination.

2. Materials and methods

(i) Plant materials

Among *Triticeae*, it is difficult to find several pairs of sister species differing by their mating systems because of phylogenetic and mating system uncertainties. The phylogenetic relationships among the four species we chose are well supported even if gene flow among

the *Triticum*/*Aegilops* group cannot be excluded. This point is discussed below. We also chose species for which the mating system is well known, and to avoid incorrect assignation of ancestral mating systems, we assumed that mating systems are known for the terminal branches only. We used five species: two selfers, T. monococcum (accession DV-92-4-1, collection, NSF Jan Dvorak) and T. urartu (DV-1792-3, collection NSF Jan Dvorak); and two outcrossers, Ae. speltoides (S1 collection INRA Rennes, J. Jahier and S. cereale (PI 561793, USDA). H. spontaneum (PI 282585, USDA) was used as an outgroup. When sequences from H. spontaneum were not available, we used sequences of H. vulgare, the domesticated form of H. spontaneum, extracted from the Expressed Sequence Tag (EST) database of GenBank (National Center for Biotechnology Information, NCBI; http:// www.ncbi.nlm.nih.gov).

(ii) Sampled locus

We sequenced a set of 46 genes expected to cover the homoeologous group 3 chromosome, to allow separation into hypothesized regions differing in recombination rate. PCR primers for gene sequencing

were designed using either a 19400 BAC End Sequences dataset (Paux et al., 2006) generated from a chromosome 3B-specific BAC (Safar et al., 2004), or with barley and wheat ESTs matching rice chromosome 1. We used the location of rice orthologues (on chromosome 1) as a proxy for their chromosomal position. Although synteny between rice chromosome 1 and wheat chromosome 3 is sometimes broken, gene repertoire and order is mostly conserved (Munkvold et al., 2004; Rota & Sorrells, 2004; Sorrells et al., 2003). Because we used two raw recombination categories only, the expected degree of synteny between the two chromosomes seems to be sufficient for our analyses. The relative distance to the centromere was computed for each chromosome arm, assuming that the centromere is located around 17000000 bp from the telomere of the short arm in rice (Sasaki et al., 2002). This defined the short arm versus long arm groups of genes according to their position relative to the centromere. Genes located further than 75% of the arm length from the centromere were grouped into the so-called telomeric group, while other genes were grouped into the centromeric group. We thus assumed that gross recombination patterns remain unchanged between species. Correspondence between rice and wheat chromosome locations (assumed to reflect the position in the four species studied here) was partly through EST physical mapping in wheat nullisomic-tetrasomic, ditelosomic and deletion lines (Sears, 1954; Sears & Sears, 1978; Endo & Gill, 1996), either by direct homology of the studied locus with a mapped EST (Qi et al., 2004; http:// wheat.pw.usda.gov/NSF/progress mapping.html), or by mapping of molecular markers from the same BAC or BAC contig (Paux et al., 2006).

We added five nuclear genes (*Crtiso*, *Eif4e*, *Eifiso4e*, *Gsp-1*, *Psy2*) from different genome locations and one chloroplastic gene (*matK*). No specific location was attributed to these genes in our analysis. Primers, PCR and sequencing conditions are described in Supplementary Table S2. Sequences were aligned manually with the Staden Package (Staden *et al.*, 2001). Coding regions were identified by comparison with EST data and the annotated rice genome (http:// www.gramene.org). In what follows, we analysed only coding regions.

(iii) Protein evolution

We used the maximum likelihood method of the CODEML program in the PAML package (Yang, 1997) to test for variation in the d_N/d_S ratio (hereafter \vee) under different scenarios (see below). For such analyses, we used the best phylogenetic tree found by maximum likelihood reconstruction (model GTR+ C) using the PHYML software (Guindon & Gascuel, 2003) on all loci concatenated (Fig. 1). This tree is

consistent with existing literature (for example see Huang *et al.*, 2002).

To test the effect of mating system on v, we considered several nested models of sequence evolution (Fig. 2). Because we investigated selection patterns on the four focal species, we used Hordeum as an outgroup to distinguish internal and external branches. When this distinction was not necessary, we excluded Hordeum sequences because we wanted to detect selection events on the four focal species only. First, we tested for differences in v between branches according to their mating system. Here, we thus included Hordeum sequences and we ran the following models: (i) the null model (M0), which assumes the same v ratio for all branches; (ii) a second model (Mb-2), which assumes two ratios, one for internal branches plus the outgroup *Hordeum*, V_0 , and one for the external branches leading to the four focal species, v_1 ; (iii) a third model (model Mb-3), which assumes three ratios, one for internal branches plus the outgroup *Hordeum*, v_0 , and two different ratios for the external branches, v out for outcrossing species and v self for self-fertilizing ones; (iv) a free ratio model (Mb-8), in which each branch has its own v ratio (8 v). We used v_0 in Mb-2 and Mb-3 because the mating systems of internal branches are unknown. To test for the effect of mating systems, Mb-3 was compared with Mb-2. These analyses were performed on every gene independently, on all genes concatenated, and on groups of genes concatenated according to their putative chromosome location (long vs short arm; telomeric vs centromeric region). To test for differences in V between genomic regions (long vs short chromosome arm, telomeric vs centromeric region), we used the codon models for multiple gene categories as described by Yang & Swanson (2002) (Fig. 2): we compared a model Mc-0, assuming the same $V(V_0)$ for both categories (and for all branches), with a model Mc-1, assuming one v (for all branches) for each category (v_{high} , v_{low}). In this analysis we excluded the Hordeum outgroup.

Under purifying selection (v < 1), v increases when selection is relaxed, but the reverse is expected under adaptive evolution (v > 1). Few adaptive substitutions can thus lead to patterns with higher v values in outcrossing than in selfing lineages, which is misleading relative to our assumption of accumulating more deleterious mutations in selfers. We thus tested for the possible occurrence of positive selection in the dataset (Fig. 2). First, we ran the nested site models M7 and M8 of CODEML (Yang, 1998). Second, we ran the clade model that allows the $d_{\rm N}/d_{\rm S}$ ratio to vary among sites and among lineages to detect lineage-specific variation in selection pressures and compared it with the null nearly neutral model (M1a) (Bielawski & Yang, 2004). In these tests, we excluded the Hordeum sequences to avoid detecting



One ω (and one transition/transversion ratio) for each region. Codon frequencies and branch lengths can differ.



Fig. 2. Schematic representation of the different branch and branch-site models tested with codeml. Branches with identical thickness lines have the same v value (M0, Mb-2, Mb-3, Mb-8, Mc-0 and Mc-1), or the same distribution of v categories (M1a, *clade*). In models Mc-0 and Mc-1, the two gene categories are allowed to have their own branch lengths as drawn. Model Mb-3 is tested against model Mb-2 (*d.f.*= 1); Model Mc-1 is tested against model Mc-0 (*d.f.*= 2); the clade model is tested against model M1a (*d.f.*= 3).

a signature of positive selection occurring in the *Hordeum* branch.

The maximum likelihoods of the models were computed and their significance was tested by likelihood ratio tests (LRT) with the appropriate degrees of freedom (Yang, 1998; Yang & Nielsen, 1998).

(iv) Base composition and codon usage evolution

For every gene except the chloroplastic gene *matk*, the GC contents of the total, first, second and third codon positions were computed (GC, GC1, GC2 and GC3,

respectively). We also computed the equilibrium GC content a sequence would eventually reach if the substitution pattern occurring in a given lineage remained constant (denoted GC* hereafter). Given the rate of the two classes of substitutions, ATp GC (u) and GCp AT (v), the GC content should converge to GC*= u/(u+v) (Sueoka, 1962). We used the phylogenetic tree of Fig. 1 and estimated the two categories of substitutions on every branch by parsimony using the DNAPARS procedure of the PHYLIP package (Felsenstein, 1989). To achieve sufficient power we pooled the substitutions occurring on external

Same ω (and transition/transversion ratio) for the two regions. Codon frequencies and branch lengths can differ

branches leading to *S. cereale* plus *Ae. speltoides* (bold lines in Fig. 1) and *T. monococcum* plus *T. urartu* (dashed lines in Fig. 1). We computed GC* for all sites, and separately for the first (GC1*), second (GC2*) and third (GC3*) codon positions. Substitutions were pooled for all genes or by genomic region (long vs short arm, telomeric vs centromeric region). Significance of differences in GC* between lineages or between genomic regions were assessed by chi-square tests.

We also examined the pattern of codon usage in the four studied species. The list of preferred codons was determined on both T. aestivum and H. vulgare (Kawabe & Miyashita, 2003; Liu & Xue, 2005; Wang & Roossinck, 2006). As the preferred codons are almost identical in these two species, we assumed that the four species studied here share the same preferred codons as T. aestivum. We estimated the frequency of optimal codons (Fop; Ikemura, 1985) for every gene. Letting U be the substitution rate from preferred to non-preferred codons and P the reverse substitution rate, we estimated the stationary Fop values as $Fop^* = P/(P + U)$. P and U were estimated by parsimony as described for GC and AT substitutions (see above). To distinguish between the effect of selection efficacy on codon usage and the effect of biased gene conversion, we used the fact that in four-fold and sixfold degenerated codons (used for some amino acids), C is preferred over G (Kawabe & Miyashita, 2003; Liu & Xue, 2005; Wang & Roossinck, 2006). For these amino acids, we computed C* and G* values as described above. Substitutions were also pooled for all genes or by genomic regions and chi-square tests were performed.

3. Results

We obtained the complete sequences of 46 gene fragments expected to belong to chromosome 3 (17 designed on BES, 29 based on rice homology), and for 6 other gene fragments located elsewhere in the genome (Supplementary Table S1). Fragments have a minimum length of 700 bp. A total of 31 218 bp of coding regions was available for analyses. Twelve genes (6531 bp) were assigned to the telomeric region versus 34 to the centromeric region (22 437 bp); 35 were assigned to the long arm (19 161 bp) versus 11 to the short arm (8592 bp). The location of seven genes is uncertain. Genes are unevenly located but the average synonymous divergences in each genomic region are very similar (telomeric region, 0·116; centromeric region, 0·109; short arm, 0·117; long arm, 0·105).

(i) Rate of protein evolution

The average v ratio computed on all concatenated genes is $0.160 \pmod{M0}$ and varied between 0.088



Fig. 3. Distribution of the differences $v_{self} - v_{out}$ estimated in model Mb-3 for the 52 genes fragments sequenced. This model has three v ratios: one for the internal branches+ *Hordeum*; one for the external outcrossing branches, v_{out} ; and one for the external selfing branches, v_{self} .

and 0.264 in the free ratio model (Mb-8). None of the 52 single gene fragments, when analysed separately, revealed a significant difference between selfing and outcrossing lineages: v_{self} is higher than v_{out} in no more than about half the genes (Fig. 3). In the concatenated gene analyses, contrary to expectations, v_{self} (from 0.077 to 0.115) is lower than v_{out} (from 0.114 to 0.159), but the difference is significant only for centromeric genes (*P*= 0.048; Table 1).

On the whole concatenated sequence dataset, we detected a proportion of 2.5% of sites under possible positive selection with v = 2.88 on average (P= 0.031) (M8 vs M7; Table 2). Among these sites, none has a posterior probability of being under positive selection higher than 0.95 (Yang et al., 2005). Excluding these sites from the dataset did not change previous conclusions (data not shown). Under the clade model we detected a significant difference between outcrossing and self-fertilizing lineages (P= 0.021): a proportion $P_2 = 3\%$ of sites are under positive selection in outcrossers (v_{2out} = 3.02), while these sites evolve almost neutrally in selfers (v_{2self} = 1.01) (Table 2). In summary, we found no clear evidence that the two self-fertilizing species have fixed more slightly deleterious alleles than outcrossers, but positive selection seems to be more efficient in outcrossers than in selfers. This might obscure the putative difference in fixation rates of slightly deleterious alleles.

Contrasts in substitution patterns in short versus long arms of chromosome 3 and telomeric versus centromeric regions were used to estimate the impact of variation in local recombination rates on rates of protein evolution. We found no difference between short arm and long arm regions, but the v ratio is higher in centromeric than in telomeric regions both in selfing (0·104 and 0·077, respectively; Table 1) and in outcrossing lineages (0·169 and 0·125, respectively; Table 1). However, in both cases, model Mc-1 is not significantly better than model Mc-0 (Table 1).

Table 1. Summary of maximum likelihood estimates of the ratio of non-synonymous to synonymous substitutions (v) in different models contrasting selfing and outcrossing lineages (model Mb-3 against model Mb-2, 1 d.f.) and highly and weakly recombining regions (model Mc-1 against model Mc-2, 2 d.f.)

Sequence set No. of genes Size (bp)		All genes 52 31 218	Chr. 3 telomeric 12 6531	Chr. 3 centromeric 34 22 425	Chr. 3 short arm 11 8592	Chr. 3 long arm 35 19 161
Model M0	v	0·160	0·136	0·161	0·159	0·158
	lnL	x 59042·70	x 11043·52	x 38485·80	x 14757·85	x 32805·47
Model Mb-2	V0	0·175	0·159	0·169	0·171	0·167
	V1	0·140	0·110	0·150	0·144	0·146
	lnL	x 59040·22	x 11042·40	x 38485·33	x 14757·47	x 32804·95
Model Mb-3	V ₀	0·174	0·158	0·168	0·170	0·167
	V _{self}	0·105	0·074	0·101	0·099	0·113
	V _{out}	0·150	0·120	0·164	0·155	0·156
	lnL	× 59038·56	× 11041·90	x 38483·36	× 14756·85	x 32804·14
LRT		3·33	1·01	3·94	1·22	1·63
<i>P</i> value		0·068	0·316	0·047	0·269	0·202
Model Mc-0	۷ ₀ lnL		0·155 x 46002·94		0·159 x 44164·85	
Model Mc-1	$v_{ m high}/v_{ m low}$ lnL		0·136/0·161 x 46002·42		0·150/0·164 x 44164·59	
LRT P value			1·03 0·309		0·53 0·468	

Table 2. Summary of maximum likelihood of site models and branch-site models. Model M8 is tested against model M7 (1 d.f.). The clade model is tested against model M1a (3 d.f.). In models M7 and M8 the \vee ratio of sites under purifying selections follows a beta distribution with parameters p and q (column 2, Beta: p/q)

Beta: <i>p/q</i> lnL	0·013/0·070 x 54007·67
Beta : p/q $p_1/v_1 > 1$	9·72/99·00 2·5 %/2·88
lnL	x 54005·34
	4.66
	0.031
$p_0 / V_0 < 1$	84%/0.00
$p_1/v_1 = 1$	16%
lnL	x 54006.58
$n_{\rm o}/{\rm V} \ge 1$	97%/0.085
$p_{1}/v_{1} = 1$	0
p_{2}/V_{2} out/	$3 \frac{9}{10} \frac{1}{2000} = 3.01/$
V welf	$v_{\text{asclf}} = 1.01$
lnL	x 54001.73
	9.69
	0.021
p_0/V_0	86%/0.003
$p_1/v_1 = 1$	11.80%
$p_2/V_{2out}/$	$2.1 \% /v_{2001} = 2.73/$
V _{2self}	$v_{\text{2self}} = 0.00$
lnL	x 54002·10
	8.95
	0.030
	Beta : p/q lnL Beta : p/q $p_1/v_1 > 1$ lnL $p_0/v_0 < 1$ $p_1/v_1 = 1$ lnL $p_0/v_0 < 1$ $p_1/v_1 = 1$ $p_2/v_{2out}/v_{2sclf}$ lnL $p_0/v_0 = 1$ $p_2/v_{2out}/v_{2sclf}$ lnL $p_2/v_{2out}/v_{1} = 1$ $p_2/v_{2out}/v_{1} = 1$ $p_2/v_{2out}/v_{1} = 1$ $p_2/v_{2out}/v_{1} = 1$ $p_2/v_{2out}/v_{1} = 1$ $p_2/v_{2out}/v_{1} = 1$ $p_2/v_{2out}/v_{2sclf}$ lnL

(ii) Base composition and codon usage

Current GC and GC3 values are very similar in all species, being only slightly higher in the two outcrossers than in the two selfers (Table 3). GC* values for each branch estimated on all concatenated genes are presented in Fig. 1. For each codon position (GC1*, GC2* and GC3*) and for each genomic region, outcrossing lineages have higher GC* values than selfing lineages, but in numerous cases the number of variable sites is not large enough to result in a significant difference (Table 2). Highly recombining regions (telomeric region, short arm) also show higher GC* and GC3* values than weakly recombining ones (centromeric region, long arm), in both selfing and outcrossing lineages (Table 2). When pooling substitutions from all branches, GC3* is significantly higher (P=0.02) on the short arm of chromosome 3 (0.53) than on the long arm (0.43). The same trend is observed between telomeric (GC3*= 0.53) and centromeric regions (GC3*= 0.45), but the difference is not significant (P=0.097). Overall, these results suggest that GC enrichment is positively correlated with the efficacy of recombination, both among (outcrossing vs selfing) and within genomes (recombination gradient).

Because preferred codons in *Triticeae* species typically end in G or C (Kawabe & Miyashita, 2003; Liu & Xue, 2005; Wang & Roossinck, 2006), results on GC3* can be due to higher codon bias in highly recombining genomes and regions. Accordingly, *Fop* is

Table 3. GC content and GC evolution (GC*) in the two self-fertilizing (T. monococcum and T. urartu) and the two outcrossing species (S. cereale and Ae. speltoides)

Sequence	MS lineages	GC1	GC2	GC3	GC	GC1*	GC2*	GC3*		GC*	
All nuclear genes (29 673 bp)	Outcrossing Self-fertilizing	0.535	0.403	0.480	0.473	0·454 (81) 0·362	0.375 (62) 0.249	0·480 (352) 0·366	p = 0.06	0·453 (495) 0·353	<i>p</i> = 0.03
	Sen-ter thizing	g 0 <i>555</i> 0402 04	0475	0.475 0.471	(18)	(15)	(109)	1	(142)	•	
Chr. 3 centromeric	Outcrossing	0.548	0.384	0.487	0.473	0.426	0.384	0.477		0.443	
(22437 bp)	Self-fertilizing	0.548	0.383	0.487	0.473	(66) 0.310 (11)	(50) 0.210 (10)	(253) 0.358 (78)	<i>p</i> = 0.08	(369) 0.326 (99)	<i>p</i> = 0.04
Chr. 3 telomeric	Outcrossing	0.542	0.413	0.488	0.481	0.611	0.375	0.528		0.511	
(6531 bp)	Self-fertilizing	0.543	0.413	0.482	0.479	(14) 0.283 (4)	(13) 0.588 (3)	(82) 0.463 (25)	<i>p</i> = 0·62	(109) 0.449 (32)	<i>p</i> = 0·53
Chr. 3 long arm	Outcrossing	0.542	0.408	0.494	0.481	0.461	0.418	0.464		0.452	
(19161 bp)	Self-fertilizing	0.541	0.407	0.491	0.480	(52) 0.336 (10)	(35) 0.369 (13)	(213) 0.352 (72)	<i>p</i> = 0·10	(300) 0.352 (95)	<i>p</i> = 0.08
Chr. 3 short arm	Outcrossing	0.554	0.395	0.478	0.476	0.473	0.376	0.529		0.496	
(632 UP)	Self-fertilizing	0.553	0.395	0.475	0.474	(24) 0·202 (6)	(23) 0.000 (2)	(112) 0.455 (29)	<i>p</i> = 0.50	(101) 0.385 (37)	<i>p</i> = 0·20

Values within parentheses indicate the number of substitutions used to compute GC*. Chi-square tests were performed using contingency tables of the number of ATp GC and GCp AT substitutions in the two categories tested.

Table 4. *Stationary optimal codon frequency (Fop*) in the two self-fertilizing* (T. monococcum *and* T. urartu) *and the two outcrossing species* (S. cereale *and* Ae. speltoides)

	Fop*			
Sequence set	All branches	Outcrossing	Self-fertilizing	
All genes		0.339	0.235	p = 0.057
e		(242)	(77)	1
Chr. 3 centromeric	0.300	0.336	0.202	p = 0.041
	(501)	(163)	(50)	•
Chr. 3 telomeric	0.384	0.375	0.377	p = 0.988
	(149)	(55)	(17)	1
	p = 0.044	p = 0.585	p = 0.123	
Chr. 3 long arm	0.301	0.324	0.218	p = 0.130
C	(439)	(138)	(44)	1
Chr. 3 short arm	0·359	0·385	0.313	p = 0.496
	(210)	(80)	(24)	1
	p = 0.118	p = 0.344	p = 0.341	

Values within parentheses indicate the number of substitutions used to compute *Fop**. Chi-square tests were performed using contingency tables of the number of preferredp unpreferred and unpreferredp preferred substitutions in the two categories tested.

slightly higher in outcrossing species (35.4%) than in self-fertilizing ones (35.1%). As for GC content, these values are very similar, but *Fop** values are much more contrasted (Table 4). As theoretically predicted, *Fop** is higher in outcrossing than in selfing lineages but the difference is only significant for the centromeric genes (*P*= 0.04), and marginally significant when all genes are concatenated (P= 0.057). Fop* is also higher in highly than in weakly recombining regions. Pooling substitutions from all the branches, the difference is significant between centromeric and telomeric regions (P= 0.04), but not significant between the long and short arm (P= 0.118) (Table 4).



Fig. 4. Theoretical V ratio for slightly deleterious alleles, Fop^* and GC*, computed in an outcrossing species and in a highly self-fertilizing species (selfing rate: S=95%) with different effective population sizes. Effective population sizes are $N_e = N_{out}$ and $N_e = N_{self}(1 \times S/2)$ for outcrossing and self-fertilizing species, respectively. N_{self} can be lower than N_{out} if factors other than the automatic effect of selfing $(1 \times S/2)$ reduce N_e (bottlenecks, hitch-hiking effects). For co-dominant mutations and $N_{self} = N_{out}$, selection is independent of S (see Charlesworth, 1992), and $V = \frac{4N_S}{e^{4N_X}1}$ (Charlesworth, 1992), and $Fop^* = \frac{e^{4N_X}N}{e^{4N_X}1}$ (Bulmer, 1991); $GC^* = \frac{e^{4N_X}N_S(2 \times S)}{u/y^4}$ (Marais *et al.*, 2004). N is N_{out} in outcrossers and N_{self} in selfers; s is the selection coefficient against deleterious alleles (V) or in favour of preferred codons (Fop^*); C is the intensity of bGC; u is the mutation rate towards unpreferred codons (Fop^*) or towards AT alleles (GC*); and v is the reverse mutation rate. For the two graphs, S= 0.95, s= c= 0.0002, and u/v = 1.5. (A) $N_{out} = 1000$ and $N_{self} = 400$, that is $N_e(\text{self})B$ 200. (B) $N_{out} = 1000$ and $N_{self} = 900$, that is $N_e(\text{self})B$ 450.

Given the strong correlation between Fop and GC, these results could be due to the sole effect of biased gene conversion (or biased mutation) rather than selection on codon usage. If selection on codon usage occurs, we would predict that C*, but not G*, should vary with mating system and recombination in four-fold and six-fold degenerated codons because C is the preferred base at the third position in those codons. Using all concatenated genes, G* is almost constant (P=0.87) in outcrossing ($G^{*}=0.29$) and in selfing lineages (G*= 0.28). C* is more variable (0.20 and 0.13, respectively), but the difference is not significant (P=0.20). G* does not differ significantly either between short and long arm ($G^{*=} 0.24$; $G^{*=}$ 0.25; P= 0.87) or between telomeric and centromeric regions (G*= 0.32; G*= 0.23; P= 0.46). C* differs significantly between short and long arm ($C^*= 0.25$; C*= 0.16; P= 0.02), but not between telomeric and centromeric regions (C*= 0.23; C*= 0.17; P= 0.15). These results weakly support selection being less efficient in weakly recombining than in highly recombining genomes or regions.

(iii) Protein and GC evolution

GC1* and GC2* shared a pattern similar to GC3*, suggesting that bGC should be effective and could affect protein evolution, partially masking selective effects. To test this hypothesis, we computed GC* on the whole tree for every gene. We then separated the genes into two groups: genes having lower GC* than the median value (45%) and genes having higher GC* than the median. The two groups had similar size (16725 and 13776 bp, respectively). First, we re-ran models Mc-0 and Mc-1 to compare these two gene categories, substituting high and low GC* categories for high and low recombination categories. v is significantly (P=0.018) lower in the low GC* group (v=0.127) than in the high GC* group (v=0.172). Second, we re-ran the *branch* model (model 2) analysis on the two GC* groups. In the low GC* group, there is no significant difference (P=0.664) between $v_{self}=0.111$ and $v_{out}=0.126$. On the contrary, in the high GC* group, $v_{self}=0.099$ and $v_{out}=0.180$ are significantly different (P=0.032). These results suggest that GC evolution affects protein evolution, at least in outcrossers.

4. Discussion

We studied the impact of selfing and recombination on molecular evolution patterns in four Triticeae species. We aimed to test whether selfing and low recombination reduce the efficacy of selection and the strength of bGC. Selection effects scale in $N_e s$, where s is the selection coefficient, while bGC effects scale in $N_e \mathbf{c}(1-S)$, where **c** is the intensity of bGC and S the selfing rate (Marais et al., 2004). Selfing on the whole-genome scale and recombination at a more local scale affect selection efficacy through N_e variation only, but they affect bGC through both N_e and the effective intensity of bGC, c(1 - S). BGC depends on the efficacy of recombination (Marais, 2003) and is likely to vary dramatically from c to 0 depending on the mating system of the species. Fig. 4 illustrates how v, Fop* and GC* are affected by mating systems when reduction in N_e due to selfing is large (Fig. 3A: $N_e(\text{self})/N_e(\text{out}) = 0.2$), or small (Fig. 3B: $N_e(\text{self})/N_e(\text{out})=0.45$, that is just below the $\frac{1}{2}$ threshold). We thus expect a stronger effect of mating system and recombination on bGC than on selection efficacy.

(i) Mating systems and recombination affect base composition

Results on GC content and codon usage fit well with the theoretical predictions on mating system and recombination effects. GC* and Fop* are higher in outcrossing than in self-fertilizing species, and in highly than in weakly recombining regions. However, differences were not all significant because of the lack of statistical power for some comparisons (Tables 3 and 4). Patterns of GC3* and Fop* can be explained either by selection on codon usage or by bGC (or both). Because GC1* and GC2* patterns fit theoretical predictions as well as GC3*, bGC probably drives, at least partly, the evolution of GC content in outcrossing species and in genomic regions with high recombination rates. We also found some evidence for selection on codon usage by contrasting G* and C* in four-fold and six-fold codons. However, this effect is weak: we found no significant mating system effect on preferred codon selection, and a significant but weak effect of recombination only when contrasting centromeric versus telomeric regions. Finally, we cannot rule out mutational bias to explain our results. However, such a bias should be mating system dependent, which has not been documented. Overall, our results are better explained by bGC than by other factors, but polymorphism data in non-coding regions would be necessary to disentangle the different hypotheses.

The bGC interpretation is consistent with the highly significant differences in GC content found between self-fertilizing and outcrossing species detected among Gramineae in both coding and noncoding regions (Glémin et al., 2006) and among Triticeae species in EST (Akhunov et al., 2003a). It is surprising, however, to have found a recombination effect even in self-fertilizing species (Table 3). Marais et al. (2004) showed that GC* should almost not vary with recombination in highly self-fertilizing species, which could explain the lack of variation in GC content observed along the A. thaliana genome. We hypothesize that the *Triticum* species studied here exhibit a lower selfing rate than A. thaliana, and that very strong recombination gradients, such as those observed in several Triticeae species (Lukaszewski, 1992; Lukaszewski & Curtis, 1993), leave a genomic signature even in self-fertilizing species. For example, RFLP polymorphism is 12 to 25 times higher in telomeric than in centromeric regions in some selffertilizing Aegilops species (Dvorak et al., 1998).

(ii) Mating systems and recombination do not clearly affect protein evolution

Contrary to expectations, we found no evidence that the two self-fertilizing species T. monococcum

and T. urartu have fixed more slightly deleterious mutations than the two outcrossing species S. cereale and Ae. speltoides. Neither gene-by-gene analysis (Fig. 3) nor analyses of concatenated genes (Table 1) showed a significantly higher v_{self} than v_{out} , and the reverse tendency is even observed. Several reasons can be invoked to explain the lack of evidence of reduced selection efficacy in the self-fertilizing species studied. First, the four Triticeae species studied here diverged recently. Polymorphism for deleterious alleles, especially in outcrossing species, can therefore account for a part of the divergence estimates, since only one sequence per gene was produced for each species. In such cases we would expect higher v ratios in terminal than in internal branches because p_N/p_S ratios are expected to be higher than d_N/d_S ratios under purifying selection. This could especially be the case for rye because of recent increased drift due to domestication. However, we found the reverse pattern: rye has a lower v than Ae. speltoides (see Fig. 1). Intraspecific polymorphism alone cannot explain our results. The recent origin of selfing was also invoked to explain similar results in the comparison of A. thaliana and A. lyrata (Wright et al., 2002). We cannot rule out this hypothesis for the two Triticum species. As selfing can rapidly and recurrently evolve, transition from outcrossing to selfing may have occurred independently and recently in the two species, although this scenario is not the most parsimonious. Similarly, the differences in branch lengths between selfing and outcrossing lineages can bias the results. However, if selfing was that recent (and correspondingly the selfing branch lengths too short), it should not have affected base composition dynamics. Gene flow among these species complex may also have obscured the expected pattern, but once again GC content dynamics should also have been obscured as well. Together, results on protein and base composition evolution suggest that Fig. 4Bmatches our dataset better than Fig. 4A. Mating systems apparently weakly affect the efficacy of selection whereas GC dynamics is more strongly affected, at least partly through bGC effects.

Finally, we suggest that other processes can obscure the predictions. First, bGC could affect protein evolution (see below). Second, we also found an excess of non-synonymous substitutions in outcrossing lineages (p_2 = 3% with v_{2out} = 3.02 in outcrossers vs v_{2self} = 1.01 in selfers), suggesting that positive selection is more efficient for these species than in the two self-fertilizing species. Short branches of selffertilizing lineages could explain why we did not detect positive selection in these lineages. However, if we consider all internal branches as selfers (Fig. 2) we still detect positive selection in the outcrossing lineages only (p_2 = 2.1% with v_{2out} = 2.73 vs v_{2self} = 0 in selfers, *p* value= 0.030). This suggests that

107

outcrossing and self-fertilizing lineages actually have different selection patterns. Few adaptive substitution events would explain why v_{out} is slightly higher than v_{self} . Given these values, we can roughly estimate a mean corrected v ratio, $v k_{ot}$, by excluding sites under putative positive selection. We can use $v_{tot} = p_2 v_2 +$ $(1 - p_2) v k_{ot}$ with v_{tot} standing for v_{self} and v_{out} of model Mb-2 (Table 1). It gives, $v k_{out} = 0.060$ instead of $v_{out} = 0.150$, and $v k_{elf} = 0.078$ instead of $v_{self} = 0.105$, which matches theoretical predictions slightly better.

We also tried to detect higher rates of protein evolution in low recombining chromosomic regions than in higher recombining ones by contrasting both long versus short arm and proximal versus distal regions. The v ratio is very similar between short and long arms, while the difference is higher, although non-significant, between telomeric and centromeric regions. Few studies have demonstrated a positive relation between recombination and the efficacy of selection through d_N/d_S measures (see for instance Haddrill et al., 2007; Pal et al., 2001). In Drosophila, genomic regions with no crossing-over experience a severe reduction in the efficacy of selection, but even a low frequency of crossing over in other regions seems to be enough to maintain the efficacy of selection (Haddrill et al., 2007). In Triticeae species, recombination gradients are very steep with virtually no recombination occurring around the centromeres (Lukaszewski, 1992; Lukaszewski & Curtis, 1993). This would explain why selection seems to be somewhat weaker in centromeric than in telomeric regions, but not between short and long chromosome arms, between which differences in average recombination rates are lower. These results must be viewed with caution because of possible assignation errors. Synteny between rice chromosome 1 and wheat chromosome 3 is sometimes broken and physical assignation of our locus using wheat deletion lines revealed a few discrepancies between rice and wheat locations, especially in the telomeric region of the short arm (Supplementary Table S1). This synteny erosion in the genome of wheat ancestors appears to be linked to recombination intensity and mating system (Akhunov et al., 2003b). Another limitation of this study is the under-representation of telomeric genes. Extensive gene sequencing along the same chromosome would allow a more robust analysis of the effect of recombination. Despite these limitations, it is worth noting that, contrary to protein evolution, GC content differences between genomic regions match well the theoretical predictions, suggesting that gene location assignation should be mainly correct.

(iii) *Does bGC affect protein evolution?*

We found that mating system and recombination affect GC content evolution even on first and second

ingly, v ratios are higher in regions of high than low GC*, especially in outcrossing lineages. Protein evolution could affect GC content but it is difficult to explain why changes in amino acid would preferentially result in ATp GC substitutions. We thus favour the reverse explanation that GC evolution drives protein evolution. Fast-evolving genomic regions exhibiting mostly ATP GC changes have also been observed in highly recombining regions in humans (Pollard et al., 2006) and in mice (Galtier & Duret, 2007). Positive selection is mostly invoked to explain such a lineage-specific acceleration of substitution rates in comparison with close relatives. Recently, Galtier & Duret (2007) proposed a neutral alternative explanation. The bGC molecular drive could overcome purifying selection, and lead to the fixation of G and C weakly deleterious mutations. BGC can thus increase the mutation load, as previously suggested by Bengtsson (1990). If high GC* is a consequence of bGC hotspots rather than selection on codon usage, our results are consistent with a bGC-associated mutation load in the two outcrossing species S. cereale and Ae. speltoides. According to Galtier & Duret (2007), outcrossing species, but not self-fertilizing ones, would suffer from a 'genomic Achille's heel'. BGC could also lead to spurious detection of sites under positive selection in outcrossing lineages. This unexpected effect contributes to obscuring the effect of mating system and recombination on patterns of selection. Including non-coding regions in such molecular studies would help to characterize parameters, such as the strength of bGC, to be included in the null model to test more accurately the effect of selection.

codon positions, and thus protein evolution. Accord-

(iv) Genomic approaches on mating system evolution

Such genomic approaches shed light on the evolution of mating systems. Self-fertilization has been suggested to be an evolutionary dead end because of the accumulation of slightly deleterious mutations, and because low genetic diversity may preclude adaptation (see review in Takebayashi & Morrell, 2001). Very few studies have provided support for the premise of this hypothesis, namely reduced selection efficacy in selfers. Our results gave no clear support for an increased drift load in selfers, partly because other processes may interfere. We suggest that, surprisingly, outcrossers may also suffer from an increase mutation load due to bGC, not to drift. We also found that positive selection could be more efficient in outcrossing than in self-fertilizing species, in agreement with the dead end theory. Wright et al. (2002) did not find any evidence for an increased load in the selfer A. thaliana compared with the outcrosser A. lyrata, but they did not test for difference in adaptive evolution nor for interference with bGC. Similar studies in other species are needed to evaluate the generality of these results. Larger genomic data and better knowledge of the distribution of local recombination rates would also help in estimating the genetic load associated with each mating system. The dead end theory also posits that self-fertilizing species should be of recent origin, and that transitions from selfing to outcrossing are rare (Takebayashi & Morrell, 2001). Inferring shifts in mating systems and their timing would thus be useful. If GC* values are well correlated with effective recombination rates (Meunier & Duret, 2004) and thus to mating systems, we hypothesize that GC3* or GC* computed on noncoding sequences could be useful statistics to infer mating system evolution along phylogenies.

We thank A. Tsitrone and J. Ronfort for helpful discussion, and N. Galtier, B. Mable, and two anonymous reviewers for comments on the manuscript. S.G. also thanks D. Charlesworth for her help on his first study on molecular evolution. This work was supported by the French Agence Nationale de la Recherche (ANR 'Exegese-Blé' project) and by the French Institut National de la Recherche Agronomique (INRA 'Tritipol' project). This is publication ISE-M 2007-152 of the Institut des Sciences de l'Evolution de Montpellier.

References

- Akashi, H. (1995). Inferring weak selection from patterns of polymorphism and divergence at 'silent' sites in *Drosophila* DNA. *Genetics* **139**, 1067–1076.
- Akhunov, E. D., David, J. L., Chao, S., Lazo, G., Anderson, O. D., et al. (2003a). GC composition and codon usage in genes of inbreeding and outcrossing *Triticeae* species. In *Tenth International Wheat Genetics Symposium*, pp. 203–206. Italy: Paestum.
- Akhunov, E. D., Goodyear, A. W., Geng, S., Qi, L. L., Echalier, B., Gill, B. S., Miftahudin, T., Gustafson, J. P., Lazo, G., Chao, S., *et al.* (2003*b*). The organization and rate of evolution of wheat genomes are correlated with recombination rates along chromosome arms. *Genome Research* 13, 753–763.
- Baker, H. G. (1955). Self-compatibility and establishment after 'long-distance' dispersal. *Evolution* **9**, 347–348.
- Barakat, A., Carels, N. & Bernardi, G. (1997). The distribution of genes in the genome of gramineae. *Proceedings of the National Academy of Sciences of the USA* 94, 6857–6861.
- Bengtsson, B. O. (1990). The effect of biased conversion on the mutation load. *Genetical Research* 55, 183–187.
- Bielawski, J. P. & Yang, Z. (2004). A maximum likelihood method for detecting functional divergence at individual codon sites, with application to gene family evolution. *Journal of Molecular Evolution* **59**, 121–132.
- Bulmer, M. (1991). The selection-mutation-drift theory of synonymous codon usage. *Genetics* **129**, 897–907.
- Carels, N. & Bernardi, G. (2000). Two classes of genes in plants. *Genetics* **154**, 1819–1825.
- Charlesworth, B. (1992). Evolutionary rates in partially selffertilizing species. *The American Naturalist* **140**, 126–148.
- Charlesworth, B., Morgan, M. T. & Charlesworth, D. (1993). The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**, 1289–1303.

- Charlesworth, D. & Wright, S. I. (2001). Breeding systems and genome evolution. *Current Opinion in Genetics & Development* 11, 685–690.
- Dvorak, J., Diterlizzi, P., Zhang, H. B. & Resta, P. (1993). The evolution of polyploid wheats: identification of the A-genome donor species. *Genome* **36**, 21–31.
- Dvorak, J., Luo, M. C. & Yang, Z. L. (1998). Restriction fragment length polymorphism and divergence in the genomic regions of high and low recombination in selffertilizing and cross-fertilizing *Aegilops* species. *Genetics* 148, 423–434.
- Endo, T. R. & Gill, B. S. (1996). The deletion stocks of common wheat. *Journal of Heredity* **87**, 295–307.
- Felsenstein, J. (1989). PHYLIP: phylogeny inference package (version 3.2). *Cladistics* 5, 164–166.
- Galtier, N. & Duret, L. (2007). Adaptation or biased gene conversion? Extending the null hypothesis of molecular evolution. *Trends in Genetics* 23, 273–277.
- Glémin, S., Bazin, E. & Charlesworth, D. (2006). Impact of mating systems on patterns of sequence polymorphism in flowering plants. *Proceedings of the Royal Society of London, Series B* 273, 3011–3019.
- Guindon, S. E. P. & Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology* 52, 696–704.
- Haddrill, P. R., Halligan, D. L., Tomaras, D. & Charlesworth, B. (2007). Reduced efficacy of selection in regions of the *Drosophila* genome that lack crossing over. *Genome Biology* 8, R18.
- Hamrick, J. L. & Godt, M. J. W. (1996). Effects of life history traits on genetic diversity in plants species. *Philosophical Transactions of the Royal Society of London, Series B* 351, 1291–1298.
- Huang, S., Sirikhachornkit, A., Su, X., Faris, J., Gill, B., Haselkorn, R. & Gornicki, P. (2002). Genes encoding plastid acetyl-CoA carboxylase and 3-phosphoglycerate kinase of the *Triticum/Aegilops* complex and the evolutionary history of polyploid wheat. *Proceedings of the National Academy of Sciences of the USA* **99**, 8133–8138.
- Ikemura, T. (1985). Codon usage and tRNA content in unicellular and multicellular organisms. *Molecular Bio*logy and Evolution 2, 13–35.
- Ingvarsson, P. K. (2002). A metapopulation perspective on genetic diversity and differentiation in partially selffertilizing plants. *Evolution* 56, 2368–2373.
- Kawabe, A. & Miyashita, N. T. (2003). Patterns of codon usage bias in three dicot and four monocot plant species. *Genes and Genetic Systems* **78**, 343–352.
- Kunzel, G., Korzun, L. & Meister, A. (2000). Cytologically integrated physical restriction fragment length polymorphism maps for the barley genome based on translocation breakpoints. *Genetics* **154**, 397–412.
- Liu, Q. & Xue, Q. (2005). Comparative studies on codon usage pattern of chloroplasts and their host nuclear genes in four plant species. *Journal of Genetics* **84**, 55–62.
- Lukaszewski, A. J. (1992). A comparision of physical distribution of recombination in chromosome 1R in diploid rye and in hexaploid triticale. *Theoretical and Applied Genetics* 83, 1048–1053.
- Lukaszewski, A. J. & Curtis, C. A. (1993). Physical distribution of recombination in B-genome chromosomes of tetraploid wheat. *Theoretical and Applied Genetics* 86, 121–127.
- Lundqvist, A. (1954). Studies on self-sterility in rye, Secale cereale L. Hereditas 40, 278–294.
- Marais, G. (2003). Biased gene conversion: implications for genome and sex evolution. *Trends in Genetics* **19**, 330–338.

- Marais, G., Charlesworth, B. & Wright, S. I. (2004). Recombination and base composition: the case of the highly self-fertilizing plant *Arabidopsis thaliana*. *Genome Biology* 5, R45.
- Maynard-Smith, J. & Haigh, D. (1974). The hitch-hiking effect of a favourable gene. *Genetical Research* 23, 23–35.
- Meunier, J. & Duret, L. (2004). Recombination drives the evolution of GC-content in the human genome. *Molecular Biology and Evolution* **21**, 984–990.
- Munkvold, J. D., Greene, R. A., Bertmudez-Kandianis, C. E., La Rota, C. M., Edwards, H., Sorrells, S. F., Dake, T., Benscher, D., Kantety, R., Linkiewicz, A. M., et al. (2004). Group 3 chromosome bin maps of wheat and their relationship to rice chromosome 1. Genetics 168, 639–650.
- Nordborg, M. (2000). Linkage disequilibrium, gene trees and selfing: an ancestral recombination graph with partial self-fertilization. *Genetics* **154**, 923–929.
- Nybom, H. (2004). Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. *Molecular Ecology* **13**, 1143–1155.
- Pal, C., Papp, B. & Hurst, L. D. (2001). Does the recombination rate affect the efficiency of purifying selection? The yeast genome provides a partial answer. *Molecular Biology and Evolution* 18, 2323–2326.
- Paux, E., Roger, D., Badaeva, E., Gay, G., Bernard, M., Sourdille, P. & Feuillet, C. (2006). Characterizing the composition and evolution of homoeologous genomes in hexaploid wheat through BAC-end sequencing on chromosome 3B. *The Plant Journal* 48, 463–474.
- Pollak, E. (1987). On the theory of partially inbreeding finite populations. I. Partial selfing. *Genetics* 117, 353–360.
- Pollard, K. S., Salama, S. R., King, B., Kern, A. D., Dreszer, T., Katzman, S., Siepel, A., Pedersen, J. S., Bejerano, G., Baertsch, R., *et al.* (2006). Forces shaping the fastest evolving regions in the human genome. *PLoS Genetics* 2, e168.
- Qi, L. L., Echalier, B., Chao, S., Lazo, G. R., Butler, G. E., Anderson, O. D., Akhunov, E. D., Dvorak, J., Linkiewicz, A. M., Ratnasiri, A., *et al.* (2004). A chromosome bin map of 16000 expressed sequence tag loci and distribution of genes among the three genomes of polyploid wheat. *Genetics* 168, 701–712.
- Rota, M. & Sorrells, M. E. (2004). Comparative DNA sequence analysis of mapped wheat ESTs reveals the complexity of genome relationships between rice and wheat. *Functional and Integrative Genomics* **4**, 34–46.
- Safar, J., Bartos, J., Janda, J., Bellec, A., Kubalakova, M., Valarik, M., Pateyron, S., Weiserova, J., Tuskova, R., Cihalikova, J., *et al.* (2004). Dissecting large and complex genomes: flow sorting and BAC cloning of individual chromosomes from bread wheat. *The Plant Journal* 39, 960–968.
- Sasaki, T., Matsumoto, T., Yamamoto, K., Sakata, K., Baba, T., Katayose, Y., Wu, J., Niimura, Y., Cheng, Z., Nagamura, Y., *et al.* (2002). The genome sequence and structure of rice chromosome 1. *Nature* **420**, 312–316.
- Schoen, D. J. & Brown, A. H. D. (1991). Intraspecific variation in population gene diversity and effective population size correlates with the mating system in plants. *Proceedings of the National Academy of Sciences of the* USA 88, 4494–4497.

- Sears, E. R. (1954). The aneuploid of common wheat. Missouri Agricultural Experiment Station Research Bulletin 572, 1–58.
- Sears, E. R. & Sears, L. (1978). The telocentric chromosomes of common wheat. In *Proceedings of the Fifth International Wheat Genetics Symposium* (ed. S. Ramanujams), pp. 389–407. New Delhi, India: Indian Agricultural Research Institute.
- Sorrells, M. E., La Rota, M., Bermudez-Kandianis, C. E., Greene, R. A., Kantety, R., Munkvold, J. D., Miftahudin, T., Mahmoud, A., Ma, X. F., Gustafson, P. J., *et al.* (2003). Comparative DNA sequence analysis of wheat and rice genomes. *Genome Research* 13, 1818– 1827.
- Staden, R., Judge, D. P. & Bonfield, J. K. (2001). Sequence assembly and finishing methods. *Methods of Biochemical Analysis* 43, 303–322.
- Sueoka, N. (1962). On the genetic basis of variation and heterogeneity of DNA base composition. *Proceedings* of the National Academy of Sciences of the USA **48**, 582–592.
- Takebayashi, N. & Morrell, P. L. (2001). Is self-fertilization an evolutionary dead end? Revisiting an old hypothesis with genetic theories and a macroevolutionary approach. *American Journal of Botany* **88**, 1143–1150.
- Wang, L. J. & Roossinck, M. J. (2006). Comparative analysis of expressed sequences reveals a conserved pattern of optimal codon usage in plants. *Plant Molecular Biology* 61, 699–710.
- Wong, G. K., Wang, J., Tao, L., Tan, J., Zhang, J., Passey, D. A. & Yu, J. (2002). Compositional gradients in *Gramineae* genes. *Genome Research* 12, 851–856.
- Wright, S. I., Lauga, B. & Charlesworth, D. (2002). Rates and patterns of molecular evolution in inbred and outbred *Arabidopsis*. *Molecular Biology and Evolution* 19, 1407–1420.
- Wright, S. I., Iorgovan, G., Misra, S. & Mokhtari, M. (2007). Neutral evolution of synonymous base composition in the *Brassicaceae*. *Journal of Molecular Evolution* 64, 136–141.
- Yamane, K. & Kawahara, T. (2005). Intra- and interspecific phylogenetic relationships among diploid *Triticum-Aegilops* species (*Poaceae*) based on base-pair substitutions, indels, and microsatellites in chloroplast noncoding sequences. *American Journal of Botany* 92, 1887–1898.
- Yang, Z. (1998). Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Molecular Biology and Evolution* 15, 568–573.
- Yang, Z. & Nielsen, R. (1998). Synonymous and nonsynonymous rate variation in nuclear genes of mammals. *Journal of Molecular Evolution* **46**, 409–418.
- Yang, Z. H. (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. *Computer Applications in the Biosciences* 13, 555–556.
- Yang, Z. & Swanson, W. J. (2002). Codon-substitution models to detect adaptive evolution that account for heterogeneous selective pressures among site classes. *Molecular Biology and Evolution* 19, 49–57.
- Yang, Z., Wong, W. S. & Nielsen, R. (2005). Bayes empirical bayes inference of amino acid sites under positive selection. *Molecular Biology and Evolution* 22, 1107–1118.

3.3. Discussion

Contrairement aux processus neutres (BGC et biais de mutation) qui peuvent en théorie affecter tous les nucléotides du génome, la sélection agit sur les nucléotides pouvant modifier la valeur sélective des individus. Ceux-ci peuvent se trouver dans la partie codante des gènes (mutations non synonymes et générant des codons préférés), dans leurs régions régulatrices (5' et 3' UTR), voire dans les introns. On ne s'attend pas à ce qu'une sélection sur les mutations non synonymes affecte grandement la composition en bases du génome, contrairement à la sélection sur les codons préférés qui enrichit le génome en GC. Dans le Tableau 3.2., les processus moléculaires, neutres et sélectifs, sont différenciés selon leurs cibles. Il sera utilisé comme trame pour la discussion de nos résultats.

3.3.1. Comment interpréter les patrons de GC ?

L'enrichissement en GC engendré par le biais de réparation des hétérozygotes AT/GC par la BGC peut se confondre avec un biais mutationnel en faveur de bases GC. La sélection sur l'usage du code peut également être responsable d'un enrichissement en GC, si les codons préférés sont majoritairement riches en GC (c'est notamment le cas pour la drosophile, Marais, Mouchiroud, and Duret 2001; l'orge et le blé, Kawabe and Miyashita 2003; Liu and Xue 2005; Wang and Roossinck 2006).

Dans notre étude, nous avons montré que les autogames et les zones faiblement recombinantes avaient des valeurs d'équilibre en GC (GC^*) plus faibles que les allogames, et les régions fortement recombinantes. Conformément à l'attendu, les espèces allogames évoluent vers un équilibre en GC ($GC^*=0.453$) plus élevé que les espèces autogames ($GC^*=0.353$). La sélection sur le biais d'usage du code ne semble pas influencer fortement l'évolution en bases des génomes à l'échelle de notre étude. Pour avancer sur cette question, nous pourrions rechercher les traces d'une sélection sur le biais d'usage du code au travers d'une corrélation positive entre $GC3^*$, Fop^* (estimateurs du biais d'usage du code) et le niveau d'expression des gènes. En effet, la force de sélection sur l'usage du code semble corrélée au niveau d'expression des gènes (Ikemura 1985).

Tableau 3.2. : Prédictions des processus neutres et sélectifs sur différents paramètres d'évolution des séquences.

	BGC	Biais de Mutation		Sélection	
			Sur la Protéine	Sur la Régulation	Sur l'Usage du Code
cibles	tous sites	tous sites	mutations non synonymes	nucléotides fonctionnels	mutations synonymes
signatures moléculaires					
enrichissement en GC	oui	oui, si biais vers GC	non	non	oui, si codons préférés en G et/ou C
accumulation mutations non synonymes (d_N/d_S)	oui	non	oui	non	non
balayage sélectif	oui, de la longueur de la conversion	non	oui, avec effets d'entraînement	oui	oui en théorie, mais en général la sélection est trop faible pour qu'un site fasse un balayage. Interférences Hill- Robertson
Paramètres dont dépend le processus	γ , S, N_e	u/v, N _e	s, N_e	s', N _e	s '', N_e
Indicateurs	<i>GC*</i> , spectre de fréquence	<i>GC</i> *, spectre de fréquence des polymorphisme,	d_N/d_S	?	Fop*

Tableau modifié d'après Galtier and Duret (2007).

 N_e : la taille efficace de la population ; γ : la force de la BGC, quantifie le biais GC : AT ; S: le taux d'autofécondation ; u/v : rapport des taux de mutation de AT vers GC et de GC vers AT ; s: le coefficient de sélection ; s': le coefficient de sélection sur la régulation et s'': le coefficient de sélection sur le biais d'usage du code

Les données d'expression ne sont pas encore faciles à obtenir mais grâce à la quantité d'ESTs disponibles chez l'orge et le blé en provenance de banques de cDNA pour de nombreux organes, on pourrait par exemple estimer le niveau d'expression des gènes en utilisant la redondance des ESTs correspondant au gène considéré dans les banques de données publiques.

L'apport de données de polymorphisme pour interpréter les patrons de GC

Sous l'hypothèse d'un biais de mutation en faveur de G ou C, un plus grand nombre de mutations en G ou C se fixent dans la population, uniquement parce qu'elles apparaissent en plus forte fréquence. Lorsque les mutations sont apparues, leur devenir dans la population est uniquement fonction de la dérive. Si le biais de mutation est constant, la proportion relative des polymorphismes AT et GC est homogène pour toutes les classes de fréquences alléliques (Galtier, Bazin, and Bierne 2006). Le mécanisme d'enrichissement en GC lié à la BGC est différent. La BGC ne favorise pas l'apparition de mutations GC, mais agit sur les sites hétérozygotes en favorisant l'allèle G (ou C) par rapport à l'allèle A (ou T). Le spectre de fréquence de sites polymorphes sera biaisé vers des fortes fréquences pour les allèles G et C (Eyre-Walker 1999). Les mutations G et C voient donc augmenter leur probabilité de fixation sous l'effet de la BGC.

La distribution des fréquences des sites polymorphes GC vs. AT a été réalisée sur un échantillon de cinq gènes chez le blé situés dans une région sub-télomérique où le taux de recombinaison local est élevé (Figure 3.4., Haudry 2004). Les allèles G et C étaient significativement plus fréquents que les allèles A et T sur les sites en ségrégation dans les régions non codantes (p<0.005). De plus, les gènes étudiés sont situés dans une région sub-télomérique où le taux de recombinaison local est attendu très fort. La BGC étant fortement corrélée à la recombinaison, ces résultats suggèrent l'existence et l'efficacité de la BGC chez le blé. Plusieurs programmes de séquençage sont actuellement en cours qui vont amener une grande quantité d'informations sur le polymorphisme chez le blé. D'après la localisation des locus séquencés dans des bin de délétion, on pourrait contraster les spectres de fréquence des allèles GC vs. AT dans différentes zones du génome.

Fréquence des allèles GC

Figure 3.4. : Distribution des fréquences des allèles G ou C chez le blé

Le graphique représente la distribution de fréquences des allèles G ou C estimées sur les sites en ségrégation dans les régions non codantes au voisinage de cinq gènes.



3.3.2. De la BGC chez les Triticeae autogames ?

Marais et al. (2004) ont montré que l'efficacité de la BGC dépend à la fois de la taille efficace des populations et du taux d'autofécondation. Le modèle prédit l'inefficacité de la BGC dans des populations fortement autogames. L'absence de corrélation entre le GC^* et le taux de recombinaison local faite sur A. thaliana est conforme à ces attendus (Marais, Charlesworth, and Wright 2004). Or, notre étude de la distribution des polymorphismes GC vs. AT dans les populations de blé suggère l'existence de la BGC chez cette espèce autogame (taux d'allofécondation estimé entre 4 et 10%, Tsegaye 1996; Enjalbert and David 2000). Nous avons également mis en évidence l'impact des variations du taux de recombinaison sur les patrons d'enrichissement en GC au sein de la tribu des Triticeae, à la fois dans les espèces allogames S. cereale et Ae. speltoides et dans leurs apparentées autogames T. monococcum et T. urartu. Nos résultats suggèrent que les patrons de GC chez les Triticeae sont, au moins partiellement, construits par l'action de la BGC. Si l'intensité de ce mécanisme se révèle plus forte chez les espèces allogames, il semble qu'elle reste efficace dans les zones recombinantes dans les espèces autogames. Nous avons en effet observé chez les espèces autogames que les régions FR ($GC^* = 0.449$ et 0.385 dans les gènes télomériques et du bras court, respectivement) avaient de plus fortes valeurs de GC* que les régions fr faiblement recombinantes ($GC^* = 0.326$ et 0.352 dans les gènes centromériques et du bras long, respectivement).

Le fort gradient de recombinaison intra-chromosomique des *Triticeae* associé à un taux d'autofécondation plus faible que celui d'*A*. *thaliana* pourrait expliquer que l'on observe
une efficacité de la BGC chez les espèces autogames étudiées. La concentration des appariements dans les régions sub-télomériques permettrait à la BGC d'être efficace lors des rares événements d'allopollinisation enrichissant ainsi en GC les régions fortement recombinantes des espèces autogames de *Triticeae*.

3.3.3. Effets de la BGC sur l'efficacité de la sélection

Nos résultats suggèrent qu'en plus d'influencer la composition en bases des génomes, la BGC pourrait entraîner la fixation de mutations non synonymes, faiblement délétères en G ou C, interagissant ainsi avec la sélection. Si la BGC affecte les régions codantes et augmente les vitesses d'évolution (d_N/d_S) , elle mime paradoxalement les signatures moléculaires de sélection positive. La sélection positive est généralement invoquée pour expliquer des accélérations du taux de substitutions non synonymes dans une espèce en comparaison à des espèces apparentées (Yang 2002). Récemment l'hypothèse neutre alternative de la BGC a été proposée par Galtier et Duret (2007) pour expliquer l'augmentation du taux de substitution observée dans des régions non codantes de l'homme (Pollard et al. 2006) ou des régions codantes de la souris (Perry and Ashworth 1999). Le fait que les substitutions fixées soient plus souvent GC que AT suggère l'implication d'un processus neutre tel que la BGC (Pollard et al. 2006; Galtier and Duret 2007). Pour détecter un réel épisode de sélection positive, il conviendrait donc d'estimer l'intensité de la BGC et l'intégrer dans le modèle neutre.

Il est notamment possible d'estimer l'intensité de la BGC (γ) sur des données de polymorphisme en mesurant l'écart au spectre de fréquence attendu des bases AT *vs.* GC (Galtier, Bazin, and Bierne 2006). A partir de données de polymorphisme intraspécifique des gènes à fort *GC** de notre échantillon (chap 3.2.) chez *S. cereale, Ae. speltoides, T. monococcum* et *T. urartu*, il serait alors possible d'estimer la force de la BGC, l'intégrer dans la formalisation de l'attendu du d_N/d_S proposée par Charlesworth (1992) telle que :

$$\omega = \frac{4N_e(s-\gamma)}{e^{4N_e(s-\gamma)}-1}$$

où *s* est le coefficient de sélection et γ la force de la BGC.

La difficulté est d'arriver à estimer γ indépendamment des effets de la sélection. Les données de polymorphisme pourraient être récoltées sur les nucléotides synonymes des exons mais leur neutralité n'est pas assurée en raison de la sélection sur l'usage du code. Dans ce cas on estimerait γ (1+*s*'') (Tableau 3.2.). L'alternative serait d'utiliser les introns, en faisant l'hypothèse que la force de la BGC y soit identique. Chez les monocotylédones, on peut toutefois douter que ce soit le cas, en raison de l'existence du gradient de GC intragénique (Wong et al 2002).

Une autre façon de s'affranchir de l'effet de la BGC sur la détection de la sélection sur les gènes des *Triticeae* consisterait à contraster les d_N/d_S le long du gradient intra-génique. Sous l'hypothèse que ce gradient de GC traduit l'intensité relative de la recombinaison et conséquemment de la BGC, les régions 3' des gènes seraient moins soumises à l'action de la BGC en comparaison des régions 5'. C'est donc dans la région 3' que l'effet du régime de reproduction sur l'efficacité de la sélection purificatrice serait le moins masqué par la BGC.

3.3.4. BGC et implications sur l'étude du régime de reproduction

L'intensité de la BGC dépend à la fois de la taille efficace de la population et du taux d'autofécondation (Marais, Charlesworth, and Wright 2004). Dans le cas de faibles différences d'effectif efficace entre populations autogame et allogame, le taux d'autofécondation entraîne des différences importantes du GC^* , tandis que les patrons de sélection purificatrice ou d'usage du code seront similaires (chap. 4.2.). La BGC est un processus très sensible au taux d'hétérozygotie, et par conséquent aux changements de système de reproduction. De telles transitions peuvent donc être identifiées grâce au GC^* , même dans le cas de faible différence de N_e .

Prenons l'exemple de deux populations dérivant du même ancêtre, l'une autogame, l'autre allogame. Leur histoire évolutive différant, la population allogame peut souffrir d'une taille efficace plus faible que la population autogame. Pour tester l'impact de telles différences de tailles efficaces malgré le changement de régime de reproduction, j'ai calculé les valeurs que prendraient GC^* , Fop^* et ω (ratio d_N/d_S) dans les populations suivantes : N_{allogame} = 1000 ; N_{autogame} = 2000 ; S, le coefficient d'autofécondation = 0.95 ; s, le coefficient de sélection = 0.0002 et γ , l'intensité de la BGC = 0.0002. La Figure 3.5.A présente les résultats. L'efficacité de la sélection ($N_e s$) purificatrice et sur l'usage du code augmente avec N_e , elle est donc meilleure ici dans la population autogame. La fréquence des codons préférés est plus importante dans la population autogame ($Fop^* = 0.83$) que dans la population allogame ($Fop^* = 0.69$), tandis que la population allogame fixe plus de mutations non synonymes faiblement délétères ($\omega = 0.65$) que la population autogame ($GC^* = 0.40$). Seul le GC^* illustre la réduction du taux d'homozygotie dans la population autogame ($GC^* = 0.51$ vs. 0.69 dans la population allogame).

Si l'on s'intéresse maintenant à l'effet de la taille efficace (NI = 1000 et N2 = 2000, Figure 3.5.B), lorsque le régime de reproduction est autogame (S = 0.95), il apparaît que la population de grande taille est plus efficace en sélection ($\omega = 0.40$ et $Fop^* = 0.83$) que la population de taille faible efficace ($\omega = 0.65$ et $Fop^* = 0.69$) (Figure 3.5.B). Par contre, le GC^* de ces deux populations est similaire ($GC^* = 0.505$ pour NI et $GC^* = 0.510$ pour N2).



Figure 3.5. : Valeurs théoriques attendues pour GC^* , Fop^* et ω dans une population allogame et une population fortement autogame de grande taille.

Soient les tailles efficaces des populations telles que : $N_e = N_{\text{allogame}}$ et $N_e = N_{\text{autogame}}$ (1 - S/2), S étant le coefficient d'autofécondation. Pour des mutations codominantes, l'efficacité de la sélection est indépendante de S et $\omega = \frac{4Ns}{e^{4Ns} - 1}$ (Charlesworth 1992) ; $Fop^* = \frac{e^{4Ns}}{u/v + e^{4Ns}}$ (Bulmer 1991); $GC^* = \frac{e^{2N\gamma(1-S)(2-S)}}{u/v + e^{2N\gamma(1-S)(2-S)}}$ (Marais et al. 2004).

Où *N* est N_{allogame} chez les allogames et N_{autogame} chez les autogames, *s* le coefficient de selection contre les mutations délétères (ω) ou en faveur des codons préférés (Fop^*); γ l'intensité de la BGC ; *u* le taux de mutation vers les codons non préférés (Fop^*) ou vers les allèles AT (GC^*) et *v* le taux de mutation réciproque.

Pour les graphes, S = 0.95, $s = \gamma = 0.0002$, et u / v = 1.5. A: $N_{\text{allogame}} = 1000$ et $N_{\text{autogame}} = 2000$, *i.e.*, N_e (autogame) $\approx N_{\text{allogame}} = 1000$. B: $NI_{\text{autogame}} = 1000$ (à gauche) et $N2_{\text{autogame}} = 2000$ (à droite).

Le GC^* apparaît comme un bon marqueur des transitions de système de reproduction puisqu'il est peu sensible à la taille efficace des populations, mais particulièrement affecté par le taux d'autofécondation de la population. Le GC^* représente donc un bon indicateur de changements de régime de reproduction sur les branches d'une phylogénie, et ce malgré des perturbations de taille efficace des espèces concernées. Perspectives.

Perspectives

Le but de cette thèse était de comprendre comment les variations de taille efficace, qu'elles soient liées à la domestication ou à une transition du système de reproduction, avaient affecté l'évolution des gènes au sein des Triticées. Le régime de reproduction jouant sur le taux de recombinaison efficace, l'impact des variations globales et locales de la recombinaison sur l'évolution des gènes a également été étudié.

De la compréhension de la domestication à la gestion de l'agriculture actuelle...

Les évènements de domestication représentent de bons exemples de modifications récentes à la fois à l'échelle morphologique et génétique. Les transitions des formes sauvages vers les formes cultivées se sont traduites par des goulots d'étranglement dont l'intensité varie fortement d'une espèce à l'autre (Figure 1.3.). L'étude du polymorphisme des formes sauvages et cultivées de blé a mis en évidence d'une part la faible quantité de polymorphisme présent chez la forme sauvage, et d'autre part la très forte érosion de la diversité génétique associée à la domestication et aux autres épisodes historiques vécus par l'espèce. De même que pour les autres espèces cultivées, le réservoir de diversité représenté dans les formes sauvages pourrait être mobilisé pour répondre aux défis d'une agriculture soumise à des bouleversements climatiques et techniques. L'amidonnier sauvage T. dicoccoides a par exemple été utilisé pour améliorer la résistance à la rouille jaune (Grama and Gerechter-Amitai 1974). La précision croissante des cartes génétiques devrait permettre d'introduire des fragments chromosomiques plus petits, voire uniquement les gènes d'intérêt par des introgressions assistées par marqueurs. Cependant, le niveau de diversité quantifié au sein de la forme sauvage de blé est relativement faible par rapport à celui d'autres espèces, ce qui représente une limitation dans les possibilités d'identification et d'introgression de gènes d'intérêt agronomique dans les variétés nouvelles. Dans une optique d'amélioration variétale, il pourrait être plus avantageux d'envisager de développer les croisements avec des espèces plus éloignées phylogénétiquement présentant des niveaux de diversité plus importants, malgré les difficultés d'obtenir des descendants fertiles dans ce type de croisement. Il existe déjà de nombreux exemples d'introgression de gènes étrangers dans le blé concernant essentiellement des gènes de résistance aux maladies : résistance à l'oïdium venant de *T. timophevii* ou de la translocation entre le chromosome 1BL avec le 1RS du seigle, résistance au piétin-verse venant d'*Aegilops ventricosa* (Auriau et al. 1992). Cependant ces introgressions ont généralement entraîné d'importants fragments chromosomiques, pouvant ainsi introduire des gènes étrangers altérant la qualité d'utilisation du blé. En effet, les génomes divergeant entre les espèces, le taux de recombinaison efficace diminue et nous confronte à des problèmes génétiques pour introgresser uniquement les gènes d'intérêt. Une meilleure compréhension des patrons de recombinaison et des gènes contrôlant l'appariement des chromosomes pourrait permettre d'assurer de la recombinaison entre espèces proches.

En plus d'apporter une meilleure compréhension des évènements démographiques associés à l'histoire de l'agriculture, la comparaison des niveaux de diversité nucléotidique entre les formes sauvages et cultivées représente une méthode alternative de détection de gènes sous sélection (chap.2.3.2). L'identification des gènes impliqués dans les transitions agricoles devrait être permise par le développement des technologies modernes de reséquençage facilitant l'acquisition des données de séquences. Là où mon étude a porté sur une vingtaine de gènes, elle pourrait aujourd'hui certainement porter sur plusieurs milliers, grâce à l'augmentation exponentielle du nombre de séquences géniques disponibles dans les bases de données mondiales. Ainsi la distribution empirique des valeurs prises par les statistiques de diversité pour les gènes neutres permettrait d'estimer plus précisément les paramètres du scénario démographique de domestication, et par conséquent permettre une identification plus puissante de gènes sous sélection positive lors de la domestication et des autres transitions culturales (en réduisant l'incertitude sur la valeur moyenne et la distribution des paramètres). La complémentation de cette approche moléculaire par des études phénotypiques (génétique d'association, génétique quantitative) apparaît comme une étape indispensable pour identifier les caractères associés aux gènes sélectionnés.

Il y a 10 000 ans, les sociétés humaines ont connu un changement majeur avec la naissance de l'agriculture, de l'élevage et la sédentarisation. Une tradition de sélection des céréales et des autres espèces domestiquées a alors commencé et pour perdurer jusqu'à nos jours. Si les premiers agriculteurs du Néolithique ont sélectionné, plus ou moins consciemment, les plantes sur des composantes qu'ils pensaient liées à la productivité (comme la taille des grains), ils n'avaient pas les moyens d'évaluer d'autres propriétés agronomiques telles que l'apport nutritionnel. Ils ont sélectionné différentes variétés adaptées

aux différents usages domestiques (galettes, bouillies, ou encore alimentation des animaux) modifiant indirectement les compositions protéiques. Une réduction globale de 50% de la teneur en protéines des céréales cultivées par rapport aux formes sauvages a accompagné l'augmentation de la productivité (Doebley, Gaut, and Smith 2006). Outre l'effet de dilution de la quantité d'azote dans un grain plus gros qui diminue les teneurs protéiques, des allèles favorisant l'assimilation d'azote dans les grains, et ainsi la teneur en protéines, ont pu être perdus par dérive. De tels allèles ont en effet été identifiés et cartographiés au sein de la forme sauvage de blé , transférés dans des lignes de blés durs, ils ont permis un gain de 3% de la teneur en protéines (revue Shewry 2007). Grâce à l'identification de nombreux couples gènes/fonctions et aux outils de sélection moderne, il semble légitime d'envisager une amélioration des variétés à la fois d'un point de vue productivité et qualité nutritionnelle, tout en limitant l'impact de cette agriculture sur l'environnement.

Sexe et efficacité de la sélection : des relations conflictuelles

Le système de reproduction et le taux de recombinaison sont des facteurs connus pour affecter l'évolution du génome et l'efficacité de la sélection naturelle via les variations de taille efficace qu'ils entraînent (Charlesworth and Wright 2001). L'étude comparative des taux l'évolution des gènes dans de quatre espèces proches autogames et allogames de Triticeae a pu mettre en évidence deux résultats majeurs. Tandis que l'évolution de la teneur en bases G et C, traduisant l'intensité de la conversion génique biaisée vers GC (BGC), est clairement affectée par le régime de reproduction et le taux de recombinaison local, les résultats sont moins évidents en ce qui concerne l'efficacité de la sélection. Afin de tester l'impact des variations locales du taux de recombinaison sur l'évolution des gènes, nous avons contrasté des régions fortement et faiblement recombinantes (respectivement nommées FR et fr). Or, un gradient de recombinaison intrachromosomique a été identifié sur les chromosomes de blé et d'orge (Lukaszewski 1992; Lukaszewski 1993; Akhunov et al. 2003b; (Kunzel, Korzun, and Meister 2000) et pourrait être la norme chez l'ensemble des espèces de Triticeae, suggérant qu'il existe toute une gamme d'intensités de recombinaison le long des chromosomes. Le programme de séquençage du génome du blé, ainsi que le programme français de cartographie physique et génétique du chromosome 3B (ANR Exegese-Blé), devrait permettre dans un futur proche de plus précises estimations du taux de recombinaison

local. Il sera alors possible de tester de manière plus fine l'impact de la recombinaison sur l'évolution des gènes, de leur dynamique d'évolution en GC, en l'intégrant comme une variable continue plutôt qu'en distinguant juste deux classes (FR et fr). Afin de pouvoir généraliser les observations réalisées au niveau du chromosome 3B à l'ensemble du génome, il faudrait compléter l'analyse de l'impact des variations locales de la recombinaison en échantillonnant des gènes dans différents chromosomes.

Enfin, un excès de mutations non synonymes a été détecté chez les espèces allogames par rapport aux autogames, affectant les gènes présentant une importante intensité de BGC. Ces résultats soutiennent l'hypothèse d'une interaction entre la BGC et la sélection proposée par Bengtsson (1990), qui propose que les espèces avec une forte efficacité de la BGC pourraient être sujettes à un fardeau de mutation. La confirmation d'un fardeau associé à la recombinaison *via* la BGC apporterait un nouvel élément de discussion sur l'évolution des systèmes de reproduction. Comme Galtier et Duret (2007) l'ont proposé pour l'homme, les espèces allogames auraient un « talon d'Achille génomique », inexistant ou plus faible chez les espèces autogames présentant une intensité de BGC réduite du fait de leur homozygotie. Afin de comprendre comment ce potentiel fardeau de mutation lié à la BGC pourrait affecter les transitions de régime de reproduction, il serait intéressant d'intégrer la force de la BGC aux modèles de dépression de consanguinité et de fardeau de mutation.

De manière plus générale, la BGC mimant les effets moléculaires de la sélection positive en favorisant les allèles G et C par rapport aux allèles A et T, l'identifier et quantifier son intensité apparaissent une priorité afin de la prendre en compte dans les futures études de détection de gènes sous sélection positive.

La BGC, un phénomène universel ?

Nos résultats suggèrent que l'évolution de la teneur en bases des génomes chez les *Triticeae* est en partie déterminée par la BGC, chez les espèces allogames ainsi que chez les autogames. L'action de la BGC est pour l'instant considérée comme la force responsable de la structure en isochores des génomes des Mammifères et des Oiseaux (Galtier et al. 2001; Meunier and Duret 2004; Duret, Eyre-Walker, and Galtier 2006) et a été observée dans d'autres espèces telles que la drosophile ou la levure (revue Marais 2003). Mais peut-on en

déduire qu'il s'agit d'une force universelle ? De nombreuses espèces (bactéries, levure, hamster, homme) montrent un excès de teneur en GC observée par rapport à celle attendue sous un modèle d'évolution de la composition en bases des génomes conduit par la mutation seule (Lynch 2007). Un modèle incorporant le biais de GC par la conversion génique tend à montrer que l'excès de GC observé dans ces espèces pourrait être la résultante de l'action de la conversion génique biaisée (Lynch 2007).

Chez les plantes, des patrons de teneur en GC très distincts ont été mis en évidence entre des espèces de Graminées et de Dicotylédones (Carels and Bernardi 2000). Les Graminées sont à la fois plus riches en GC et plus hétérogènes que les Dicotylédones (Carels and Bernardi 2000; Wong et al. 2002). Ces différents patrons de composition en bases pourraient refléter à la fois une activité plus importante de la BGC chez les Graminées ainsi qu'une plus forte hétérogénéité des patrons de recombinaison. **Il se pourrait que la BGC soit suffisamment intense pour influencer l'évolution de la teneur en GC des génomes et refléter les patrons de recombinaison uniquement chez les Graminées.** La raison pour laquelle l'intensité de la BGC serait plus importante ou que les patrons de recombinaison seraient particuliers chez les Graminées reste cependant une question sans réponse pour l'heure.

Le modèle Brassicaceae

Grâce à la récente publication de plusieurs génomes complets au sein de la famille, les *Brassicaceae* offrent une excellente opportunité de comparer l'évolution moléculaire à l'échelle du génome chez des plantes différant pour leur régime de reproduction. Une analyse d'évolution génomique comparative pourrait en particulier être menée entre l'autogame *Arabidopsis thaliana* et l'espèce sœur allogame *A. lyrata*, en orientant les transitions grâce à l'utilisation de *Capsella rubella* comme groupe externe (Figure 4.1.). Grâce à ce complexe de trois espèces séquencées, de nombreuses questions concernant l'évolution des génomes en fonction du système de reproduction pourront être approfondies.



Figure 4.1. : Espèces de la famille Brassicaceae dont le génome est entièrement séquencé.

- 1) l'allogame Arabidopsis lyrata
- 2) l'autogame Arabidopsis thaliana
- 3) leur proche apparentée autocompatible Capsella rubella

La réduction de la taille des fleurs chez *A.thaliana* associée à la transition vers l'autogamie est illustrée en 4) (photo tirée de (Nasrallah et al. 2000). Les pétales d'*A. lyrata* (A) sont approximativement 30 fois plus grands que ceux d'*A. thaliana* (B).

Une étude précédente basée sur les taux de substitution et le biais d'usage du code estimés au niveau de 24 locus n'a pas mis en évidence de différence d'efficacité de la sélection entre *A. thaliana* et *A. lyrata* (Wright, Lauga, and Charlesworth 2002). Cependant, la puissance de détection de variations des patrons d'évolution moléculaire pouvait être trop faible avec un tel jeu de données (Wright et al. 2008). Cette limitation de puissance devrait être écartée par l'utilisation des données de séquences des génomes complets. A partir des plus de 25 000 gènes codant pour des protéines (The Arabidopsis Genome Initiative 2000), serait-il possible de détecter une augmentation de d_N/d_S sur la branche d' *A. thaliana* associée à la transition de système de reproduction vers l'autogamie ?

Notre étude sur l'impact du système de reproduction a également mis en évidence l'apport de puissance permis par l'utilisation de statistiques dynamiques (GC^* et Fop^*) plutôt que de statistiques stationnaires (GC% et Fop) pour détecter des variations de patrons d'évolution. Les valeurs d'équilibre étant estimées uniquement sur la base des sites ayant subi des substitutions sur les branches étudiées, elles sont plus puissantes pour révéler les changements des processus impliqués. En effet, si les teneurs en GC sont similaires entre *Triticeae* autogames et allogames, les valeurs d'équilibre (GC^*) vers lesquelles tendent les espèces allogames sont significativement plus fortes que pour les espèces autogames (Haudry et al. 2008). A ce jour, aucune différence significative n'a pu être observée entre *A. thaliana* et *A. lyrata* en ce qui concerne la teneur en GC et le biais d'usage du code (Wright, Lauga, and Charlesworth 2002). En utilisant l'espèce *Capsella rubella* afin d'orienter les transitions sur la phylogénie, il serait possible d'augmenter encore la puissance de détection de l'impact du système de reproduction sur l'évolution des génomes chez les *Brassicaceae* en comparant les valeurs d'équilibre GC^* et *Fop** estimées pour chaque espèce.

Le rôle majeur de la recombinaison et de la BGC sur l'évolution de la composition en bases des génomes se trouvera-t-il confirmé au sein des *Brassicaceae* ? Une étude théorique suggère que l'intensité de la BGC serait insuffisante chez une espèce aussi fortement autogame qu'*A. thaliana* du fait de son homozygotie (Marais, Charlesworth, and Wright 2004). Dans ce cas, et si la BGC est une force qui affecte également le génome des *Brassicaceae*, il devrait être possible de détecter un changement de la teneur en GC d'équilibre sur la branche menant à *A. thaliana* par rapport à *A. lyrata*. Même s'il n'existe qu'une faible différence entre les tailles efficaces de *A. thaliana* et de *A. lyrata* du fait de différentes histoires démographiques (hypothèse suggérée par Wright, Lauga, and Charlesworth 2002), une réduction de la teneur d'équilibre en GC est attendue chez l'espèce autogame par rapport à l'espèce allogame sous le modèle de BGC (Marais, Charlesworth, and Wright 2004). En effet, les estimateurs *GC** sont beaucoup plus sensibles aux variations du taux d'hétérozygotie qu'à la taille efficace des populations (chap. 3.3.4.). L'absence de différence de patrons d'évolution de la teneur en GC entre *A. thaliana* et de *A. lyrata* pourrait argumenter en défaveur de l'existence de la BGC chez les *Brassicaceae*.

Enfin, sur un jeu de données composé de plus de 25 000 gènes, il serait possible de distinguer différentes classes de gènes en fonction du type de pression sélective auquel ils sont soumis (estimés par des ratios d_N/d_S moyens entre autogames et allogames), de la position sur le chromosome, de la teneur en GC, ou encore du taux de recombinaison local. Parmi ces différentes classes de gènes, lesquelles permettent de mettre en évidence l'effet du régime de reproduction ? Par exemple, si l'action de la BGC se trouve confirmée chez A. *lyrata*, celle-ci pourrait interagir avec la sélection naturelle et brouiller le signal de relâchement de la sélection prédit chez A. *thaliana*. Dans ce cas, **les gènes présentant des intensités plus faibles de BGC pourraient subir l'augmentation des d_N/d_S associée au fardeau d'autofécondation tandis que les gènes avec une forte activité de BGC pourraient confirmer l'existence d'un fardeau génétique lié à la BGC chez les allogames.**

L'accès à l'information de séquence du génome complet devrait nous permettre d'étudier l'impact du système de reproduction sur **l'évolution des génomes**, autant que sur l'évolution des gènes. En facilitant l'obtention de données de séquences pour les régions non codantes, mais également en nous plaçant à une échelle d'observation supérieure à celle du gène, de nouvelles perspectives seront offertes. Il sera alors possible d'étudier l'évolution des introns, des éléments transposables, et les changements dans l'organisation du génome via la position des gènes dans le génome, le nombre de duplications... Par exemple, le relâchement de l'efficacité de la sélection attendu chez les autogames devrait entraîner l'accumulation d'éléments transposables dans le génome d'*A. thaliana* par rapport à *A. lyrata* (Wright et al. 2008).

Les différentes études présentées dans ce document permettent d'illustrer **comment interpréter les patrons de polymorphisme et de divergence de séquences nucléotidiques pour documenter l'histoire évolutive des espèces.** Les données de séquences connaissant un réel « boom » depuis ces dernières années, les perspectives qui s'offrent actuellement pour les études d'évolution moléculaire sont multiples. La possibilité d'élargir à la fois le nombre d'espèces et de gènes analysés ainsi qu'une meilleure connaissance des patrons de recombinaison grâce aux avancées de la cartographie physique permettront d'affiner notre compréhension des mécanismes neutres et sélectifs évoqués dans cette thèse et de leurs interactions. Bibliographie.

Bibliographie

- Akashi H, Kliman R, Eyre-Walker A. 1998. Mutation pressure, natural selection, and the evolution of base composition in Drosophila. Genetica 102-103:49-60.
- Akhunov ED, David JL, Chao S, Lazo G, Anderson OD, Qi L, Echalier B, Gill BS, Linkiewicz AM, Dubcovsky J, Miftahudin, Gustafson JP, La Rota CM, Sorrells ME, Zhang DS, Nguyen HT, Hossain K, Kianian SF, Peng J, Lapitan NLV, Sidhu D, Gill KS, McGuire PE, Qualset CO, Dvorak J. 2003a. GC composition and codon usage in genes of inbreeding and outcrossing *Triticeae* species. Pp. 203-206. Tenth international wheat genetics symposium, Paestum, Italy.
- Akhunov ED, Goodyear AW, Geng S, Qi LL, Echalier B, Gill BS, Miftahudin, Gustafson JP, Lazo G, Chao SM, Anderson OD, Linkiewicz AM, Dubcovsky J, La Rota M, Sorrells ME, Zhang DS, Nguyen HT, Kalavacharla V, Hossain K, Kianian SF, Peng JH, Lapitan NLV, Gonzalez-Hernandeiz JL, Anderson JA, Choi DW, Close TJ, Dilbirligi M, Gill KS, Walker-Simmons MK, Steber C, McGuire PE, Qualset CO, Dvorak J. 2003b. The organization and rate of evolution of wheat genomes are correlated with recombination rates along chromosome arms. Genome Res. 13:753-763.
- Allaby RC, Banerjee M, Brown TA. 1999. Evolution of the high molecular weight glutenin loci of the A, B, D, and G genomes of wheat. Genome 42:296-307.
- Auriau P, Doussinault G, Jahier J, Lecomte C, Pierre J, Pluchard P, Rousset M, Saur L, Trottet M. 1992. Le blé tendre. In: Editions INRA, ed. Amélioration des espèces végétales cultivées. Objectifs et critères de sélection. Ouvrage collectif coordonné par A. Gallais et H. Bannerot, Paris. p. 22-38.
- Bar-Yosef O. 1998. The Natufian culture in the Levant, threshold to the origins of agriculture. Evolutionary Anthropology 6:159-177.
- Begun DJ, Aquadro CF. 1992. Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. Nature 356:519-520.
- Birdsell JA. 2002. Integrating genomics, bioinformatics, and classical genetics to study the effects of recombination on genome evolution. Mol. Biol. Evol. 19:1181-1197.
- Braverman JM, Hudson RR, Kaplan NL, Langley CH, Stephan W. 1995. The Hitchhiking Effect on the Site Frequency-Spectrum of DNA Polymorphisms. Genetics 140:783-796.
- Brown AHD, Zohary D, Nevo E. 1978. Outcrossing rates and heterozygosity in natural populations of Hordeum spontaneum Koch in Israel. Heredity 41.
- Buckler E, Thornsberry JM, Kresovich S. 2001. Molecular diversity, structure and domestication of grasses. Genet. Res. 77.
- Bulmer M. 1991. The Selection-Mutation-Drift Theory of Synonymous Codon Usage. Genetics 129:897-907.
- Caicedo AL, Williamson SH, Hernandez RD, Boyko A, Fledel-Alon A, York TL, Polato NR, Olsen KM, Nielsen R, McCouch SR, Bustamante CD, Purugganan MD. 2007. Genome-Wide Patterns of Nucleotide Polymorphism in Domesticated Rice. PLoS Genet. 3:e163.
- Caldwell KS, Russell J, Langridge P, Powell W. 2006. Extreme population-dependent linkage disequilibrium detected in an inbreeding plant species, *Hordeum vulgare*. Genetics 172:557-567.
- Carels N, Bernardi G. 2000. Two classes of genes in plants. Genetics 154:1819-1825.
- Cenci A, Chantret N, Haudry A, Poirier S, Hochu I, Santoni S, Gautier M-F, Joudrier P, Bataillon T, David J. in prep. Comparative analysis at the *Ha* locus of the two

tetraploid sister species, T. turgidum ssp. dicoccoides and T. timopheevii ssp. armeniacum.

- Chamary JV, Parmley JL, Hurst LD. 2006. Hearing silence: non-neutral evolution at synonymous sites in mammals. Nature Rev. Genet. 7:98-108.
- Chantret N, Salse J, Sabot F, Rahman S, Bellec A, Laubin B, Dubois I, Dossat C, Sourdille P, Joudrier P, Gautier MF, Cattolico L, Beckert M, Aubourg S, Weissenbach J, Caboche M, Bernard M, Leroy P, Chalhoub B. 2005. Molecular basis of evolutionary events that shaped the hardness locus in diploid and polyploid wheat species (*Triticum* and *Aegilops*). Plant Cell 17:1033-1045.
- Charlesworth B. 1996. Background selection and patterns of genetic diversity in *Drosophila melanogaster*. Genet. Res. 68:131-149.
- Charlesworth B. 1992. Evolutionary Rates in Partially Self-Fertilizing Species. Am. Nat. 140:126-148.
- Charlesworth B. 1994. The Effect of Background Selection against Deleterious Mutations on Weakly Selected, Linked Variants. Genet. Res. 63:213-227.
- Charlesworth B, Morgan MT, Charlesworth D. 1993. The Effect of Deleterious Mutations on Neutral Molecular Variation. Genetics 134:1289-1303.
- Charlesworth D, Wright SI. 2001. Breeding systems and genome evolution. Current Opinion in Genetics & Development 11:685-690.
- Cockram J, Jones H, Leigh FJ, O'Sullivan D, Powell W, Laurie DA, Greenland AJ. 2007. Control of flowering time in temperate cereals: genes, domestication, and sustainable productivity. J. Exp. Bot. 58:1231-1244.
- Coop G, Przeworski M. 2007. An evolutionary view of human recombination. Nature Rev. Genet. 8:23-34.
- Cruden RW. 1977. Pollen-ovule ratios: a conservative indicator of breeding systems in flowering plants. Evolution 31.
- Darwin C. 1883. The Variation of Animals and Plants under Domestication. D.Appleton & Co, New York.
- Daud HM, Gustafson JP. 1996. Molecular evidence for *Triticum speltoides* as a B-genome progenitor of wheat (*Triticum aestivum*). Genome 39:543-548.
- de Visser JAGM, Elena SF. 2007. The evolution of sex: empirical insights into the roles of epistasis and drift. Nature Rev. Genet. 8:139-149.
- DeSalle R, Templeton AR. 1988. Founder Effects and the Rate of Mitochondrial DNA Evolution in Hawaiian Drosophila. Evolution 42:1076-1084.
- Doebley J. 1989. Isozymic evidence and the evolution. of crop plants. In: Soltis D. E. Soltis and P. S. ed. Isozymes in plant biology. Dioscorides, Portland, Oregon. p. 165-191.
- Doebley JF, Gaut BS, Smith BD. 2006. The Molecular Genetics of Crop Domestication. Cell 127:1309-1321.
- Dubcovsky J, Dvorak J. 2007. Genome Plasticity a Key Factor in the Success of Polyploid Wheat Under Domestication. Science 316:1862-1866.
- Dvorak J, Akhunov ED. 2005. Tempos of Gene Locus Deletions and Duplications and Their Relationship to Recombination Rate During Diploid and Polyploid Evolution in the *Aegilops-Triticum* Alliance. Genetics 171:323-332.
- Dvorak J, Diterlizzi P, Zhang HB, Resta P. 1993. The Evolution of Polyploid Wheats Identification of the a-Genome Donor Species. Genome 36:21-31.
- Dvorak J, Luo M-C, Yang Z-L. 1998. Restriction Fragment Length Polymorphism and divergence in the genomic regions of high and low recombination in self-fertilizing and cross-fertilizing *Aegilops* species. Genetics 148:423-434.
- Dvorak J, Luo M-C, Yang Z-L, Zhang H-B. 1998. The structure of the *Aegilops tauschii* genepool and the evolution of hexaploid wheat. Theor. Appl. Genet. 97:657-670.

- Dvorak J, Zhang H. 1990. Variation in Repeated Nucleotide Sequences Sheds Light on the Phylogeny of the Wheat B and G Genomes. Proceedings of the National Academy of Science 87:9640-9644.
- Eckardt NA. 2001. A sense of self: The role of DNA sequence elimination in allopolyploidization. Plant Cell 13:1699-1704.
- Endo TR, Gill BS. 1996. The Deletion Stocks of Common Wheat. J. Hered. 87:295-307.
- Enjalbert J, David JL. 2000. Inferring the recent history of outcrossing using multilocus individual heterozygosity (MIH). Application to naturally evolving populations of wheat. Genetics 156:1973-1982.
- Eyre-Walker A. 1999. Evidence of selection on silent site base composition in mammals: Potential implications for the evolution of isochores and junk DNA. Genetics 152:675-683.
- Eyre-Walker A, Gaut RL, Hilton H, Feldman DL, Gaut BS. 1998. Investigation of the bottleneck leading to the domestication of maize. Proc. Natl Acad. Sci. USA 95:4441-4446.
- Eyre-Walker A, Hurst LD. 2001. The evolution of isochores. Nature Rev. Genet. 2:549-555.
- Felsenstein J. 2006. Accuracy of Coalescent Likelihood Estimates: Do We Need More Sites, More Sequences, or More Loci? Mol. Biol. Evol. 23:691-700.
- Frankel O, Brown AHD, Burdon J. 1995. The Conservation of Plant Biodiversity. Cambridge University Press, Cambridge, UK.
- Galtier N, Bazin E, Bierne N. 2006. GC-biased segregation of noncoding polymorphisms in *Drosophila*. Genetics 172:221-228.
- Galtier N, Duret L. 2007. Adaptation or biased gene conversion? Extending the null hypothesis of molecular evolution. Trends Genet. 23:273-277.
- Galtier N, Piganeau G, Mouchiroud D, Duret L. 2001. GC-content evolution in mammalian genomes: The biased gene conversion hypothesis. Genetics 159:907-911.
- Gaut BS, Clegg MT. 1993. Nucleotide Polymorphism in the *Adh1* Locus of Pearl Millet (*Pennisetum glaucum*) (*Poaceae*). Genetics 135:1091-1097.
- Gerton JL, DeRisi J, Shroff R, Lichten M, Brown PO, Petes TD. 2000. Inaugural Article: Global mapping of meiotic recombination hotspots and coldspots in the yeast Saccharomyces cerevisiae. Proceedings of the National Academy of Sciences 97:11383-11390.
- Glémin S. 2007. Mating Systems and the Efficacy of Selection at the Molecular Level. Genetics 177:905-916.
- Glémin S, Bazin E, Charlesworth D. 2006. Impact of mating systems on patterns of sequence polymorphism in flowering plants. Proc. Biol. Sci. 273:3011-3019.
- Glinka S, Ometto L, Mousset S, Stephan W, De Lorenzo D. 2003. Demography and natural selection have shaped genetic variation in *Drosophila melanogaster*: A multi-locus approach. Genetics 165:1269-1278.
- Goldman N, Yang Z. 1994. A codon-based model of nucleotide substitution for proteincoding DNA sequences. Mol. Biol. Evol. 11:725-736.
- Grama A, Gerechter-Amitai ZK. 1974. Inheritance of resistance to stripe rust (Puccinia striiformis) in crosses between wild emmer (Triticum dicoccoides) and cultivated tetraploid and hexaploid wheats. II. Triticum aestivum. Euphytica 23:393-398.
- Griffiths S, Sharp R, Foote T, Bertin I, Wanous M, Reader S, Colas I, Moore G. 2006. Molecular characterization of *Ph1* as a major chromosome pairing locus in polyploid wheat. Nature 439:749-752.
- Hackauf B, Wehling P. 2005. Approaching the self-incompatibility locus Z in rye (*Secale cereale* L.) via comparative genetics. Theor. Appl. Genet. 110:832-845.

Haddrill P, Halligan D, Tomaras D, Charlesworth B. 2007. Reduced efficacy of selection in regions of the *Drosophila* genome that lack crossing over. Genome Biol. 8:R18.

- Hamrick JL, Godt MJW. 1996. Effects of life history traits on genetic diversity in plant species. Philos. Trans. R. Soc. Lond. B Biol. Sci. 351:1291-1298.
- Haudry A. 2004. Evolution du polymorphisme de séquence au cours de l'histoire évolutive des blés domestiques. Mémoire de DEA UMII, Montpellier.
- Haudry A, Cenci A, Ravel C, Bataillon T, Brunel D, Poncet C, Hochu I, Poirier S, Santoni S, Glemin S, David J. 2007. Grinding up Wheat: A Massive Loss of Nucleotide Diversity Since Domestication. Mol. Biol. Evol. 24:1506-1517.
- Heslop-Harrison JS. 1991. The Molecular Cytogenetics of Plants. J. Cell Sci. 100:15-21.
- Hill WG, Robertson A. 1966. The effect of linkage on limits to artificial selection. Genet. Res. 8:269–294.
- Huang S, Sirikhachornkit A, Su XJ, Faris J, Gill B, Haselkorn R, Gornicki P. 2002a. Genes encoding plastid acetyl-CoA carboxylase and 3-phosphoglycerate kinase of the Triticum/Aegilops complex and the evolutionary history of polyploid wheat. Proc. Natl Acad. Sci. USA 99:8133-8138.
- Huang SX, Sirikhachornkit A, Faris JD, Su XJ, Gill BS, Haselkorn R, Gornicki P. 2002b. Phylogenetic analysis of the acetyl-CoA carboxylase and 3-phosphoglycerate kinase loci in wheat and other grasses. Plant Molecular Biology 48:805-820.
- Hubby JL, Lewontin RC. 1966. A Molecular Approach to the Study of Genic Heterozygosity in Natural Populations.I. The Number of Alleles at Different Loci in Drosophila Pseudoobscura. Genetics 54:577-594.
- Hudson RR. 2002. Generating samples under a Wright-Fisher neutral model of genetic variation. Bioinformatics 18:337-338.
- Hudson RR. 1990. Gene genealogies and the coalescent process. In: Futuyma D. J., Antonovics J., eds. Oxford surveys in evolutionary biology. Oxford University Press, New York. p. 1-44.
- Hudson RR, Kaplan NL. 1995. Deleterious Background Selection With Recombination. Genetics 141:1605-1617.
- Hyten DL, Song Q, Zhu Y, Choi I-Y, Nelson RL, Costa JM, Specht JE, Shoemaker RC, Cregan PB. 2006. Impacts of genetic bottlenecks on soybean genome diversity. Proceedings of the National Academy of Science of United States of America 103:16666-16671.
- Ikemura T. 1985. Codon usage and tRNA content in unicellular and multicellular organisms. Molecular Biology and Evolution 2:13–35.
- Ingvarsson PK. 2002. A metapopulation perspective on genetic diversity and differentiation in partially self-fertilizing plants. Evolution 56:2368-2373.
- Innan H, Kim Y. 2004. Pattern of polymorphism after strong artificial selection in a domestication event. Proc. Natl Acad. Sci. USA 101:10667-10672.
- Innan H, Nordborg M. 2003. The extent linkage disequilibrium and haplotype sharing around a polymorphic site. Genetics 165:437-444.
- Jensen-Seaman MI, Furey TS, Payseur BA, Lu Y, Roskin KM, Chen C-F, Thomas MA, Haussler D, Jacob HJ. 2004. Comparative Recombination Rates in the Rat, Mouse, and Human Genomes. Genome Res. 14:528-538.
- Johnson KP, Seger J. 2001. Elevated rates of nonsynonymous substitution in island birds. Mol. Biol. Evol. 18:874-881.
- Kaplan NL, Hudson RR, Langley CH. 1989. The ``Hitchhiking Effect" Revisited. Genetics 123:887-899.
- Kawabe A, Miyashita NT. 2003. Patterns of codon usage bias in three dicot and four monocot plant species. Genes & Genetic Systems 78:343-352.

- Keightley PD, Eyre-Walker A. 2000. Deleterious nutations and the evolution of sex. Science 290:331-333.
- Kihara H. 1944. Discovery of the DD-analyser, one of the ancestors of *Triticum vulgare*. Agric. Horticulture (Tokyo) 19:13-14.
- Kilian B, Ozkan H, Deusch O, Effgen S, Brandolini A, Kohl J, Martin W, Salamini F. 2007. Independent Wheat B and G Genome Origins in Outcrossing Aegilops Progenitor Haplotypes. Mol. Biol. Evol. 24:217-227.
- Kilian B, Ozkan H, Kohl J, von Haeseler A, Barale F, Deusch O, Brandolini A, Yucel C, Martin W, Salamini F. 2006. Haplotype structure at seven barley genes: relevance to gene pool bottlenecks, phylogeny of ear type and site of barley domestication. Mol. Genet. Genomics 276:230-241.
- Kim S, Plagnol V, Hu TT, Toomajian C, Clark RM, Ossowski S, Ecker JR, Weigel D, Nordborg M. 2007. Recombination and linkage disequilibrium in Arabidopsis thaliana. Nature Genet. 39:1151-1155.
- Kimura M. 1983. The neutral theory of molecular evolution, Cambridge, UK.
- Kimura M, Ohta T. 1971. Theoretical Aspects of Population Genetics. Princeton University Press, New Jersey.
- Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, Richardsson B, Sigurdardottir S, Barnard J, Hallbeck B, Masson G, Shlien A, Palsson ST, Frigge ML, Thorgeirsson TE, Gulcher JR, Stefansson K. 2002. A high-resolution recombination map of the human genome. Nature Genet. 31:241-247.
- Kraft T, Sall T, Magnusson-Rading I, Nilsson N-O, Hallden C. 1998. Positive Correlation Between Recombination Rates and Levels of Genetic Variation in Natural Populations of Sea Beet (Beta vulgaris subsp. maritima). Genetics 150:1239-1244.
- Kunzel G, Korzun L, Meister A. 2000. Cytologically integrated physical restriction fragment length polymorphism maps for the barley genome based on translocation breakpoints. Genetics 154:397-412.
- Lamb BC. 1984. The properties of meiotic gene conversion important in its effects on evolution. Heredity 53.
- Lev-Yadun S, Gopher A, Abbo S. 2000. Archaeology Che cradle of agriculture. Science 288:1602-1603.
- Lewontin RC, Hubby JL. 1966. A Molecular Approach to the Study of Genic Heterozygosity in Natural Populations.II. Amount of Variation and Degree of Heterozygosity in natural populations of Drosophila Pseudoobscura. Genetics 54:595-609.
- Liu A, Burke JM. 2006. Patterns of Nucleotide Diversity in Wild and Cultivated Sunflower. Genetics 173:321-330.
- Liu QP, Xue QZ. 2005. Comparative studies on codon usage pattern of chloroplasts and their host nuclear genes in four plant species. Journal of Genetics 84:55-62.
- Lukaszewski AJ. 1992. A comparison of physical distribution of recombination in chromosome 1R in diploid rye and in hexaploid triticale. Theor. Appl. Genet. 83:1048-1053.
- Lukaszewski AJ, and C.A. Curtis. 1993. Physical distribution of recombination in B genome chromosome of tetraploid wheat. Theor. Appl. Genet. 86:121-127.
- Lundqvist A. 1954. Studies on self-sterility in rye, Secale cereale L. Hereditas 40:278-294.
- Luo M-C, Yang Z-L, You FM, Kawahara T, Waines JG, Dvorak J. 2007. The structure of wild and domesticated emmer wheat populations, gene flow between them, and the site of emmer domestication. Theor. Appl. Genet. 114:947-959.
- Lynch M. 2007. The Nucleotide Composition Landscape. The Origins of Genome Architecture. Sinauer Associates, Sunderland, Massachusetts. p. 121-150.

- Marais G. 2003. Biased gene conversion : implications for genome and sex evolution. Trends Genet. 19:330-338.
- Marais G, Charlesworth B, Wright SI. 2004. Recombination and base composition: the case of the highly self-fertilizing plant Arabidopsis thaliana. Genome Biol. 5:R45.
- Marais G, Mouchiroud D, Duret L. 2001. Does recombination improve selection on codon usage? Lessons from nematode and fly complete genomes. Proc. Natl Acad. Sci. USA 98:5688-5692.
- Maynard Smith J, Haigh J. 1974. The hitch-hiking effect of a favourable gene. Genet. Res. 23:23-55.
- McFadden ES, Sears ER. 1946. The origine of *Triticum spelta* and its free-threshing hexaploid relatives. J. Hered. 37:81-89, 107-116.
- McVean GAT, Charlesworth B. 2000. The effects of Hill-Robertson interference between weakly selected mutations on patterns of molecular evolution and variation. Genetics 155:929-944.
- Meunier J, Duret L. 2004. Recombination Drives the Evolution of GC-Content in the Human Genome. Mol. Biol. Evol. 21:984-990.
- Miyashita NT, Mori N, Tsunewaki K. 1994. Molecular Variation in Chloroplast DNA Regions in Ancestral Species of Wheat. Genetics 137:883-889.
- Moeller DA, Tenaillon MI, Tiffin P. 2007. Population Structure and Its Effects on Patterns of Nucleotide Polymorphism in Teosinte (Zea mays ssp. parviglumis). Genetics 176:1799-1809.
- Mori N, Liu YG, Tsunewaki K. 1995. Wheat Phylogeny Determined by Rflp Analysis of Nuclear-DNA .2. Wild Tetraploid Wheats. Theor. Appl. Genet. 90:129-134.
- Muller M-H, Poncet C, Prosperi JM, Santoni S, Ronfort J. 2006. Domestication history in the Medicago sativa species complex: inferences from nuclear sequence polymorphism. Mol. Ecol. 15:1589-1602.
- Munkvold JD, Greene RA, Bertmudez-Kandianis CE, La Rota CM, Edwards H, Sorrells SF, Dake T, Benscher D, Kantety R, Linkiewicz AM, Dubcovsky J, Akhunov ED, Dvorak J, Mifahudin, Gustafson JP, Pathan MS, Nguyen HT, Matthews DE, Chao S, Lazo GR, Hummel DD, Anderson OD, Anderson JA, Gonzalez-Hernandez JL, Peng JH, Lapitan N, Qi LL, Echalier B, Gill BS, Hossain KG, Kalavacharla V, Kianian SF, Sandhu D, Erayman M, Gill KS, McGuire PE, Qualset CO, Sorrells ME. 2004. Group 3 chromosome bin maps of wheat and their relationship to rice chromosome 1. Genetics 168:639-650.
- Nachman MW. 2001. Single nucleotide polymorphisms and recombination rate in humans. Trends Genet. 17:481-485.
- Nachman MW. 1997. Patterns of DNA Variability at X-Linked Loci in Mus domesticus. Genetics 147:1303-1316.
- Nachman MW, Bauer VL, Crowell SL, Aquadro CF. 1998. DNA Variability and Recombination Rates at X-Linked Loci in Humans. Genetics 150:1133-1141.
- Nachman MW, Churchill GA. 1996. Heterogeneity in Rates of Recombination Across the Mouse Genome. Genetics 142:537-548.
- Nagylaki T. 1983. Evolution of a Finite Population under Gene Conversion. Proceedings of the National Academy of Science of United States of America 80:6278-6281.
- Nasrallah ME, Yogeeswaran K, Snyder S, Nasrallah JB. 2000. Arabidopsis species hybrids in the study of species differences and evolution of amphiploidy in plants. Plant Physiol 124:1605-1614.
- Nesbitt M, Samuel D. 1998. Wheat domestication: Archaeobotanical evidence. Science 279:1433-1433.

- Nesbitt M, Samuel D. 1996. From staple crop extinction? The archaeology and history of the hulled wheats. In: Padulosi S Hammer K, Heller J, ed. Hulled Wheat: Promoting the Conservation and Use of Underutilized and Neglected Crops. International Plant Genetic Resources Institute, Rome. p. 41-100.
- Nordborg M. 2000. Linkage disequilibrium, gene trees and selfing : an ancestral recombination graph with partial self-fertilization. Genetics 154:923-929.
- Nordborg M. 2003. Coalescent theory. In: Balding D Bishop M, Cannings C, ed. Handbook of Statistical Genetics. Wiley, Chichester, UK. p. 602-635.
- Nybom H. 2004. Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. Mol. Ecol. 13:1143-1155.
- Ohta T. 1992. The Nearly Neutral Theory of Molecular Evolution. Annu. Rev. Ecol. Syst. 23:263-286.
- Ohta T, Kimura M. 1969. Linkage Disequilibrium at Steady State Determined by Random Genetic Drift and Recurrent Mutation. Genetics 63:229-238.
- Ometto L, Glinka S, De Lorenzo D, Stephan W. 2005. Inferring the effects of demography and selection on Drosophila melanogaster populations from a chromosome-wide scan of DNA variation. Mol. Biol. Evol. 22:2119-2130.
- Ozkan H, Brandolini A, Pozzi C, Effgen S, Wunder J, Salamini F. 2005. A reconsideration of the domestication geography of tetraploid wheats. TAG Theoretical and Applied Genetics 110:1052-1060.
- Paland S, Lynch M. 2006. Transitions to asexuality result in excess amino acid substitutions. Science 311:990-992.
- Papa R, Acosta J, Delgado-Salinas A, Gepts P. 2005. A genome-wide analysis of differentiation between wild and domesticated Phaseolus vulgaris from Mesoamerica. Theor. Appl. Genet. 111:1147-1158.
- Perry J, Ashworth A. 1999. Evolutionary rate of a gene affected by chromosomal position. Current Biology 9:987-989.
- Piffanelli P, Ramsay L, Waugh R, Benabdelmouna A, D'Hont A, Hollricher K, Jorgensen JH, Schulze-Lefert P, Panstruga R. 2004. A barley cultivation-associated polymorphism conveys resistance to powdery mildew. Nature 430:887-891.
- Pollak E. 1987. On the Theory of Partially Inbreeding Finite Populations. I. Partial Selfing. Genetics 117:353-360.
- Pollard KS, Salama SR, King B, Kern AD, Dreszer T, Katzman S, Siepel A, Pedersen JS, Bejerano G, Baertsch R, Rosenbloom KR, Kent J, Haussler D. 2006. Forces Shaping the Fastest Evolving Regions in the Human Genome. PLoS Genet. 2:e168.
- Provine WB. 1986. Sewall Wright and evolutionary biology. University of Chicago Press, Chicago.
- Qi L, Echalier B, Friebe B, Gill B. 2003. Molecular characterization of a set of wheat deletion stocks for use in chromosome bin mapping of ESTs. Functional & Integrative Genomics 3:39-55.
- Rafalski A, Morgante M. 2004. Corn and humans: recombination and linkage disequilibrium in two genomes of similar size. Trends Genet. 20:103-111.
- Ravel C, Praud S, Murigneux A, Canaguier A, Sapet F, Samson D, Balfourier F, Dufour P, Chalhoub B, Brunel D, Beckert M, Charmet G. 2006. Single-nucleotide polymorphism frequency in a set of selected lines of bread wheat (Triticum aestivum L.). Genome 49:1131-1139.
- Roselius K, Stephan W, Stadler T. 2005. The Relationship of Nucleotide Polymorphism, Recombination Rate and Selection in Wild Tomato Species. Genetics 171:753-763.

- Ross-Ibarra J, Morrell PL, Gaut BS. 2007. Colloquium Papers: Plant domestication, a unique opportunity to identify the genetic basis of adaptation. Proc. Natl Acad. Sci. USA 104:8641-8648.
- Salamini F, Ozkan H, Brandolini A, Schâfer-Pregl R, Martin W. 2002. genetics and geography of wild cereal domestication in the near east. Nature Rev. Genet. 3:429-441.
- Shewry PR. 2007. Improving the protein content and composition of cereal grain. Journal of Cereal Science 46:239-250.
- Smith BD. 1998. The Emergence of Agriculture. Diane Pub Co.
- Sorrells ME, La Rota M, Bermudez-Kandianis CE, Greene RA, Kantety R, Munkvold JD, Miftahudin, Mahmoud A, Ma XF, Gustafson PJ, Qi LLL, Echalier B, Gill BS, Matthews DE, Lazo GR, Chao SM, Anderson OD, Edwards H, Linkiewicz AM, Dubcovsky J, Akhunov ED, Dvorak J, Zhang DS, Nguyen HT, Peng JH, Lapitan NLV, Gonzalez-Hernandez JL, Anderson JA, Hossain K, Kalavacharla V, Kianian SF, Choi DW, Close TJ, Dilbirligi M, Gill KS, Steber C, Walker-Simmons MK, McGuire PE, Qualset CO. 2003. Comparative DNA sequence analysis of wheat and rice genomes. Genome Res. 13:1818-1827.
- Stajich JE, Hahn MW. 2005. Disentangling the effects of demography and selection in human history. Mol. Biol. Evol. 22:63-73.
- Stephan W, Langley CH. 1998. DNA Polymorphism in Lycopersicon and Crossing-Over per Physical Length. Genetics 150:1585-1593.
- Sueoka N. 1962. On the Genetic Basis of Variation and Heterogeneity of DNA Base Composition. Proc. Natl Acad. Sci. USA 48:582-592.
- Tajima F. 1983. Evolutionary relationship of DNA sequences in finite populations. Genetics 105:437-460.
- Tajima F. 1989a. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585-595.
- Tajima F. 1989b. The Effect of Change in Population Size on DNA Polymorphism. Genetics 123:597-601.
- Tanno K, Willcox G. 2006. How fast was wild wheat domesticated? Science 311:1886-1886.
- Tenaillon MI, Sawkins MC, Anderson LK, Stack SM, Doebley J, Gaut BS. 2002. Patterns of diversity and recombination along chromosome 1 of maize (Zea mays ssp mays L.). Genetics 162:1401-1413.
- Tenaillon MI, U'Ren J, Tenaillon O, Gaut BS. 2004. Selection versus demography: A multilocus investigation of the domestication process in maize. Mol. Biol. Evol. 21:1214-1225.
- Teshima KM, Coop G, Przeworski M. 2006. How reliable are empirical genomic scans for selective sweeps? Genome Res. 16:702-712.
- Thornton K, Andolfatto P. 2006. Approximate Bayesian Inference Reveals Evidence for a Recent, Severe Bottleneck in a Netherlands Population of Drosophila melanogaster. Genetics 172:1607-1619.
- Thornton KR, Jensen JD. 2007. Controlling the False-Positive Rate in Multilocus Genome Scans for Selection. Genetics 175:737-750.
- Thuillet A-C, Bataillon T, Poirier S, Santoni S, David JL. 2005. Estimation of Long-Term Effective Population Sizes Through the History of Durum Wheat Using Microsatellite Data. Genetics 169:1589-1599.
- Tsegaye S. 1996. Estimation of outcrossing rate in landraces of tetraploid wheat (*Triticum turgidum*). Plant Breeding 115:195-197.
- Wakeley J, Hey J. 1997. Estimating ancestral population parameters. Genetics 145:847-855.

- Wang LJ, Roossinck MJ. 2006. Comparative analysis of expressed sequences reveals a conserved pattern of optimal codon usage in plants. Plant Mol. Biol. 61:699-710.
- Watterson GA. 1975. On the number of segregating sites in genetical models without recombinaison. Theor. Popul. Biol. 7:188-193.
- Whitt SR, Wilson LM, Tenaillon MI, Gaut BS, Buckler ES. 2002. Genetic diversity and selection in the maize starch pathway. Proc. Natl Acad. Sci. USA 99:12959-12962.
- Wong GKS, Wang J, Tao L, Tan J, Zhang JG, Passey DA, Yu J. 2002. Compositional gradients in Gramineae genes. Genome Res. 12:851-856.
- Wright S. 1931. Evolution in Mendelian populations. Genetics:97-159.
- Wright SI, Bi IV, Schroeder SG, Yamasaki M, Doebley JF, McMullen MD, Gaut BS. 2005. The Effects of Artificial Selection on the Maize Genome. Science 308:1310-1314.
- Wright SI, Foxe JP, DeRose-Wilson L, Kawabe A, Looseley M, Gaut BS, Charlesworth D. 2006. Testing for Effects of Recombination Rate on Nucleotide Diversity in Natural Populations of Arabidopsis lyrata. Genetics 174:1421-1430.
- Wright SI, Gaut BS. 2005. Molecular population genetics and the search for adaptive evolution in plants. Mol. Biol. Evol. 22:506-519.
- Wright SI, Iorgovan G, Misra S, Mokhtari M. 2007. Neutral Evolution of Synonymous Base Composition in the Brassicaceae. J. Mol. Evol. 64:136-141.
- Wright SI, Lauga B, Charlesworth D. 2002. Rates and Patterns of Molecular Evolution in Inbred and Outbred Arabidopsis. Mol. Biol. Evol. 19:1407-1420.
- Wright SI, Ness RW, Foxe JP, Barrett SCH. 2008. Genomic Consequences of Outcrossing and Selfing in Plants. Int. J. Plant Sci. 169:105-118.
- Yahiaoui N, Brunner S, Keller B. 2006. Rapid generation of new powdery mildew resistance genes after wheat domestication. Plant J. 47:85-98.
- Yamasaki M, Tenaillon MI, Bi IV, Schroeder SG, Sanchez-Villeda H, Doebley JF, Gaut BS, McMullen MD. 2005. A large-scale screen for artificial selection in maize identifies candidate agronomic loci for domestication and crop improvement. Plant Cell 17:2859-2872.
- Yang Z. 2002. Inference of selection from multiple species alignments. Current Opinion in Genetics & Development 12:688-694.
- Yang ZH. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Computer Applications in the Biosciences 13:555-556.
- Yang ZH, Bielawski JP. 2000. Statistical methods for detecting molecular adaptation. Trends in Ecology & Evolution 15:496-503.
- Yu A, Zhao CF, Fan Y, Jang WH, Mungall AJ, Deloukas P, Olsen A, Doggett NA, Ghebranious N, Broman KW, Weber JL. 2001. Comparison of human genetic and sequence-based physical maps. Nature 409:951-953.
- Zamora E. 2005. Etude du régime de reproduction des Triticées diploïdes. Flux de gènes interspécifiques entre le blé dur et Aegilops ovata : barrières à l'hybridation. Mémoire de fin d'études Institut Universitaire de Technologie d'Avignon.
- Zeder MA, Emshwiller E, Smith BD, Bradley DG. 2006. Documenting domestication: the intersection of genetics and archaeology. Trends Genet. 22:139-155.
- Zhu Q, Zheng X, Luo J, Gaut BS, Ge S. 2007. Multilocus Analysis of Nucleotide Variation of Oryza sativa and Its Wild Relatives: Severe Bottleneck during Domestication of Rice. Mol. Biol. Evol. 24:875-888.

Annexe A.

Grinding up Wheat: A Massive Loss of Nucleotide Diversity Since Domestication

Supplementary Material

Supplementary Table S1	
Accessions of Wild and Cultivated Forms of Wheat	

	Individual	Species	Accession Id.	Source	Geographic Region
•	AE1	T. aestivum	erge 748	CRB Clermont	Afghanistan
	AE2	T. aestivum	erge 822	CRB Clermont	China
	AE3	T. aestivum	erge 964	CRB Clermont	France
	AE4	T. aestivum	erge 1110	CRB Clermont	Australia
	AE5	T. aestivum	erge 1192	CRB Clermont	Balkans
	AE6	T. aestivum	erge 1232	CRB Clermont	France
	AE7	T. aestivum	erge 1288	CRB Clermont	Hungary
	AE8	T. aestivum	erge 2135	CRB Clermont	China
	AE9	T. aestivum	erge 2153	CRB Clermont	Russia
	AE10	T. aestivum	erge 2171	CRB Clermont	Nepal
	AE11	T. aestivum	erge 2330	CRB Clermont	Turkey
	AE12	T. aestivum	erge 2353	CRB Clermont	Brazil
	AE13	T. aestivum	erge 2358	CRB Clermont	France
	AE14	T. aestivum	erge 2526	CRB Clermont	France
	AE16	T. aestivum	erge 3358	CRB Clermont	Canada
	AE17	T. aestivum	erge 3366	CRB Clermont	Hungary
	AE19	T. aestivum	erge 3942	CRB Clermont	Finland
	AE21	T. aestivum	erge 4482	CRB Clermont	Israel
	AE22	T. aestivum	erge 4645	CRB Clermont	France
	AE23	T. aestivum	erge 4874	CRB Clermont	Lebanon
	AE24	T. aestivum	erge 4901	CRB Clermont	Portugal
	AE25	T. aestivum	erge 5088	CRB Clermont	Israel
	AE26	T. aestivum	erge 5102	CRB Clermont	Nepal
	AE27	T. aestivum	erge 5116	CRB Clermont	China
	AE28	T. aestivum	erge 5166	CRB Clermont	Nepal
	AE29	T. aestivum	erge 5308	CRB Clermont	India

T. aestivum	erge 5399	CRB Clermont	Japan
T. aestivum	erge 5486	CRB Clermont	Germany
T. aestivum	erge 5748	CRB Clermont	Mexico
T. aestivum	erge 6027	CRB Clermont	France
T. aestivum	erge 6086	CRB Clermont	France
T. aestivum	erge 6529	CRB Clermont	Korea
T. aestivum	erge 8048	CRB Clermont	France
T. aestivum	erge 8058	CRB Clermont	Belgium
T. aestivum	erge 9048	CRB Clermont	Finland
T. aestivum	erge 13310	CRB Clermont	Germany
T. aestivum	erge 13471	CRB Clermont	France
T. aestivum	erge 13476	CRB Clermont	France
T. aestivum	erge 13481	CRB Clermont	France
T. aestivum	erge 13811	CRB Clermont	Mexico
T. aestivum	erge 15658	CRB Clermont	France
T. t. ssp. dicoccum	45383	Icarda	Bulgaria
T. t. ssp. dicoccum	45351	Icarda	Iran
T. t. ssp. dicoccum	45239	Icarda	Italy
T. t. ssp. dicoccum	45354	Icarda	Russia
T. t. ssp. dicoccum	45280	Icarda	Slovakia
T. t. ssp. dicoccum	45309	Icarda	Slovakia
T. t. ssp. dicoccum	PI 352365	UsDA	Germany
T. t. ssp. dicoccum	PI 355484	UsDA	Spain
T. t. ssp. dicoccum	PI 94635	UsDA	Iran
T. t. ssp. dicoccum	PI 415152	UsDA	Israel
T. t. ssp. dicoccum	PI 94648	UsDA	Italy
T. t. ssp. dicoccum	erge 4098	CRB Clermont	India
T. t. ssp. dicoccoides	113302	Icarda	Iran Ilam
T. t. ssp. dicoccoides	45963	Icarda	Jordan Amman
	T. aestivum T. t. ssp. dicoccum T. t. ssp. dicoccum	T. aestivumerge 5399 T. aestivumerge 5748 T. aestivumerge 6027 T. aestivumerge 6027 T. aestivumerge 6027 T. aestivumerge 6529 T. aestivumerge 8048 T. aestivumerge 8058 T. aestivumerge 9048 T. aestivumerge 9048 T. aestivumerge 13471 T. aestivumerge 13471 T. aestivumerge 13476 T. aestivumerge 13476 T. aestivumerge 13481 T. aestivumerge 13811 T. aestivumerge 15658 T. t. ssp. dicoccum 45351 T. t. ssp. dicoccum 45351 T. t. ssp. dicoccum 45309 T. t. ssp. dicoccum 45309 T. t. ssp. dicoccum 45309 T. t. ssp. dicoccumPI 352365 T. t. ssp. dicoccumPI 352365 T. t. ssp. dicoccumPI 355484 T. t. ssp. dicoccumPI 94635 T. t. ssp. dicoccumPI 94648 T. t. ssp. dicoccum<	T. aestivumerge 5399CRB Clermont $T.$ aestivumerge 5486CRB Clermont $T.$ aestivumerge 5748CRB Clermont $T.$ aestivumerge 6027CRB Clermont $T.$ aestivumerge 6086CRB Clermont $T.$ aestivumerge 6529CRB Clermont $T.$ aestivumerge 8048CRB Clermont $T.$ aestivumerge 8058CRB Clermont $T.$ aestivumerge 9048CRB Clermont $T.$ aestivumerge 9048CRB Clermont $T.$ aestivumerge 13310CRB Clermont $T.$ aestivumerge 13471CRB Clermont $T.$ aestivumerge 13476CRB Clermont $T.$ aestivumerge 13476CRB Clermont $T.$ aestivumerge 13476CRB Clermont $T.$ aestivumerge 13471CRB Clermont $T.$ aestivumerge 13481CRB Clermont $T.$ aestivumerge 13481CRB Clermont $T.$ aestivumerge 15658CRB Clermont $T.$ assp. dicoccum45383Icarda $T.$ t. ssp. dicoccum45354Icarda $T.$ t. ssp. dicoccum45329Icarda $T.$ t. ssp. dicoccumPI 352365UsDA $T.$ t. ssp. dicoccumPI 355484UsDA $T.$ t. ssp. dicoccumPI 94635UsDA $T.$ t. ssp. dicoccumPI 94648UsDA $T.$ t. ssp. dicoccumPI 94648UsDA $T.$ t. ssp. dicoccumPI 94648UsDA $T.$ t. ssp. dicoccumPI 94635Icarda

DD23	T. t. ssp. dicoccoides	46391	Icarda	Jordan Zarqa	
DD25	T. t. ssp. dicoccoides	46470	Icarda	Syria As Suwayda	
DD26	T. t. ssp. dicoccoides	46253	Icarda	Turkey Diyarbakir	
DD27	T. t. ssp. dicoccoides	116172	Icarda	Turkey Gaziantep	
DD28	T. t. ssp. dicoccoides	PI 467014	UsDA	Israel	
DD29	T. t. ssp. dicoccoides	PI 428133	UsDA	Lebanon	
DD30	T. t. ssp. dicoccoides	46516	Icarda	Syria As Suwayda	
DD31	T. t. ssp. dicoccoides	PI 487255	UsDA	Syria	
DD48	T. t. ssp. dicoccoides	46499	Icarda	Jordan As Salt	
DD51	T. t. ssp. dicoccoides	111002	Icarda	Jordan Irbid	
DD53	T. t. ssp. dicoccoides	115811	Icarda	Jordan Tafila	
DD56	T. t. ssp. dicoccoides	46056	Icarda	Jordan Tafila	
DD61	T. t. ssp. dicoccoides	46294	Icarda	Palestinian Territories	
DD63	T. t. ssp. dicoccoides	46310	Icarda	Palestinian Territories	
DD64	T. t. ssp. dicoccoides	46518	Icarda	Syria Dar'a	
DD65	T. t. ssp. dicoccoides	46520	Icarda	Syria Dar'a	
DD70	T. t. ssp. dicoccoides	46501	Icarda	Syria As Suwayda	
DD74	T. t. ssp. dicoccoides	116175	Icarda	Turkey Gaziantep	
DD75	T. t. ssp. dicoccoides	116179	Icarda	Turkey Gaziantep	
DD76	T. t. ssp. dicoccoides	116184	Icarda	Turkey Gaziantep	
DD77	T. t. ssp. dicoccoides	46191	Icarda	Turkey Urfa	
DD78	T. t. ssp. dicoccoides	46244	Icarda	Turkey Urfa	
DD81	T. t. ssp. dicoccoides	68263	rennes	Unknown	
DD82	T. t. ssp. dicoccoides	PI 428105	UsDA	Israel	
DD95	T. t. ssp. dicoccoides	117894	Icarda	Syria Aleppo	
DD101	T. t. ssp. dicoccoides	PI 538657	UsDA	Turkey Gaziantep	
DR36	T. t. ssp. durum	senatoreCappelli(21)	Montpellier	Italy	
DR47	T. t. ssp. durum	LRB6R5546/49/4		Bulgaria	
DR49	T. t. ssp. durum	kubanka	Montpellier	Russia	

DR50	T. t. ssp. durum	84866	Icarda	Syria
DR51	T. t. ssp. durum	95920	Icarda	Syria
DR52	T. t. ssp. durum	82697	Icarda	Turkey
DR53	T. t. ssp. durum	82726	Icarda	Turkey
DR54	T. t. ssp. durum	82715	Icarda	Turkey
DR55	T. t. ssp. durum	BRUMAIRE	GEVEs	France
DR56	T. t. ssp. durum	PRIMADUR	GEVEs	France
DR57	T. t. ssp. durum	NEODUR91	GEVEs	France
DR58	T. t. ssp. durum	IXOs9442	GEVEs	France
DR59	T. t. ssp. durum	VILLEMUR	GEVEs	France
DR60	T. t. ssp. durum	ARMET910	GEVEs	France
DR61	T. t. ssp. durum	DURENTAL	GEVEs	France
DR62	T. t. ssp. durum	LLOYD945	GEVEs	France
DR63	T. t. ssp. durum	B6Rtchir		Bulgaria
DR75	T. t. ssp. durum	erge 3173	CRB Clermont	Italy
DR77	T. t. ssp. durum	erge 7657	CRB Clermont	Spain
DR78	T. t. ssp. durum	erge 3509	CRB Clermont	Turkey
TI103	T. timopheevii	PI 560872	UsDA	Turkey
TI104	T. timopheevii	PI 560873	UsDA	Turkey
TI105	T. timopheevii	PI 560874	UsDA	Turkey
TI106	T. timopheevii	PI 560877	UsDA	Turkey
TI32	T. timopheevii	PI 560697	UsDA	Turkey

Supplementary accessions (microsatellite data)

DD43	T. t. ssp. dicoccoides	113301	Icarda	Iran Ilam
DD47	T. t. ssp. dicoccoides	46386	Icarda	Jordan Amman
DD50	T. t. ssp. dicoccoides	46491	Icarda	Jordan Irbid

DD52	T. t. ssp. dicoccoides	115802	Icarda	Jordan Irbid
DD55	T. t. ssp. dicoccoides	46385	Icarda	Jordan Irbid
DD57	T. t. ssp. dicoccoides	46323	Icarda	Jordan As Salt
DD60	T. t. ssp. dicoccoides	46275	Icarda	Palestinian Territories
DD62	T. t. ssp. dicoccoides	46309	Icarda	Palestinian Territories
DD66	T. t. ssp. dicoccoides	119437	Icarda	Syria Hama
DD68	T. t. ssp. dicoccoides	46459	Icarda	Syria As Suwayda
DD69	T. t. ssp. dicoccoides	46470	Icarda	Syria As Suwayda
DD71	T. t. ssp. dicoccoides	46453	Icarda	Syria As Suwayda
DD72	T. t. ssp. dicoccoides	46471	Icarda	Syria Damascus
DD79	T. t. ssp. dicoccoides	116182	Icarda	Turkey Gaziantep
DD83	T. t. ssp. dicoccoides	PI 466987	UsDA	Israel
DD84	T. t. ssp. dicoccoides	PI 467005	UsDA	Israel
DD86	T. t. ssp. dicoccoides	PI 481489	UsDA	Israel
DD87	T. t. ssp. dicoccoides	PI 481499	UsDA	Israel
DD88	T. t. ssp. dicoccoides	110815	Icarda	Lebanon
DD89	T. t. ssp. dicoccoides	46530	Icarda	Lebanon
DD90	T. t. ssp. dicoccoides	PI 352324	UsDA	Lebanon
DD94	T. t. ssp. dicoccoides	46289	Icarda	Syria Dar'a
DD98	T. t. ssp. dicoccoides	PI 487253	UsDA	Syria As Suwayda
DD100	T. t. ssp. dicoccoides	PI 428084	UsDA	Turkey Diyarbakir

11B		91A		AapA		AlperA		Bp3B		Bp2A	
AE1	EF095061	AE1	EF104264	DC10	EF104346	DR63	EF104415	DC12	EF108597	DC10	EF108667
AE2	EF095062	AE2	EF104265	DC11	EF104347	DD51	EF104416	DC11	EF108598	DC11	EF108668
AE3	EF095063	AE3	EF104266	DC12	EF104348	DD53	EF104417	DC13	EF108599	DC12	EF108669
AE4	EF095064	AE4	EF104267	DC13	EF104349	DD61	EF104418	DC14	EF108600	DC13	EF108670
AE6	EF095065	AE5	EF104268	DC14	EF104350	DD63	EF104419	DC15	EF108601	DC14	EF108671
AE7	EF095066	AE6	EF104269	DC15	EF104351	DD64	EF104420	DC17	EF108602	DC15	EF108672
AE8	EF095067	AE7	EF104270	DC18	EF104352	DD70	EF104421	DC19	EF108603	DC16	EF108673
AE9	EF095068	AE8	EF104271	DC19	EF104353	DD74	EF104422	DC20	EF108604	DC17	EF108674
AE10	EF095069	AE9	EF104272	DC20	EF104354	DD75	EF104423	DD21	EF108605	DC18	EF108675
AE11	EF095070	AE10	EF104273	DD21	EF104355	DD76	EF104424	DD22	EF108606	DC19	EF108676
AE12	EF095071	AE11	EF104274	DD23	EF104356	DD77	EF104425	DD23	EF108607	DC20	EF108677
AE13	EF095072	AE12	EF104275	DD25	EF104357	DD78	EF104426	DD25	EF108608	DD21	EF108678
AE14	EF095073	AE13	EF104276	DD26	EF104358	DD81	EF104427	DD26	EF108609	DD23	EF108679
AE16	EF095074	AE14	EF104277	DD27	EF104359	DD82	EF104428	DD27	EF108610	DD25	EF108680
AE17	EF095075	DR75	EF104278	DD28	EF104360	DD95	EF104429	DD28	EF108611	DD26	EF108681
AE19	EF095076	AE16	EF104279	DD29	EF104361	DD101	EF104430	DD29	EF108612	DD27	EF108682
AE22	EF095077	AE17	EF104280	DD30	EF104362	TI103	EF104431	DD30	EF108613	DD28	EF108683
AE23	EF095078	DR78	EF104281	DD31	EF104363	TI104	EF104432	DD31	EF108614	DD29	EF108684
AE24	EF095079	AE19	EF104282	TI32	EF104364	TI105	EF104433	TI32	EF108615	DD30	EF108685
AE25	EF095080	DC70	EF104283	DR47	EF104365	DC10	EF104434	DR36	EF108616	DD31	EF108686
AE27	EF095081	AE21	EF104284	DR49	EF104366	DC11	EF104435	DR47	EF108617	TI32	EF108687
AE28	EF095082	AE22	EF104285	DR50	EF104367	DC12	EF104436	DR49	EF108618	DR36	EF108688
AE29	EF095083	AE23	EF104286	DR52	EF104368	DC13	EF104437	DR50	EF108619	DR47	EF108689
AE30	EF095084	AE24	EF104287	DR53	EF104369	DC14	EF104438	DR51	EF108620	DR49	EF108690
AE32	EF095085	AE25	EF104288	DR54	EF104370	DC15	EF104439	DR52	EF108621	DR50	EF108691
AE33	EF095086	AE26	EF104289	DR55	EF104371	DC16	EF104440	DR53	EF108622	DR51	EF108692
AE34	EF095087	AE27	EF104290	DR56	EF104372	DC17	EF104441	DR54	EF108623	DR52	EF108693
AE35	EF095088	AE28	EF104291	DR58	EF104373	DC18	EF104442	DR55	EF108624	DR53	EF108694
AE38	EF095089	AE29	EF104292	DR59	EF104374	DC19	EF104443	DR56	EF108625	DR54	EF108695
AE39	EF095090	AE30	EF104293	DR60	EF104375	DC20	EF104444	DR58	EF108626	DR55	EF108696
AE40	EF095091	AE31	EF104294	DR61	EF104376	DD21	EF104445	DR59	EF108627	DR56	EF108697
AE41	EF095092	AE32	EF104295	DR62	EF104377	DD22	EF104446	DR60	EF108628	DR57	EF108698
AE42	EF095093	AE33	EF104296	DR63	EF104378	DD23	EF104447	DR61	EF108629	DR58	EF108699
AE43	EF095094	AE34	EF104297	AE1	EF104379	DD25	EF104448	DR62	EF108630	DR59	EF108700
AE44	EF095095	AE35	EF104298	AE11	EF104380	DD26	EF104449	DR63	EF108631	DR60	EF108701

11B		91A		AapA		AlperA		Bp3B		Bp2A		
AE45	EF095096	DR77	EF104299	AE12	EF104381	DD27	EF104450	AE1	EF108632	DR61	EF108702	
AE47	EF095097	AE38	EF104300	AE13	EF104382	DD28	EF104451	AE2	EF108633	DR62	EF108703	
DC10	EF095098	AE39	EF104301	AE14	EF104383	DD29	EF104452	AE3	EF108634	DR63	EF108704	
DC11	EF095099	AE40	EF104302	DR75	EF104384	DD30	EF104453	AE4	EF108635	DD48	EF108705	
DC12	EF095100	AE41	EF104303	AE16	EF104385	DD31	EF104454	AE7	EF108636	DD51	EF108706	
DC13	EF095101	AE42	EF104304	AE17	EF104386	DR36	EF104455	AE8	EF108637	DD53	EF108707	
DC14	EF095102	AE43	EF104305	DR78	EF104387	DR47	EF104456	AE9	EF108638	DD56	EF108708	
DC15	EF095103	AE44	EF104306	AE19	EF104388	DR49	EF104457	AE14	EF108639	DD61	EF108709	
DC16	EF095104	AE45	EF104307	AE2	EF104389	DR50	EF104458	DR75	EF108640	DD63	EF108710	
DC17	EF095105	AE47	EF104308	DC70	EF104390	DR51	EF104459	AE16	EF108641	DD64	EF108711	
DC18	EF095106	DC10	EF104309	AE22	EF104391	DR52	EF104460	AE17	EF108642	DD65	EF108712	
DC19	EF095107	DC11	EF104310	AE23	EF104392	DR53	EF104461	DR78	EF108643	DD70	EF108713	
DC20	EF095108	DC12	EF104311	AE24	EF104393	DR54	EF104462	DC70	EF108644	DD74	EF108714	
DC70	EF095109	DC13	EF104312	AE25	EF104394	DR55	EF104463	AE23	EF108645	DD75	EF108715	
DD21	EF095110	DC14	EF104313	AE27	EF104395	DR56	EF104464	AE24	EF108646	DD76	EF108716	
DD22	EF095111	DC15	EF104314	AE28	EF104396	DR57	EF104465	AE25	EF108647	DD77	EF108717	
DD23	EF095112	DC16	EF104315	AE3	EF104397	DR58	EF104466	AE27	EF108648	DD78	EF108718	
DD25	EF095113	DC17	EF104316	AE30	EF104398	DR59	EF104467	AE28	EF108649	DD81	EF108719	
DD26	EF095114	DC18	EF104317	AE32	EF104399	DR60	EF104468	AE29	EF108650	DD82	EF108720	
DD27	EF095115	DC19	EF104318	AE33	EF104400	DR61	EF104469	AE30	EF108651	DD95	EF108721	
DD28	EF095116	DC20	EF104319	AE34	EF104401	DR62	EF104470	AE31	EF108652	DD101	EF108722	
DD29	EF095117	DD21	EF104320	AE35	EF104402	AE1	EF104471	AE32	EF108653	TI103	EF108723	
DD30	EF095118	DD22	EF104321	DR77	EF104403	AE2	EF104472	AE33	EF108654	TI104	EF108724	
DD31	EF095119	DD23	EF104322	AE38	EF104404	AE3	EF104473	AE34	EF108655	TI105	EF108725	
DR36	EF095120	DD25	EF104323	AE39	EF104405	AE4	EF104474	AE35	EF108656	TI106	EF108726	
DR47	EF095121	DD26	EF104324	AE4	EF104406	AE5	EF104475	DR77	EF108657	AE1	EF108727	
DR49	EF095122	DD27	EF104325	AE42	EF104407	AE8	EF104476	AE38	EF108658	AE3	EF108728	
DR50	EF095123	DD28	EF104326	AE43	EF104408	AE9	EF104477	AE39	EF108659	AE4	EF108729	
DR51	EF095124	DD29	EF104327	AE44	EF104409	AE10	EF104478	AE40	EF108660	AE6	EF108730	
DR52	EF095125	DD30	EF104328	AE47	EF104410	AE11	EF104479	AE41	EF108661	AE7	EF108731	
DR53	EF095126	DD31	EF104329	AE6	EF104411	AE13	EF104480	AE42	EF108662	AE8	EF108732	
DR54	EF095127	DR36	EF104330	AE7	EF104412	DR75	EF104481	AE43	EF108663	AE9	EF108733	
DR55	EF095128	DR47	EF104331	AE8	EF104413	AE16	EF104482	AE44	EF108664	AE10	EF108734	
DR56	EF095129	DR49	EF104332	AE9	EF104414	DR78	EF104483	AE45	EF108665	AE11	EF108735	
DR57	EF095130	DR50	EF104333			AE19	EF104484	AE47	EF108666	AE12	EF108736	

11B		91A		AlperA		Bp2A		Bp5A		GdhA		
DR58	EF095131	DR5	1 EF104334	DC70	EF104485	AE13	EF108737	AE28	EF108834	AE32	EF108963	
DR59	EF095132	DR5	2 EF104335	AE21	EF104486	AE14	EF108738	AE29	EF108835	AE33	EF108964	
DR60	EF095133	DR5	3 EF104336	AE23	EF104487	DR75	EF108739	AE30	EF108836	AE34	EF108965	
DR61	EF095134	DR5	4 EF104337	AE24	EF104488	AE16	EF108740	AE31	EF108837	AE35	EF108966	
DR62	EF095135	DR5	5 EF104338	AE25	EF104489	AE17	EF108741	AE32	EF108838	DR77	EF108967	
DR63	EF095136	DR5	6 EF104339	AE26	EF104490	DR78	EF108742	AE33	EF108839	AE38	EF108968	
DR75	EF095137	DR5	7 EF104340	AE27	EF104491	AE19	EF108743	AE34	EF108840	AE39	EF108969	
DR77	EF095138	DR5	9 EF104341	AE29	EF104492	DC70	EF108744	AE35	EF108841	AE40	EF108970	
DR78	EF095139	DR6	0 EF104342	AE30	EF104493	AE21	EF108745	AE38	EF108842	AE41	EF108971	
		DR6	1 EF104343	AE17	EF104494	AE22	EF108746	AE39	EF108843	AE42	EF108972	
		DR6	2 EF104344	AE22	EF104495	AE23	EF108747	AE40	EF108844	AE43	EF108973	
		DR6	3 EF104345	AE31	EF104496	AE24	EF108748	AE41	EF108845	AE44	EF108974	
				AE39	EF104497	AE25	EF108749	AE42	EF108846	AE45	EF108975	
				AE40	EF104498	AE27	EF108750	AE43	EF108847	AE47	EF108976	
				AE32	EF104499	AE28	EF108751	AE44	EF108848	DD48	EF108977	
				AE33	EF104500	AE30	EF108752	AE45	EF108849	DD51	EF108978	
				AE34	EF104501	AE31	EF108753	AE47	EF108850	DD53	EF108979	
				AE35	EF104502	AE32	EF108754	DD48	EF108851	DD56	EF108980	
					EF104503	AE33	EF108755	DD51	EF108852	DD61	EF108981	
				AE38	EF104504	AE34	EF108756		EF108853	DD63	EF108982	
				AE41	EF104505	AE35	EF108757		EF108854	DD64	EF108983	
					EF104506		EF 1007 30		EF 100000		EF100904	
					EF104507	AE39	EF100709		EF100000		EF100900	
					EF104506		EF100700	DD75	EF100007	DD74	EF100900	
				AL47	EF 104509		EF108762		EF108850	78	EF100907	
							EF108763	0077	EF108860		EF108080	
						AL4/	LI 100705	0070	EF108861	0001	EF108909	
								2800	EF108862		EF108001	
								DD95	EF108863	DD35	EF108997	
								DD101	EF108864	TI103	EF108993	
								00101	21 100004	TI104	EF108994	
										TI105	EF108995	
										TI106	EF108996	
						4						
Bp5A		ChsA		GdhA		Gsp1A		Gsp1B		HgA		
-------	----------	------	----------	------	----------	-------	----------	-------	----------	------	----------	
TI103	EF108764	DC11	EF108865	DC10	EF108893	DC10	EF109241	DC10	EF109332	DC14	EF108997	
TI104	EF108765	DC12	EF108866	DC11	EF108894	DC11	EF109242	DC11	EF109333	DC15	EF108998	
TI105	EF108766	DC13	EF108867	DC12	EF108895	DC12	EF109243	DC12	EF109334	DC16	EF108999	
TI106	EF108767	DC14	EF108868	DC13	EF108896	DC13	EF109244	DC13	EF109335	DC17	EF109000	
TI32	EF108768	DC15	EF108869	DC14	EF108897	DC14	EF109245	DC14	EF109336	DC19	EF109001	
DC10	EF108769	DC19	EF108870	DC15	EF108898	DC15	EF109246	DC15	EF109337	DD21	EF109002	
DD64	EF108770	DD21	EF108871	DC16	EF108899	DC16	EF109247	DC16	EF109338	DD23	EF109003	
DD28	EF108771	DD23	EF108872	DC17	EF108900	DC17	EF109248	DC17	EF109339	DD26	EF109004	
DD26	EF108772	DD25	EF108873	DC18	EF108901	DC18	EF109249	DC18	EF109340	DD27	EF109005	
DD65	EF108773	DD26	EF108874	DC19	EF108902	DC19	EF109250	DC19	EF109341	DD28	EF109006	
DC11	EF108774	DD27	EF108875	DC20	EF108903	DC20	EF109251	DC20	EF109342	DD29	EF109007	
DC12	EF108775	DD28	EF108876	DD21	EF108904	DD21	EF109252	DD21	EF109343	DD30	EF109008	
DC13	EF108776	TI32	EF108877	DD22	EF108905	DD22	EF109253	DD22	EF109344	DD31	EF109009	
DC14	EF108777	DR36	EF108878	DD23	EF108906	DD23	EF109254	DD23	EF109345	TI32	EF109010	
DC15	EF108778	DR47	EF108879	DD25	EF108907	DD25	EF109255	DD25	EF109346	DR36	EF109011	
DC16	EF108779	DR49	EF108880	DD26	EF108908	DD26	EF109256	DD26	EF109347	DR47	EF109012	
DC17	EF108780	DR50	EF108881	DD27	EF108909	DD27	EF109257	DD27	EF109348	DR49	EF109013	
DC18	EF108781	DR52	EF108882	DD28	EF108910	DD28	EF109258	DD28	EF109349	DR50	EF109014	
DC19	EF108782	DR53	EF108883	DD29	EF108911	DD29	EF109259	DD29	EF109350	DR51	EF109015	
DC20	EF108783	DR55	EF108884	DD30	EF108912	DD30	EF109260	DD30	EF109351	DR52	EF109016	
DD21	EF108784	DR56	EF108885	DD31	EF108913	DD31	EF109261	DD31	EF109352	DR53	EF109017	
DD22	EF108785	DR57	EF108886	TI32	EF108914	DR36	EF109262	TI32	EF109353	DR54	EF109018	
DD23	EF108786	DR58	EF108887	DR36	EF108915	DR47	EF109263	DR49	EF109354	DR55	EF109019	
DD25	EF108787	DR59	EF108888	DR47	EF108916	DR49	EF109264	DR50	EF109355	DR56	EF109020	
DD27	EF108788	DR60	EF108889	DR49	EF108917	DR50	EF109265	DR51	EF109356	DR57	EF109021	
DD29	EF108789	DR61	EF108890	DR50	EF108918	DR51	EF109266	DR52	EF109357	DR58	EF109022	
DD30	EF108790	DR62	EF108891	DR51	EF108919	DR52	EF109267	DR53	EF109358	DR59	EF109023	
DD31	EF108791	DR63	EF108892	DR52	EF108920	DR54	EF109268	DR54	EF109359	DR60	EF109024	
DR36	EF108792			DR53	EF108921	DR55	EF109269	DR55	EF109360	DR62	EF109025	
DR47	EF108793			DR54	EF108922	DR56	EF109270	DR56	EF109361	AE4	EF109026	
DR49	EF108794			DR55	EF108923	DR57	EF109271	DR57	EF109362	AE5	EF109027	
DR50	EF108795			DR56	EF108924	DR58	EF109272	DR59	EF109363	AE6	EF109028	
DR52	EF108796			DR57	EF108925	DR59	EF109273	DR60	EF109364	AE12	EF109029	
DR53	EF108797			DR58	EF108926	DR60	EF109274	DR61	EF109365	AE14	EF109030	
DR54	EF108798			DR59	EF108927	DR61	EF109275	DR62	EF109366	DR78	EF109031	

Bp5A		GdhA		Gsp1A		Gsp1B		HgA		HiplA	
DR55	EF108799	DR60	EF108928	DR62	EF109276	DR63	EF109367	DC70	EF109032	DR62	EF109468
DR56	EF108800	DR61	EF108929	DR63	EF109277	AE1	EF109368	AE21	EF109033	DR78	EF109469
DR57	EF108801	DR62	EF108930	AE1	EF109278	AE2	EF109369	AE22	EF109034	DR63	EF109470
DR58	EF108802	DR63	EF108931	AE4	EF109279	AE3	EF109370	AE23	EF109035	DR77	EF109471
DR59	EF108803	AE1	EF108932	AE6	EF109280	AE4	EF109371	AE24	EF109036	DR47	EF109472
DR60	EF108804	AE2	EF108933	AE9	EF109281	AE5	EF109372	AE25	EF109037	DR58	EF109473
DR61	EF108805	AE3	EF108934	AE11	EF109282	AE6	EF109373	AE26	EF109038	AE2	EF109474
DR62	EF108806	AE4	EF108935	AE12	EF109283	AE7	EF109374	AE27	EF109039	AE3	EF109475
DR63	EF108807	AE5	EF108936	AE13	EF109284	AE8	EF109375	AE28	EF109040	AE4	EF109476
AE2	EF108808	AE6	EF108937	AE14	EF109285	AE9	EF109376	AE29	EF109041	AE5	EF109477
AE3	EF108809	AE7	EF108938	DR75	EF109286	AE10	EF109377	AE30	EF109042	AE6	EF109478
AE4	EF108810	AE8	EF108939	AE16	EF109287	AE11	EF109378	AE34	EF109043	AE8	EF109479
AE5	EF108811	AE9	EF108940	AE17	EF109288	AE13	EF109379	DR77	EF109044	AE11	EF109480
AE6	EF108812	AE10	EF108941	DR78	EF109289	AE14	EF109380	AE38	EF109045	AE16	EF109481
AE7	EF108813	AE11	EF108942	AE21	EF109290	DR75	EF109381	AE39	EF109046	AE22	EF109482
AE8	EF108814	AE12	EF108943	AE22	EF109291	AE16	EF109382	AE40	EF109047	AE23	EF109483
AE9	EF108815	AE13	EF108944	AE23	EF109292	AE17	EF109383	AE43	EF109048	AE27	EF109484
AE10	EF108816	AE14	EF108945	AE24	EF109293	DR78	EF109384	DD101	EF109049	AE29	EF109485
AE11	EF108817	DR75	EF108946	AE25	EF109294	AE19	EF109385	TI103	EF109050	AE35	EF109486
AE12	EF108818	AE16	EF108947	AE26	EF109295	DC70	EF109386	TI104	EF109051	AE39	EF109487
AE13	EF108819	AE17	EF108948	AE27	EF109296	AE21	EF109387	TI105	EF109052	AE7	EF109488
AE14	EF108820	DR78	EF108949	AE28	EF109297	AE22	EF109388	TI106	EF109053	AE10	EF109489
DR75	EF108821	AE19	EF108950	AE29	EF109298	AE23	EF109389	DD48	EF109054	AE12	EF109490
AE16	EF108822	DC70	EF108951	AE31	EF109299	AE24	EF109390	DD51	EF109055	AE13	EF109491
AE17	EF108823	AE21	EF108952	AE32	EF109300	AE25	EF109391	DD53	EF109056	AE17	EF109492
DR78	EF108824	AE22	EF108953	AE33	EF109301	AE26	EF109392	DD61	EF109057	AE24	EF109493
AE19	EF108825	AE23	EF108954	AE34	EF109302	AE27	EF109393	DD63	EF109058	AE32	EF109494
DC70	EF108826	AE24	EF108955	AE35	EF109303	AE28	EF109394	DD64	EF109059	AE34	EF109495
AE21	EF108827	AE25	EF108956	DR77	EF109304	AE29	EF109395	DD76	EF109060	AE38	EF109496
AE22	EF108828	AE26	EF108957	AE39	EF109305	AE30	EF109396	DD82	EF109061	AE43	EF109497
AE23	EF108829	AE27	EF108958	AE42	EF109306	AE31	EF109397			DD51	EF109498
AE24	EF108830	AE28	EF108959	AE43	EF109307	AE32	EF109398			DD53	EF109499
AE25	EF108831	AE29	EF108960	AE44	EF109308	AE33	EF109399			DD56	EF109500
AE26	EF108832	AE30	EF108961	AE45	EF109309	AE34	EF109400			DD61	EF109501
AE27	EF108833	AE31	EF108962	AE47	EF109310	AE35	EF109401			DD63	EF109502

Gsp1A		Gsp1B		HiplA		MdhA		Mdh4B		Mp7A	
AE7	EF109311	DR77	EF109402	DD64	EF109503	AE38	EF109132	AE41	EF109231	AE13	EF109589
DD48	EF109312	AE38	EF109403	DD65	EF109504	AE39	EF109133	AE42	EF109232	AE14	EF109590
DD51	EF109313	AE39	EF109404	DD70	EF109505	AE41	EF109134	AE43	EF109233	DR75	EF109591
DD53	EF109314	AE40	EF109405	DD74	EF109506	AE42	EF109135	AE44	EF109234	AE16	EF109592
DD61	EF109315	AE41	EF109406	DD75	EF109507	AE43	EF109136	AE45	EF109235	AE17	EF109593
DD63	EF109316	AE42	EF109407	DD76	EF109508	AE44	EF109137	AE47	EF109236	DR78	EF109594
DD64	EF109317	AE43	EF109408	DD77	EF109509	AE45	EF109138	AE6	EF109237	AE19	EF109595
DD65	EF109318	AE44	EF109409	DD78	EF109510	AE47	EF109139	AE7	EF109238	DC70	EF109596
DD70	EF109319	AE45	EF109410	DD81	EF109511	DD48	EF109140	AE8	EF109239	AE21	EF109597
DD74	EF109320	AE47	EF109411	DD82	EF109512	DD51	EF109141	AE9	EF109240	AE22	EF109598
DD76	EF109321	DD101	EF109412	DD95	EF109513	DD53	EF109142			AE23	EF109599
DD77	EF109322	TI103	EF109413	DD101	EF109514	DD56	EF109143			AE24	EF109600
DD78	EF109323	TI105	EF109414	TI103	EF109515	DD61	EF109144			AE25	EF109601
DD81	EF109324	TI106	EF109415	TI104	EF109516	DD63	EF109145			AE26	EF109602
DD101	EF109325	DD48	EF109416	TI105	EF109517	DD64	EF109146			AE27	EF109603
TI32	EF109326	DD51	EF109417	TI106	EF109518	DD65	EF109147			AE28	EF109604
TI103	EF109327	DD53	EF109418			DD70	EF109148			AE29	EF109605
TI104	EF109328	DD56	EF109419			DD74	EF109149			AE30	EF109606
TI105	EF109329	DD61	EF109420			DD76	EF109150			AE31	EF109607
TI106	EF109330	DD63	EF109421			DD77	EF109151			AE32	EF109608
DD75	EF109331	DD64	EF109422			DD78	EF109152			AE33	EF109609
		DD65	EF109423			DD81	EF109153			AE34	EF109610
		DD70	EF109424			DD82	EF109154			AE35	EF109611
		DD74	EF109425			DD95	EF109155			DR77	EF109612
		DD75	EF109426			DD101	EF109156			AE38	EF109613
		DD76	EF109427			TI103	EF109157			AE39	EF109614
		DD77	EF109428			TI104	EF109158			AE40	EF109615
		DD78	EF109429			TI105	EF109159			AE41	EF109616
		DD81	EF109430			TI106	EF109160			AE42	EF109617
		DD82	EF109431							AE43	EF109618
		DD95	EF109432							AE44	EF109619
										AE45	EF109620
										AE47	EF109621

HiplA		MdhA		Mdh4B		Mp7A		MybA		MybB	
DC10	EF109433	DC10	EF109062	DC10	EF109161	DC10	EF109519	DC10	EF114792	DC11	EF114875
DC11	EF109434	DC11	EF109063	DC11	EF109162	DC11	EF109520	DC11	EF114793	DC12	EF114876
DC13	EF109435	DC12	EF109064	DC12	EF109163	DC12	EF109521	DC12	EF114794	DC13	EF114877
DC14	EF109436	DC13	EF109065	DC13	EF109164	DC13	EF109522	DC13	EF114795	DC14	EF114878
DC15	EF109437	DC14	EF109066	DC14	EF109165	DC14	EF109523	DC14	EF114796	DC15	EF114879
DC16	EF109438	DC15	EF109067	DC15	EF109166	DC15	EF109524	DC15	EF114797	DC16	EF114880
DC17	EF109439	DC16	EF109068	DC16	EF109167	DC16	EF109525	DC16	EF114798	DC17	EF114881
DC18	EF109440	DC17	EF109069	DC17	EF109168	DC17	EF109526	DC17	EF114799	DC18	EF114882
DC19	EF109441	DC19	EF109070	DC18	EF109169	DC18	EF109527	DC18	EF114800	DC19	EF114883
DC20	EF109442	DC20	EF109071	DC19	EF109170	DC19	EF109528	DC19	EF114801	DC20	EF114884
DC70	EF109443	DD21	EF109072	DC20	EF109171	DC20	EF109529	DC20	EF114802	DD21	EF114885
DC12	EF109444	DD22	EF109073	DD21	EF109172	DD21	EF109530	DD21	EF114803	DD22	EF114886
DD21	EF109445	DD23	EF109074	DD22	EF109173	DD22	EF109531	DD22	EF114804	DD23	EF114887
DD22	EF109446	DD25	EF109075	DD23	EF109174	DD23	EF109532	DD23	EF114805	DD25	EF114888
DD23	EF109447	DD26	EF109076	DD25	EF109175	DD25	EF109533	DD25	EF114806	DD26	EF114889
DD26	EF109448	DD27	EF109077	DD27	EF109176	DD26	EF109534	DD26	EF114807	DD27	EF114890
DD27	EF109449	DD28	EF109078	DD29	EF109177	DD27	EF109535	DD27	EF114808	DD28	EF114891
DD28	EF109450	DD29	EF109079	DD30	EF109178	DD28	EF109536	DD28	EF114809	DD29	EF114892
DD29	EF109451	DD30	EF109080	DD31	EF109179	DD29	EF109537	DD29	EF114810	DD30	EF114893
DD30	EF109452	DD31	EF109081	DR36	EF109180	DD30	EF109538	DD30	EF114811	DD31	EF114894
DD31	EF109453	TI32	EF109082	DR47	EF109181	DD31	EF109539	DD31	EF114812	DR36	EF114895
TI32	EF109454	DR36	EF109083	DR49	EF109182	DD48	EF109540	TI32	EF114813	DR47	EF114896
DR36	EF109455	DR47	EF109084	DR50	EF109183	DD51	EF109541	DR36	EF114814	DR49	EF114897
DR49	EF109456	DR49	EF109085	DR51	EF109184	DD53	EF109542	DR47	EF114815	DR50	EF114898
DR50	EF109457	DR50	EF109086	DR52	EF109185	DD56	EF109543	DR49	EF114816	DR51	EF114899
DR51	EF109458	DR51	EF109087	DR53	EF109186	DD63	EF109544	DR50	EF114817	DR52	EF114900
DR52	EF109459	DR52	EF109088	DR54	EF109187	DD64	EF109545	DR51	EF114818	DR53	EF114901
DR53	EF109460	DR53	EF109089	DR55	EF109188	DD65	EF109546	DR52	EF114819	DR54	EF114902
DR54	EF109461	DR54	EF109090	DR56	EF109189	DD70	EF109547	DR53	EF114820	DR55	EF114903
DR55	EF109462	DR55	EF109091	DR57	EF109190	DD74	EF109548	DR54	EF114821	DR56	EF114904
DR56	EF109463	DR56	EF109092	DR58	EF109191	DD75	EF109549	DR55	EF114822	DR57	EF114905
DR57	EF109464	DR57	EF109093	DR59	EF109192	DD76	EF109550	DR56	EF114823	DR58	EF114906
DR59	EF109465	DR58	EF109094	DR60	EF109193	DD77	EF109551	DR57	EF114824	DR59	EF114907
DR60	EF109466	DR59	EF109095	DR61	EF109194	DD78	EF109552	DR58	EF114825	DR60	EF114908
DR61	EF109467	DR60	EF109096	DR62	EF109195	DD81	EF109553	DR59	EF114826	DR61	EF114909

MdhA		Mdh4B		Mp7A		MybA		MybB		NrpA	
DR61	EF109097	DR63	EF109196	DD101	EF109554	DR60	EF114827	DR62	EF114910	AE7	EF114992
DR62	EF109098	AE1	EF109197	TI32	EF109555	DR61	EF114828	DR63	EF114911	AE8	EF114993
DR63	EF109099	AE10	EF109198	TI103	EF109556	DR62	EF114829	AE10	EF114912	AE9	EF114994
AE2	EF109100	AE11	EF109199	TI104	EF109557	DR63	EF114830	AE11	EF114913	AE11	EF114995
AE3	EF109101	AE12	EF109200	TI105	EF109558	AE10	EF114831	AE12	EF114914	AE12	EF114996
AE4	EF109102	AE13	EF109201	TI106	EF109559	AE11	EF114832	AE13	EF114915	AE16	EF114997
AE5	EF109103	AE14	EF109202	DR36	EF109560	AE12	EF114833	AE14	EF114916	DR78	EF114998
AE6	EF109104	DR75	EF109203	DR47	EF109561	AE13	EF114834	DR75	EF114917	DC70	EF114999
AE7	EF109105	AE16	EF109204	DR49	EF109562	AE14	EF114835	AE16	EF114918	AE29	EF115000
AE9	EF109106	AE17	EF109205	DR50	EF109563	DR75	EF114836	AE17	EF114919	AE34	EF115001
AE11	EF109107	DR78	EF109206	DR51	EF109564	AE16	EF114837	DR78	EF114920	DR77	EF115002
AE12	EF109108	AE19	EF109207	DR52	EF109565	AE17	EF114838	AE19	EF114921	AE38	EF115003
AE13	EF109109	AE2	EF109208	DR53	EF109566	DR78	EF114839	AE1	EF114922	DD51	EF115004
AE14	EF109110	DC70	EF109209	DR54	EF109567	AE19	EF114840	DC70	EF114923	DD61	EF115005
DR75	EF109111	AE21	EF109210	DR55	EF109568	AE1	EF114841	AE21	EF114924	DD63	EF115006
AE16	EF109112	AE22	EF109211	DR56	EF109569	DC70	EF114842	AE22	EF114925	DD70	EF115007
AE17	EF109113	AE23	EF109212	DR57	EF109570	AE21	EF114843	AE23	EF114926	DD76	EF115008
DR78	EF109114	AE24	EF109213	DR58	EF109571	AE22	EF114844	AE24	EF114927	DD77	EF115009
AE19	EF109115	AE25	EF109214	DR59	EF109572	AE23	EF114845	AE25	EF114928	DD78	EF115010
DC70	EF109116	AE26	EF109215	DR60	EF109573	AE24	EF114846	AE26	EF114929	DD81	EF115011
AE21	EF109117	AE27	EF109216	DR61	EF109574	AE25	EF114847	AE27	EF114930	DD101	EF115012
AE22	EF109118	AE28	EF109217	DR62	EF109575	AE26	EF114848	AE28	EF114931		
AE23	EF109119	AE29	EF109218	DR63	EF109576	AE27	EF114849	AE29	EF114932		
AE24	EF109120	AE3	EF109219	AE1	EF109577	AE28	EF114850	AE2	EF114933		
AE25	EF109121	AE30	EF109220	AE2	EF109578	AE29	EF114851	AE30	EF114934		
AE26	EF109122	AE31	EF109221	AE3	EF109579	AE2	EF114852	AE31	EF114935		
AE27	EF109123	AE32	EF109222	AE4	EF109580	AE30	EF114853	AE32	EF114936		
AE28	EF109124	AE33	EF109223	AE5	EF109581	AE31	EF114854	AE33	EF114937		
AE29	EF109125	AE34	EF109224	AE6	EF109582	AE32	EF114855	AE34	EF114938		
AE30	EF109126	AE35	EF109225	AE7	EF109583	AE33	EF114856	AE35	EF114939		
AE31	EF109127	DR77	EF109226	AE8	EF109584	AE34	EF114857	DR77	EF114940		
AE33	EF109128	AE38	EF109227	AE9	EF109585	AE35	EF114858	AE38	EF114941		
AE34	EF109129	AE39	EF109228	AE10	EF109586	DR77	EF114859	AE39	EF114942		
AE35	EF109130	AE4	EF109229	AE11	EF109587	AE38	EF114860	AE3	EF114943		
DR77	EF109131	AE40	EF109230	AE12	EF109588	AE39	EF114861	AE40	EF114944		

	SD
AE3 EF114862 AE41 EF114945 AE47 EF115118 AE10 EF115083 AE1	EF115189
AE40 EF114863 AE42 EF114946 AE11 EF115084 AE1	E11 EF115190
AE41 EF114864 AE43 EF114947 ZdsB AE12 EF115085 AE1	12 EF115191
AE42 EF114865 AE44 EF114948 AE47 EF115224 AE13 EF115086 AE1	13 EF115192
AE43 EF114866 AE45 EF114949 AE14 EF115087 AE1	E14 EF115193
AE44 EF114867 AE47 EF114950 DR75 EF115088 DR	R75 EF115194
AE45 EF114868 AE4 EF114951 AE16 EF115089 AE1	E16 EF115195
AE47 EF114869 AE5 EF114952 AE17 EF115090 AE1	E17 EF115196
AE4 EF114870 AE6 EF114953 DR78 EF115091 DR	R78 EF115197
AE6 EF114871 AE7 EF114954 AE19 EF115092 AE	EF115198
AE7 EF114872 AE8 EF114955 DC70 EF115093 DC	C70 EF115199
AE8 EF114873 AE9 EF114956 AE21 EF115094 AE2	21 EF115200
AE9 EF114874 AE22 EF115095 AE2	EF115201
AE23 EF115096 AE2	23 EF115202
AE24 EF115097 AE2	:24 EF115203
AE25 EF115098 AE2	:25 EF115204
AE26 EF115099 AE2	26 EF115205
AE27 EF115100 AE2	:27 EF115206
AE28 EF115101 AE2	:28 EF115207
AE29 EF115102 AE2	EF115208
AE30 EF115103 AE3	30 EF115209
	31 EF110210
AE32 EF115105 AE3	32 EF110211
AE33 EF115100 AE3	24 EE115212
AE34 EF115107 AE3 AE35 EE115108 AE7	34 EF115215
	.55 EF115214
AF38 FF115110 AF	38 FF115216
AE39 EE115111 AE	39 FF115217
AF40 FF115112 AF4	40 FF115218
AF41 FF115113 AF4	41 EF115219
AE42 EF115114 AE4	42 EF115220
AE43 EF115115 AE4	43 EF115221
AE44 EF115116 AE4	44 EF115222
AE45 EF115117 AE4	45 EF115223

NrpA		PsyA		Zds	В	PsyA		ZdsB	
DC10	EF114957	DC10	EF115013	DC1	0 EF115119	DR60	EF115048	DR61	EF115154
DC12	EF114958	DC11	EF115014	DC1	1 EF115120	DR61	EF115049	DR62	EF115155
DC13	EF114959	DC12	EF115015	DC1	2 EF115121	DR62	EF115050	DR63	EF115156
DC14	EF114960	DC13	EF115016	DC1	3 EF115122	DR63	EF115051	DD22	EF115157
DC15	EF114961	DC14	EF115017	DC1	4 EF115123	DD48	EF115052	DD48	EF115158
DC16	EF114962	DC15	EF115018	DC1	5 EF115124	DD51	EF115053	DD51	EF115159
DC18	EF114963	DC16	EF115019	DC1	6 EF115125	DD53	EF115054	DD53	EF115160
DC20	EF114964	DC17	EF115020	DC1	7 EF115126	DD56	EF115055	DD56	EF115161
DD21	EF114965	DC18	EF115021	DC1	8 EF115127	DD61	EF115056	DD61	EF115162
DD23	EF114966	DC19	EF115022	DC1	9 EF115128	DD63	EF115057	DD63	EF115163
DD25	EF114967	DC20	EF115023	DC2	0 EF115129	DD64	EF115058	DD64	EF115164
DD26	EF114968	DD21	EF115024	DD2	1 EF115130	DD65	EF115059	DD65	EF115165
DD27	EF114969	DD22	EF115025	DD2	3 EF115131	DD70	EF115060	DD70	EF115166
DD28	EF114970	DD23	EF115026	DD2	5 EF115132	DD74	EF115061	DD74	EF115167
DD29	EF114971	DD25	EF115027	DD2	6 EF115133	DD75	EF115062	DD75	EF115168
DD30	EF114972	DD26	EF115028	DD2	7 EF115134	DD76	EF115063	DD76	EF115169
DD31	EF114973	DD27	EF115029	DD2	8 EF115135	DD77	EF115064	DD77	EF115170
DR47	EF114974	DD28	EF115030	DD2	9 EF115136	DD78	EF115065	DD78	EF115171
DR50	EF114975	DD29	EF115031	DD3	0 EF115137	DD81	EF115066	DD81	EF115172
DR51	EF114976	DD30	EF115032	DD3	1 EF115138	DD82	EF115067	DD82	EF115173
DR52	EF114977	DD31	EF115033	TI32	EF115139	DD95	EF115068	DD95	EF115174
DR53	EF114978	TI32	EF115034	DR3	6 EF115140	DD101	EF115069	DD101	EF115175
DR54	EF114979	DR36	EF115035	DR4	7 EF115141	TI103	EF115070	TI103	EF115176
DR55	EF114980	DR47	EF115036	DR4	9 EF115142	TI104	EF115071	TI104	EF115177
DR56	EF114981	DR49	EF115037	DR5	0 EF115143	TI105	EF115072	TI105	EF115178
DR57	EF114982	DR50	EF115038	DR5	1 EF115144	TI106	EF115073	TI106	EF115179
DR58	EF114983	DR51	EF115039	DR5	2 EF115145	AE1	EF115074	AE1	EF115180
DR59	EF114984	DR52	EF115040	DR5	3 EF115146	AE2	EF115075	AE2	EF115181
DR60	EF114985	DR53	EF115041	DR5	4 EF115147	AE3	EF115076	AE3	EF115182
DR61	EF114986	DR54	EF115042	DR5	5 EF115148	AE4	EF115077	AE4	EF115183
DR62	EF114987	DR55	EF115043	DR5	6 EF115149	AE5	EF115078	AE5	EF115184
AE2	EF114988	DR56	EF115044	DR5	7 EF115150	AE6	EF115079	AE6	EF115185
AE3	EF114989	DR57	EF115045	DR5	8 EF115151	AE7	EF115080	AE7	EF115186
AE4	EF114990	DR58	EF115046	DR5	9 EF115152	AE8	EF115081	AE8	EF115187
AE5	EF114991	DR59	EF115047	DR6	0 EF115153	AE9	EF115082	AE9	EF115188

	F-primer	F-primer sequence	R-primer
<i>11B</i>	1.1_G1	GGTGATGGCACTCGTGAGC	1.1_D1
<i>91A</i>	9.1_G1	CATCGTCAGCCTTAACTGCC	9.1_D1
AapA	AAPIN2g4A	CCTGCATTCTAATGAGATAGACA	AAPIN4d2A
AlperA	ALP4F	CTAGCCTTGACACGTGTTGG	ALPxR
Bp2A	2.3BPGIPGMEX4gA	TTCGTAGAGACAATTGAGAAC	2.3BPGIPGMEX6d1A
Вр3В	2.3BPGIPGMEX8gB	GAGCCTATGCTTGACAAGAGT	2.3BPGIPGMEX9dB
Bp5A	2.3BPGIPGMEX6gA	CACACTGGTGATATTGAAGCT	2.3BPGIPGMEX9dA
ChsA	CHS4F	GTAGACGAGGTGGGTGATCG	CHS5R
GdhA	GDHEX1g1A	GGGGGGGATTTCGTCGG	GDHEX3d1A
Gsp1A	GSP38F	ACAGAACAAAGAAAGTTAGTGGAAGAT	GSP39R
Gsp1B	GSP3F	TCACCAGTAATATCCGCTAGTG	GSP67R
HgA	HG22F	AGGGACACACGAGACCTCAC	HG23R
HiplA	HPL17F	TCATCAAAAACAAGGCAGTGA	HPL19R
Mdh4B	1MDHEX6g1	TCCTGACATTGTTGAATGCTCG	1MDHEX7d2
MdhA	3MDHEX6g	CTGACATTGTTGAATGCTCC	3MDHEX7d
Mp7A	MP7F	GAGTGCAACTGCCCAATATG	MP7R
MybA	mybIN2g2A	GGTTATCATGTACAGTATATTACCTTT	mybEX3d1A
	mybEX3g1A	CATGAAGCTCAGGGACTAAA	mybEX4d1
MybB	mybEX1g1	TCCCTCCCTGAAAATTCTCC	mybIN2d3B
	mybIN2g2B	CTACAGATTTTTATTTTTATCTATC	mybIN2d2B
	mybIN2g5Bg1	TTTGTTGATAGTCTAGAGCAGC	mybEX3d1B
	mybEX3g1B	GAGTTCAAGTTCCTCTGTGAGC	mybEX4d2
NrpA	NRP29F	CTTTCAGTGTTATTGTCTTCAGTTCC	NRP31R
PsyA	PSYA1F	CTGATTCGTAACTGACACTTTCCA	PSY4R
ZdsB	ZDSB2F	GAACCAATCAACTCAAGGTAATCAG	ZDSB1R

R-primer sequence	genome	expected size	bufferamp	dNTPamp	primeramp
GCCCATGCTACGGCCTATG	В	700 bp	buffer PCR 1x	100µM dNTP	10pmole primerF 10pmole primerR
CCAAACAAAAGTTCACAGATCA	А	1200 bp	buffer PCR 1x	100µM dNTP	10pmole primerF 10pmole primerR
GCTGTTATGCCTTGAAAAAAA	А	1000 bp	buffer PCR 1x	100µM dNTP	10pmole primerF 10pmole primerR
ACAGTTCAGCCCAACAGTTTGG	А	1500 bp	buffer PCR 1x	200µM dNTP	6.25pmole primerF 6.25pmole primerR
CTTGCAAGCAACAACTGTA	А	1500 bp	buffer PCR 1x	100µM dNTP	10pmole primerF 10pmole primerR
CATATAGCTGAGCAAAAATCG	В	500 bp	buffer PCR 1x	100µM dNTP	10pmole primerF 10pmole primerR
GCGATCATTCCTTGACCA	А	600 bp	buffer PCR 1x	100µM dNTP	10pmole primerF 10pmole primerR
TGGTTGTTTATTTGCTTCTGGTG	А	776 bp	buffer PCR 1x	200µM dNTP	10pmole primerF 10pmole primerR
GCGTTGGTGCCCATATCT	А	1100 bp	buffer PCR 1x	100µM dNTP	10pmole primerF 10pmole primerR
CCATGTATACCAAAACGCAATG	А	1075 bp	buffer PCR 1x	200µM dNTP	10pmole primerF 10pmole primerR
CCTCAACTTTCATCTTCG	В	720 bp	buffer PCR 1x	200µM dNTP	10pmole primerF 10pmole primerR
CGGATGACATTCCGATCTTT	А	989 bp	buffer PCR 1x	200µM dNTP	10pmole primerF 10pmole primerR
CATTGGTGTACAATCCACCG	А	920 bp	buffer PCR 1x	200µM dNTP	10pmole primerF 10pmole primerR
GGGTTTGCTATAATCTGCAAAATGA	В	1500 bp	buffer PCR 1x	100µM dNTP	10pmole primerF 10pmole primerR
GGCCTAACTAACTTGCATTT	А	900 bp	buffer PCR 1x	100µM dNTP	10pmole primerF 10pmole primerR
TCCTGATGAGCCTGAAAGCG	А	1000 bp	buffer PCR 1x	200µM dNTP	10pmole primerF 10pmole primerR
TTCCCAATAGTCCTGACAGAT	А	900 bp	buffer PCR 1x	100µM dNTP	10pmole primerF 10pmole primerR
CAAGGTTCGGCAAGGAATT	А	600 bp	buffer PCR 1x	100µM dNTP	10pmole primerF 10pmole primerR
CAGTTCATCACATAAAAAATATAGTC	В	1400 bp	buffer PCR 1x	100µM dNTP	10pmole primerF 10pmole primerR
ACAACAGAGGTGAACTTTTTT	В	700 bp	buffer PCR 1x	100µM dNTP	10pmole primerF 10pmole primerR
TTCCCAATATTCCTGACAGAG	В	1100 bp	buffer PCR 1x	100µM dNTP	10pmole primerF 10pmole primerR
GTGAGGGATACCAAGAATCTCC	В	600 bp	buffer PCR 1x	100µM dNTP	10pmole primerF 10pmole primerR
TGCGGTGACTCCTTGTCATA	А	1117 bp	buffer PCR 1x	200µM dNTP	10pmole primerF 10pmole primerR
CCTCCTCCATTTATCCGTCA	А	1208 bp	buffer PCR 1x	200µM dNTP	10pmole primerF 10pmole primerR
GGAATTGTTCTGAAGTCCCATAGAC	В	1112 bp	buffer PCR 1x	200µM dNTP	10pmole primerF 10pmole primerR

ADNamp	Taqamp	H2O for total volume	Tmamp
25 ng	1 u	25 μL	65°C to 70°C
25 ng	1 u	25 µL (sol.Q 1x)	60°C to 70°C
25 ng	1 u	25 μL	55°C to 65°C
50 ng	1 u	25 μL	56°C
25 ng	1 u	25 μL	60°C to 70°C
25 ng	1 u	25 μL	55°C to 65°C
25 ng	1 u	25 μL	55°C to 65°C
50 ng	1 u	25 μL	57°C
25 ng	1 u	25 μL	58°C to 68°C
50 ng	1 u	25 μL	56°C
50 ng	1 u	25 μL	56°C
50 ng	1 u	25 μL	57°C
50 ng	1 u	25 μL	57°C
25 ng	1 u	25 µL (sol.Q 1x)	60°C to 70°C
25 ng	1 u	25 μL	60°C to 70°C
50 ng	1 u	25 μL	56°C
25 ng	1 u	25 μL	55°C to 65°C
25 ng	1 u	25 μL	56°C to 66°C
25 ng	1 u	25 μL	55°C to 65°C
25 ng	1 u	25 μL	55°C to 65°C
25 ng	1 u	25 μL	55°C to 65°C
25 ng	1 u	25 μL	55°C to 65°C
50 ng	1 u	25 µL	57°C
50 ng	1 u	25 µL	56°C
50 ng	1 u	25 µL	56°C

amplification cycle

94°C 4min [94°C 1min 70°C 1min 72°C 2min]4x -1°C/cycle [94°C 1min 65°C 1min 72°C 2min]30x 72°C 5min 15°C end 94°C 4min [94°C 1min 70°C 1min 72°C 2min]9x -1°C/cycle [94°C 1min 60°C 1min 72°C 2min]30x 72°C 5min 15°C end 94°C 4min [94°C 1min 65°C 1min 72°C 2min]9x -1°C/cycle [94°C 1min 55°C 1min 72°C 2min]30x 72°C 5min 15°C end 94°C 4min [94°C 45sec 65°C 45sec 72°C 2min30]12x -0.7°C/cvcle [94°C 45sec 56°C 45sec 72°C 2min30]25x 72°C 4min 15°C 94°C 4min [94°C 1min 70°C 1min 72°C 2min]9x -1°C/cycle [94°C 1min 60°C 1min 72°C 2min]30x 72°C 5min 15°C end 94°C 4min [94°C 1min 65°C 1min 72°C 2min]9x -1°C/cycle [94°C 1min 55°C 1min 72°C 2min]30x 72°C 5min 15°C end 94°C 4min [94°C 1min 65°C 1min 72°C 2min]9x -1°C/cycle [94°C 1min 55°C 1min 72°C 2min]30x 72°C 5min 15°C end 94°C 4min [94°C 45sec 66°C 45sec 72°C 1min]12x -0.7°C/cvcle [94°C 45sec 57°C 45sec 72°C 1min]25x 72°C 4min 15°C 94°C 4min [94°C 1min 68°C 1min 72°C 2min]8x [94°C 1min 60°C 1min 72°C 2min]30x 72°C 5min 15°C end 94°C 4min [94°C 45sec 65°C 45sec 72°C 1min]12x -0.7°C/cvcle [94°C 45sec 56°C 45sec 72°C 1min]25x 72°C 4min 15°C 94°C 4min [94°C 45sec 65°C 45sec 72°C 1min]12x -0,7°C/cycle [94°C 45sec 56°C 45sec 72°C 1min]25x 72°C 4min 15°C 94°C 4min [94°C 45sec 66°C 45sec 72°C 1min]12x -0,7°C/cycle [94°C 45sec 57°C 45sec 72°C 1min]25x 72°C 4min 15°C 94°C 4min [94°C 45sec 66°C 45sec 72°C 1min]12x -0.7°C/cycle [94°C 45sec 57°C 45sec 72°C 1min]25x 72°C 4min 15°C 94°C 4min [94°C 1min 70°C 1min 72°C 2min]9x -1°C/cycle [94°C 1min 60°C 1min 72°C 2min]30x 72°C 5min 15°C end 94°C 4min [94°C 1min 70°C 1min 72°C 2min]9x -1°C/cycle [94°C 1min 60°C 1min 72°C 2min]30x 72°C 5min 15°C end 94°C 4min [94°C 45sec 65°C 45sec 72°C 1min30]12x -0,7°C/cycle [94°C 45sec 56°C 45sec 72°C 1min30]25x 72°C 4min 15°C 94°C 4min [94°C 1min 65°C 1min 72°C 2min]9x -1°C/cycle [94°C 1min 55°C 1min 72°C 2min]30x 72°C 5min 15°C end 94°C 4min [94°C 1min 65°C 1min 72°C 2min]9x -1°C/cycle [94°C 1min 55°C 1min 72°C 2min]30x 72°C 5min 15°C end 94°C 4min [94°C 1min 65°C 1min 72°C 2min]9x -1°C/cycle [94°C 1min 55°C 1min 72°C 2min]30x 72°C 5min 15°C end 94°C 4min [94°C 1min 65°C 1min 72°C 2min]9x -1°C/cycle [94°C 1min 55°C 1min 72°C 2min]30x 72°C 5min 15°C end 94°C 4min [94°C 1min 65°C 1min 72°C 2min]9x -1°C/cycle [94°C 1min 55°C 1min 72°C 2min]30x 72°C 5min 15°C end 94°C 4min [94°C 1min 65°C 1min 72°C 2min]9x -1°C/cycle [94°C 1min 55°C 1min 72°C 2min]30x 72°C 5min 15°C end 94°C 4min [94°C 45sec 66°C 45sec 72°C 1min]12x -0,7°C/cycle [94°C 45sec 57°C 45sec 72°C 1min]25x 72°C 4min 15°C 94°C 4min [94°C 45sec 65°C 45sec 72°C 1min30]12x -0,7°C/cycle [94°C 45sec 56°C 45sec 72°C 1min30]25x 72°C 4min 15°C 94°C 4min [94°C 45sec 65°C 45sec 72°C 1min30]12x -0,7°C/cycle [94°C 45sec 56°C 45sec 72°C 1min30]25x 72°C 4min 15°C

PCR program name	bufferseq	mixseq	primerseq	ADN	Tmseq
SPA65/70	buffer 1x	1 µl BigDye Terminator V3.1	10 pmole	25ng	55°C
SPA70	buffer 1x	1 µl BigDye Terminator V3.1	10 pmole	25ng	60°C
SPA65	buffer 1x	1 µl BigDye Terminator V3.1	10 pmole	25ng	55°C
TD56	buffer 1x	2 µl BigDye Terminator V3.1	3,2 pmole	20ng	50°C
SPA70	buffer 1x	1 µl BigDye Terminator V3.1	10 pmole	25ng	60°C
SPA65	buffer 1x	1 µl BigDye Terminator V3.1	10 pmole	25ng	55°C
SPA65	buffer 1x	1 µl BigDye Terminator V3.1	10 pmole	25ng	55°C
TD57	buffer 1x	2 µl BigDye Terminator V1.1	3,2 pmole	10ng	50°C
SPA68	buffer 1x	1 µl BigDye Terminator V3.1	10 pmole	25ng	55°C
TD56	buffer 1x	2 µl BigDye Terminator V1.1	3,2 pmole	15ng	50°C
TD56	buffer 1x	2 µl BigDye Terminator V1.1	3,2 pmole	15ng	50°C
TD57	buffer 1x	2 μl BigDye Terminator V1.1	3,2 pmole	15ng	50°C
TD57	buffer 1x	2 µl BigDye Terminator V1.1	3,2 pmole	15ng	50°C
SPA70	buffer 1x	1 µl BigDye Terminator V3.1	10 pmole	25ng	60°C
SPA70	buffer 1x	1 µl BigDye Terminator V3.1	10 pmole	25ng	60°C
TD56	buffer 1x	2 µl BigDye Terminator V1.1	3,2 pmole	15ng	50°C
SPA65	buffer 1x	1 µl BigDye Terminator V3.1	10 pmole	25ng	55°C
SPA66	buffer 1x	1 µl BigDye Terminator V3.1	10 pmole	25ng	55°C
SPA65	buffer 1x	1 µl BigDye Terminator V3.1	10 pmole	25ng	55°C
SPA65	buffer 1x	1 µl BigDye Terminator V3.1	10 pmole	25ng	55°C
SPA65	buffer 1x	1 µl BigDye Terminator V3.1	10 pmole	25ng	55°C
SPA65	buffer 1x	1 µl BigDye Terminator V3.1	10 pmole	25ng	55°C
TD57	buffer 1x	2 μl BigDye Terminator V1.1	3,2 pmole	15ng	50°C
TD56	buffer 1x	2 µl BigDye Terminator V3.1	3,2 pmole	15ng	50°C
TD56	buffer 1x	2 µl BigDye Terminator V1.1	3,2 pmole	15ng	50°C

sequence cycle
94°C 5min [94°C 10sec 55°C 4min]20x 10°C end
94°C 5min [94°C 10sec 60°C 4min]20x 10°C end
94°C 5min [94°C 10sec 55°C 4min]20x 10°C end
92°C 5min [92°C 30sec 50°C 20sec 60°C 4min30]60x 15°C rampes : 1°C/sec
94°C 5min [94°C 10sec 60°C 4min]20x 10°C end
94°C 5min [94°C 10sec 55°C 4min]20x 10°C end
94°C 5min [94°C 10sec 55°C 4min]20x 10°C end
92°C 5min [92°C 30sec 50°C 20sec 60°C 4min30]60x 15°C rampes : 1°C/sec
94°C 5min [94°C 10sec 55°C 4min]20x 10°C end
92°C 5min [92°C 30sec 50°C 20sec 60°C 4min30]60x 15°C rampes : 1°C/sec
92°C 5min [92°C 30sec 50°C 20sec 60°C 4min30]60x 15°C rampes : 1°C/sec
92°C 5min [92°C 30sec 50°C 20sec 60°C 4min30]60x 15°C rampes : 1°C/sec
92°C 5min [92°C 30sec 50°C 20sec 60°C 4min30]60x 15°C rampes : 1°C/sec
94°C 5min [94°C 10sec 60°C 4min]20x 10°C end
94°C 5min [94°C 10sec 60°C 4min]20x 10°C end
92°C 5min [92°C 30sec 50°C 20sec 60°C 4min30]60x 15°C rampes : 1°C/sec
94°C 5min [94°C 10sec 55°C 4min]20x 10°C end
94°C 5min [94°C 10sec 55°C 4min]20x 10°C end
94°C 5min [94°C 10sec 55°C 4min]20x 10°C end
94°C 5min [94°C 10sec 55°C 4min]20x 10°C end
94°C 5min [94°C 10sec 55°C 4min]20x 10°C end
94°C 5min [94°C 10sec 55°C 4min]20x 10°C end
92°C 5min [92°C 30sec 50°C 20sec 60°C 4min30]60x 15°C rampes : 1°C/sec
92°C 5min [92°C 30sec 50°C 20sec 60°C 4min30]60x 15°C rampes : 1°C/sec
92°C 5min [92°C 30sec 50°C 20sec 60°C 4min30]60x 15°C rampes : 1°C/sec

Distribution of Mean Values of Ln Likelihood for Each Run Simulating Subdivision Of *aestivum* in k Groups.



Figure Representing the Estimated Proportion of Each Individual Originating from Each of the two Groups Defined within *aestivum* using the Structure software for k=2. The group I is represented in grey and the group II in black.



S4 A





Figure Representing the Estimated Proportion of Each Individual Originating from Each of the two Groups Defined within *dicoccoides* using the Structure software for k=2. One group is represented in grey and the other group in black.



S4 B

S4 C

Results of Spatial Structure Analysis in dicoccoides with microsatellite data



r is the correlation coefficient between genetic and geographic distances.

r, U and L values are adjusted by subtracting the Mean permuted r. Bootstrap mean, Ur, Lr are also adjusted by the Mean permuted r.

Upper (U) and lower (L) confidence limits bound the 95% confidence interval about the null hypothesis of No spatial structure for the combined data set as determined by permutation.

S4 D

Likelihood (y-axis, logarithmic scale) profiles as a function of bottleneck intensity, α (x-axis) for each cultivated group ($\alpha = Ne_{dicoccoides} / Ne_{cultivated}$).



Supplementary Table S5 Diversity in Wild and Cultivated Forms of Wheat

_	Wild group						Cultivated groups													
			<i>T. t.</i> ssp.	. dicoccoi	des			<i>T. t.</i> ss	sp. dicoccu	ım			<i>T. t.</i> ss	sp. durun	n			Т. а	estivum	
Locus	n ^a	S ^b	π_{total}^{c}	π_{silent}^{c}	$\mathbf{D_{Taj}}^{d}$	n ^a	S ^D	π_{total}^{c}	π_{silent}^{c}	$\mathbf{D_{Taj}}^{d}$	n ^a	S ^b	π_{total}^{c}	π_{silent}^{c}	$\mathbf{D_{Taj}}^{d}$	n ^a	S ^D	π_{total}^{c}	π_{silent}^{c}	$\mathbf{D_{Taj}}^{d}$
<i>11B</i>	10	10	3.28	2.77	-1.59	12	2	0.48	0.58	-1.45	20	3	1.87	2.24	1.42	37	4	2.19	2.62	1.42
<i>91A</i>	10	7	2.18	3.52	0.44	12	2	0.27	0.46	-1.45	19	7	1.37	2.10	-0.49	40	3	0.12	0.14	-1.71
AapA	9	2	0.60	0.76	-0.58	10	0	0	0	_	17	1	0.30	0.38	0.09	32	1	0.38	0.48	0.85
AlperA	10	7	2.28	3.24	0.33	12	5	0.83	1.18	-1.53	19	4	1.45	2.07	1.33	30	5	1.24	1.54	0.32
					-1.72					-1.75										
Bp2A	27	18	1.67	2.44	0.1>p>0.5	12	4	0.47	1.10	0.1>p>0.5	20	0	0	0	_	33	0	0	0	_
Вр3В	10	1	1.04	1.39	1.3	9	1	0.44	0.59	-1.09	19	0	0	0	_	31	0	0	0	_
Bp5A	26	5	0.83	1.38	-1.7	12	0	0	0	_	19	0	0	0	_	40	1	0.08	0.13	-1.12
ChsA	6	11	11.27	15.91	0.06	6	1	1.22	0	0.86	15	0	0	0	_	0				
GdhA	26	11	2.40	2.94	-0.08	12	3	0.99	1.26	0.77	19	3	0.74	0.94	0.17	41	6	1.53	1.94	0.92
														_						2.1
GspA	25	26	4.20	6.09	-1.58	11	8	1.86	2.80	-1.49	19	0	0	0		30	5	2.38	2.71	p<0.05
GspB	28	7	2.23	3.14	-1.25	11	4	3.23	6.66	0.43	19	2	0.45	0.84	-1.51	40	0	0	0	_
HgA	14	38	11.63	13.6	-1.03	6	2	1.10	1.29	0.31	16	4	0.83	0.79	-1.31	11	0	0	0	_
HiplA	26	6	1.80	3.92	-0.87	12	2	0.54	1.17	-1.45	20	2	0.62	1.33	-0.77	23	4	2.57	4.27	1.21
MdhA	27	7	1.16	1.30	-1.4	11	0	0	0	_	20	0	0	0	_	36	0	0	0	_
Mdh4B	8	6	1.72	1.56	-0.22	12	2	0.65	0.70	1.35	20	0	0	0	_	40	3	0.43	0.47	-0.19
Mp7A	24	4	0.97	1.20	-0.65	12	1	0.19	0.23	-1.14	20	2	0.76	0.93	0.41	40	3	0.63	0.70	-0.48
MybA	10	3	0.61	0.45	-0.66	12	2	0.23	0	0.33	20	0	0	0	_	39	0	0	0	_
															-1.72					2.35
MybB	10	16	1.37	2.33	-0.86	11	14	1.38	2.13	-0.12	20	3	0.09	0.10	p<0.05	38	18	2.18	3.58	p<0.05
NrpA	17	15	2.58	3.55	-1.69	9	7	2.13	4.22	-0.9	16	3	0.39	0.86	-1.7	13	5	2.13	4.25	0.97
PsyA	28	7	1.84	2.53	-1.18	12	2	0.56	0.82	-1.45	22	1	0.29	0.43	-0.64	42	1	0.73	1.08	1.25
ZdsB	28	5	0.82	0.90	-1.19	12	2	0.73	0.80	-0.05	21	1	0.11	0.12	-1.16	41	1	0.05	0.06	-1.12
Average	18	10.1	2.69	3.57	-0.77**	11	3	0.82	1.24	-0.55*	19	1.7	0.44	0.63	-0.45 ns	32	3	0.83	1.20	0.47*

^a Number of sequences sampled.

^b Total number of segregating sites.

^c Average number of pairwise differences calculated for all sites(π_{total}) and silent sites (π_{silent}) expessed in 10⁻³.

^d Tajima's D statistic for all sites and its significance.

Mating system and recombination affect molecular evolution in four *Triticeae* species

Supplementary Material

Table S1 legend

Gene Name

The LOC prefix refers to the TIGR nomenclature. *Crtiso*: carotenoid isomerase, *Eif4e* and *Eifiso4E* are both translation initiation factors, *Gsp-1* is the grain softness protein gene, *Psy2*: phytoene synthase (caroteinoid biosynthesis), *matK*: maturase K, a chloroplastic gene.

Template

Genomic resources on which PCR primers were designed. BES : Coding sequences identified in BAC End Sequences extracted from a BAC library built on the 3B chromosome of wheat (Paux et al, 2006).

Rice: the locus was chosen on the rice genome pseudo-molecule of the chromosome 1.

Chr1 Rice position

Number of base pairs from the start of the short arm of pseudomolecule of chromosome 1 of Rice (TIGR)

Relative Distance to centromere (%)

Distance to the centromere relatively to the total arm length

Wheat marker ID

Genbank EST ID or INRA Clermont-Ferrand BES ID.

For BES primers, physical assignation is realized by assigning any marker that could be developed from the BAC end (INRA Clermont-Ferrand BES ID) itself or from others BACs which were showed to be physically linked to the considered BAC (unpub. data, INRA Clermont Ferrand).

For Rice primers, physical assignation was realized either by using a previously mapped EST unambiguously identical to the gene fragment studied (bold) or by using a mapped EST (html) having a strong identity with a gene in Rice physically very close (30-50 kb) to the gene under study (italic).

Origin

CF assignation realised by INRA Clermont Ferrand, West assignation realized by the Wheat Est deletion mapping project (http://wheat.pw.usda.gov/ggpages/NSF_Wheat_Resources.html)

Deletion lines and deletions bins

A set of hexaploid wheat deletion lines were used to assign the physical position of the locus on the wheat genome. Their breakpoints define a suite of 'bins' along the chromosome, which are given relatively to the arm chromosome length.

In bold locus which physical assignation on the wheat genome were not consistent with the Rice location.

Fragment length

Total length of the amplified segment(s)

Codons

Number of codons finally used in the analysis (gaps, missing data and unreadable nucleotides being removed)

Tigr Gene ID	Template	Chr1 Rice position	Centromere Relative distance (%)	Arm	Group	Wheat marker ID	origin	Deletion line & deletion bin	Fragment length	Codons
LOC_Os01g06470	BES	3,044,234	0.82	3S	т	3B_033_J05_FM1	CF	3BS1-0.33-0.55	1670	109
LOC_Os01g07370	BES	3,476,198	0.80	3S	т	3B_017_K04_FM1	CF	3BS9-0.57-0.75	1205	349
LOC_Os01g07760	BES	3,721,117	0.78	3S	Т	3B_054_N07_FM2	CF	3BS1-0.33-0.55	996	544
LOC_Os01g10450	BES	5,507,324	0.68	3S	С	3B_032_C01_FM1	CF	3BS4-0.55-0.56	1240	257
LOC_Os01g12710	BES	7,021,695	0.59	3S	С	3B_015	CF	C-3BS1-0.33	850	44
LOC_Os01g17390	BES	10,000,561	0.41	3S	С	3B_013	CF	3BL1-0.31-0.38	850	87
LOC_Os01g21160	rice	11,786,626	0.31	3S	С	CD452946	West	C-3AS2-0.23	1980	259
LOC_Os01g22900	rice	12,852,721	0.24	3S	С	BF485348	West	C-3BL2-0.22	2680	436
LOC_Os01g24680	rice	13,869,753	0.18	3S	С	BE446628	West	3BS1-0.33-0.57	3440	541
LOC_Os01g24950	BES	14,042,084	0.17	3S	С	3B_054_A16_RM2	CF	C-3BL2-0.22 ?	996	332
LOC_Os01g27940	rice	15,945,091	0.06	3S	С				831	183
LOC_Os01g31360	rice	17,480,679	0.02	3L	С				740	206
LOC_Os01g33420	rice	18,721,100	0.06	3L	С	BE591684	West		2825	378
LOC_Os01g35040	rice	19,716,678	0.10	3L	С	BE591684	West	3DS	830	79
LOC_Os01g36390	rice	20,518,897	0.13	3L	С	BE500863	West	C-3BL2-0.22	3320	535
LOC_Os01g37560	rice	21,314,607	0.16	3L	С	BE499186	West	3BL7-0.63-1.00	1030	345
LOC_Os01g37960	BES	21,592,497	0.17	3L	С	3B_052_L08_RM1	CF	C-3BL2-0.22	1460	33
LOC_Os01g39000	BES	22,248,876	0.19	3L	С	3B_007_I06_RM1	CF		890	124
LOC_Os01g39310	rice	22,447,167	0.20	3L	С	BQ161938	West		2735	243
LOC_Os01g41190	rice	23,643,156	0.25	3L	С	BQ161200	West	C-3BL2-0.22	995	190
LOC_Os01g42730	BES	24,639,449	0.28	3L	С	3B_150_F24_RM1	CF	C-3BL2-0.22	1700	84
LOC_Os01g43170	BES	24,960,613	0.29	3L	С	3B_111_M24_RM1	CF	C-3BL2-0.22	1260	203
LOC_Os01g44220	rice	25,680,653	0.32	3L	С	BE517931	West	C-3BL2-0.22	1847	189
LOC_Os01g47460	BES	27,450,336	0.39	3L	С	3B_142_D15_FM1	CF	C-3BL2-0.22	2020	167
LOC_Os01g48720	rice	28,265,649	0.42	3L	С	BG607861	West	3BL2-0.22-0.50	978	57
LOC_Os01g49940	rice	29,010,718	0.44	3L	С	BE443053	West	C-3BL2-0.22	1910	129

LOC_Os01g51380	rice	29,863,554	0.48	3L	С	BG262775, BE585797	West	3BL2-0.22-0.50	2060	194
LOC_Os01g51870	BES	30,145,711	0.49	3L	С	3B_017_P23_FM1	CF	3BL2-0.22-0.28	730	210
LOC_Os01g52050	rice	30,254,432	0.49	3L	С	BE406587, BG607570	West	3BL2-0.22-0.50	1730	575
LOC_Os01g53880	BES	31,302,613	0.53	3L	С	3B_086_K21_RM1	CF	3BL7-0.63-1.00/3BL1-0.31-0.38	1000	115
LOC_Os01g55530	rice	32,318,941	0.57	3L	С	BE443568	West	3BL2-0.22-0.50	1071	358
LOC_Os01g56630	rice	32,996,390	0.59	3L	С	BE490651	West	3BL2-0.22-0.50	280	64
LOC_Os01g56670	BES	33,025,872	0.59	3L	С	3B_093_C10_FM1	CF	3BL7-0.63-1.00	1125	101
LOC_Os01g57110	rice	33,326,901	0.60	3L	С	BQ169615, BE497979	West	3BL2-0.22-0.50	3010	317
LOC_Os01g58080	rice	33,920,645	0.63	3L	С	BE499348	West	3BL2-0.22-0.50	870	290
LOC_Os01g59130	rice	34,480,207	0.65	3L	С	CD452849	West		1551	143
LOC_Os01g60660	rice	35,396,100	0.68	3L	С	BF201439, BE443770	West	3BL10-0.50-0.63	2011	98
LOC_Os01g65920	rice	38,601,921	0.80	3L	Т	BE405221	West	3BL7-0.63-1.00	808	109
LOC_Os01g67210	BES	39,327,531	0.83	3L	Т	3B_002	CF	3BS8-0.78-1.00	1320	116
LOC_Os01g67590	BES	39,616,212	0.84	3L	Т	3B_045_F06_RM2	CF		670	36
LOC_Os01g68380	rice	40,057,147	0.85	3L	Т	CD454927	West	3BL7-0.63-1.00	1665	79
LOC_Os01g68770	rice	40,274,921	0.86	3L	Т	BE442674	CF	3BL7-0.63-1.00	1067	150
LOC_Os01g69970	rice	40,789,858	0.88	3L	Т	BE637668, BF485004	West	3BL7-0.63-1.00	1183	34
LOC_Os01g70670	rice	41,242,189	0.90	3L	Т	BE497398	West	3BL7-0.63-1.00	2320	373
LOC_Os01g73790	rice	43,065,613	0.97	3L	Т	BE489841	West	3BL7-0.63-1.00	692	112
LOC_Os01g74650	rice	43,558,783	0.98	3L	Т	BM137927	West	3BL7-0.63-1.00	1603	215
Crtiso		-							1110	169
Eif4E		-							588	165
Eifiso4E		-							690	165
Gsp-1		-		5S					780	165
matK		-		chloro					2422	512
Psy2		-		5					990	154
52 genes									74,624	11,122

Locus	FPrimer Name	F Primer Sequence	RPrimer Name	R Primer Sequence
LOC_Os01g06470	J05-F	ACAAACATAGCAGCTGGGGAGC	J05-1R	ACTGAGGCACATGGCTTCGTAC
LOC_Os01g07370	K04F	AGCTCCACTGTTTCAGCCCTCGA	K04R	AGGCCCATTCATGGTGGTGGGAC
LOC_Os01g07760	N07F	AGTATCAGGTTCTGGCTGCTCCTGG	N07R	CGAGAAAGCTCGTGTTGGGCGTACC
LOC_Os01g10450	C01F	CGGTCACCATCTCCCACAGCAGC	C01R	TCACAAGCCCTCCCCGAAGC
LOC_Os01g12710	L07F	AGGAAAAGCACTTGCTCGGGAG	L07R	ACAACTCGCATTCTTGGGACCA
LOC_Os01g17390	B06F	TGGATGATTCCATCATGTCGTGG	B06R	CACATGATTCCAAGTGGCCAAC
LOC_Os01g21160	2116U1F	TGCGCTGGTTCGTCAGCGAG	2116U1R	GGCTATCATGAGGTACAACTTGACTGCCG
LOC_Os01g21160	2116U2F	AGCAATGGTGAAATCAATGAGCCTGG	2116U3R	ACTGGACTTTCTTGATGCTTGGCACC
LOC_Os01g21160	2116UF	GCAGCTCTGTGAGGTACAGAGCGAC	2116U2R	TGTTCAACAATGGATATTTACTCAGTGCC
LOC_Os01g22900	2290U1F	TGGATGTGCAAACCGGGATTAAGC	2290UR	GCCACAGTAGGGTTGGCCAAGATCC
LOC_Os01g22900	2290U2F	GGCGCATTATTACAGGCAGTGACC	2290U2R	TGATATCTGATTTGGCCGCACGTCG
LOC_Os01g22900	2290UF	ACCCCGAGTGTACGGAGATGCT	2290U1R	TGAATCTCAAGAGGATGCCCGTG
LOC_Os01g24680	2468U1F	AGCTATGAAGAAGCTCTTCAGAGGAAGG	2468U2R	AGAATCTCCCTAGCCTCACCAAGG
LOC_Os01g24680	2468U2F	TCATCACCATCTGCAACCCGCCAG	2468U1R	TCCGCTGTGAAGATCACCAAAGG
LOC_Os01g24680	2468U3F	AGGCTTTGTCTCTTCTCAGTGGTGCCC	2468UR	ACTGCATACCAGTAGCCACAGC
LOC_Os01g24680	2468U4F	CAAATTTGGAATGCCCATGGGTCC	2468U4R	GCGAGGTAGGAGCAAGGCTCG
LOC_Os01g24680	2468UF	AGAGGCACACCAACTGGGTC	2468U3R	TGGGAAGGGCAGTACTTCTCCAAGTCCG
LOC_Os01g24950	A16F	AGTTGGGATGCTACTCGACGTGCTG	A16R	TGGATTCTGCCTAGCACCCCTAG
LOC_Os01g27940	2794U1F	TGCTGCAGCTAGGAGGCACC	2794UR	TCAATGGCGCCTGTGCCGATGATC
LOC_Os01g31360	3136UF	TCTCCCAGAACCAAGGACACCAGC	3136U1R	TGTTACTGCTAGTTCTGTCACCAGC
LOC_Os01g33420	3342U1F	TGAGACGACTTCTAATACCAGACATTCTGG	3342U2R	TGCAATCCATGCACGAGCTCCAG
LOC_Os01g33420	3342U3F	ACGGTCGGCAATGGGCAGCTC	3342U3R	CCGGATCAGTTAGCCAGCCA
LOC_Os01g33420	3342UF	TGCTGCGAATAAGATCGGGGGCCAC	3342U1R	TCATTGGTGCAGGAAGTGAGACCC
LOC_Os01g35040	3504U2F	CATGAGGCATACTGTGTGAGGCAC	3504U2R	CGCACATGTTTCGAAGCCGATCTCG
LOC_Os01g36390	3639U1F	CGGCTTCTACTCAGAACCTGTCATGG	3639U1R	TCACTGTCCTCTGACTTGGTCCGA
LOC_Os01g36390	3639U2F	AGATGTCTGATAATGCCCGAAGC	3639UR	CCTCATTGCAACGTAGCCACGGGTC
LOC_Os01g36390	3639U3F	TGAGCACCAGGTCTTGGATTTGCCC	3639U3R	ACGGAGCCCTCACTCTGCAG
LOC_Os01g36390	3639UF	TGAGATACCAGAAGGTGGCACTCC	3639U2R	TGCAATGGATACAGTCTGCTGC
LOC_Os01g37560	3756UF	GGATGGCAACAAGGACGACGCTGTC	3756UR	TGTCTGGTAGGCTAGCCCCCATTGG
LOC_Os01g37960	L08F	ACTCCCTGCCATTATAGGCCCAGGTC	L08R	TGCTAAGATTCTTCAGGAGCACCGT
LOC_Os01g39000	I06F	ACCTCTCAAAGATCATGGCGAAG	I06R	GGTCGTGCTTCTCGGCTTCC
LOC_Os01g39310	3931U1F	TCTTGGTCATTTACGGACTAGAAC	3931U1R	TGGCCAGGTGTCTTCACTGGCTGA
LOC_Os01g39310	3931UF	AGGCTCATGTGGTGTCAGTTCGTGC	3931U3R	TCCACAGAGACTTGCACTGGACAGGG
LOC_Os01g39310	3931U2F	TCAAAGTGACTATGTACCAGGTTTCG	3931U2R	ATCTGGGTAGAATCGCGAGCAG

Locus	FPrimer Name	F Primer Sequence	RPrimer Name	R Primer Sequence
LOC_Os01g41190	4119UF	GCTAAAGACGATCAAGCATGGGTGC	4119U1R	TGGCTCAACTGCTGAGCTGG
LOC_Os01g41190	4119U2F	CCACTATTGGCTGGTGCAGCTGGGGC	4119U3R	ACCTTACAGGGTGCTTCACTGACTG
LOC_Os01g42730	F24F	ACCGAGTACCACAATCGTTTGCTCCAGG	F24R	TGGGAGACTTTGTACTCTCCATCCCGAC
LOC_Os01g43170	M24F	GCCAGGGAACGACATCGCTTGATGG	M24R	CGCGTCATCTGCAAAAGCTGCTCTGG
LOC_Os01g44220	4422U2F	ACCTCGCCTCGATACTTGCCACCAACA	4422UR	GGTCGCGTTCTTCTGGATCACCACG
LOC_Os01g44220	4422UF	AGGCTCATCGATATTCCCATGAGCA	4422U1R	TGGTCATGCAGAGCTCTCGGGAG
LOC_Os01g47460	D15F1	GTCAGTCGTCTTTCGCGATGG	D15R	AGATGTAGCCTTCATCGCCCGAAGC
LOC_Os01g48720	4872U3F	TGACATCATCAACAAATGGTATTTTGG	4872U4R	TGTCTCAACTGCACAACCTTGTGTGACC
LOC_Os01g49940	4994U1F	AGTGTGTCAATGTTGCTGATGCTCCAG	4994U1R	TCAGATGTAGATTCTTTTGAGGTTGTG
LOC_Os01g49940	4994U2F	TCCACCTGAAGGCAGTGAGG	4994U3R	CTGCAATATCCGAGCAAGTGCGTC
LOC_Os01g49940	4994U3F	GGTGCTGCCTATATTTCCGAGACTAG	4994U4R	ACAAGCCTGGAAAGTATCCCCATGAC
LOC_Os01g49940	4994U4F	TCAAGCACTCGATGCTTTCCCAG	4994U5R	TGCAACTTCAAGCTGTGGTGGAGG
LOC_Os01g49940	4994U5F	TGGCATACCAGACTCAACCACGCTC	4994UR	CGCAAATGGCATTGGTGCCTGCTG
LOC_Os01g51380	5138U1F	ATCCCAGCGAACGGACATACC	5138U2R	AGTCAGGATGACAAGAACATGCTTCCCAC
LOC_Os01g51380	5138U2F	GCAAATGATCCAACCATCGAGCG	5138UR	GGCGAGTCATGCCTTCGCCAATAGC
LOC_Os01g51380	5138UF	AGATGGCACAACTAGGCGTGGTCAG	5138U1R	ACCAGCCCAGCCTGACGGCA
LOC_Os01g51870	P23F	GGCTGTATGAGAAACCGTCGGAG	P23R	ACCCTTCACAAGCGCCTCCAGG
LOC_Os01g52050	5205U1F	AGCTTAGCCGCCTCCAGGACC	5205U1R	TCGTGCCAGAATGTGACCGGC
LOC_Os01g52050	5205UF	TCGCATACTGGGCAAGGCTC	5205UR	CGCGATTTTCAAATGCTCCAGCAGC
LOC_Os01g53880	K21F	TGCAGCTCCAGTGGTCGGCT	K21R1	CATGCATCTGCGGTATGC
LOC_Os01g55530	5553UF	AACAACACGCCTTCCACGGGTG	5553UR	ACTGAGGGTTTCTCGAGCCGAGACC
LOC_Os01g56630	5663U3F	ACTGTCTTGCAATTCTAGCAAACATGG	5663UR	ACGAGGATGGTTCTTGAACGGCTG
LOC_Os01g56630	5663UF	AGCGCTGAAGGTGCGACAAGTC	5663U2R	TCAACAGCATGTAGATCTGATTGGAAGTC
LOC_Os01g56670	C10F	AGAGGCTTACAAGGCTGCCATG	C10R	TCTGAAATGTGGCAGTAGCCGCAAGG
LOC_Os01g57110	5711U1F	TGTCCAGTCTCTAATTGCAGAGTCC	5711UR	ACTGTCAGTCGAGTCTGACCGCCAC
LOC_Os01g57110	5711U3F	AGTCTGAGGCTAATCCACCTGATG	5711U3R	TCTTGTAACCATTAATTGGGTTCCTGC
LOC_Os01g57110	5711U4F	GTCGCATGATGAAAGGGCGACTCT	5711U4R	TCTGCGAGAATTCCACCAGAGC
LOC_Os01g57110	5711UF	ACTTCACGAGTGAGGAGCGTGC	5711U1R	GGGCAGCCCTCACTTTAGAAGATGC
LOC_Os01g58080	5808UF	TAACCACGCGGCAGAAGACC	5808UR	TGCCTCATCGATTCGCGAGC
LOC_Os01g59130	5913U1F	GCTACTGAAGATGCCCGAAAGG	5913U2R	AGGGCCATCTGTCTGAACCG
LOC_Os01g59130	5913U2F	ACGAGACTTACTTATGCTCCAGAGC	5913U3R	TCATCGTGGTGGAATATGACTCCCTTGC
LOC_Os01g59130	5913U3F	TGACTACGCATCAAAATCCGTCCATG	5913UR	ACCGTGGACTTGATCGCCTG
LOC_Os01g60660	6066U1F	ACACTTATCAAGGAGAATGATATTGTGAGCC	6066U2R	AGAAGGTGCAGGTGCAGGTG
LOC_Os01g60660	6066U4F	TGTGTGCTTCAAATAAAGATCATACAGTTG	6066UR	TCCATTCTCGTCGGTACGGAG

Locus	FPrimer Name	F Primer Sequence	RPrimer Name	R Primer Sequence
LOC_Os01g60660	6066UF	TTCACGCAGAGCAAGCAGGCCATG	6066U1R	TCATCCTGCGGTGAGGTCACG
LOC_Os01g65920	6592UF	TGGCACAGACACCACCTTGTCAGG	6592U1R	ACTAGTGAGGTAGATCAAAGCAGCATCAC
LOC_Os01g67210	J09F	TGGAGTCCACCCCGGAGGACCA	J09R1	TGCAAGGACTCTTGCTTTGGGA
LOC_Os01g67590	F06F	GTCACTACTGAAGCTTGCCAGTCAC	F06R	AGCATGAAGCTGGGCTGGTCC
LOC_Os01g68380	6838U2F	CTGGCACTTGAGACCATCAGATCG	6838U1R	TGCTAGAAGATGCGTCTCCGAAGG
LOC_Os01g68380	6838U3F	TGGCAACGGATAACTCAGCATTG	6838U2R	TCCAATAGAATGACTTCTGCTGCGAG
LOC_Os01g68380	6838UF	ACGGTTTTGCATCACTAGAGCAGG	6838U3R	AGAGGTCAACTGGCTGATAACGTGTGC
LOC_Os01g68770	6877U2F	GCACAGGTCTTCTTCCGCTGGAG	6877UR	AGGTCAGGGCCATAGAACTGCTCG
LOC_Os01g69970	6997U2F	ACTGTGAAAGTTTGGGACATGAGTGCTGG	6997UR	AGAAACGGCCTTTTCGTGTGCATGC
LOC_Os01g70670	7067U1F	ACCAGGTGGTTCTTCTCTCCAGATGC	7067UR	ACAATTCAGCTTTGCGTCTGTGG
LOC_Os01g70670	7067U2F	TGCTCTCCTTACATCTGACGAACTGTGG	7067U2R	AGCCAAAGACCATCGTTTACAGCTTTGC
LOC_Os01g70670	7067U3F	TGGTCCATGGAGAGAAGCTGGAGC	7067U3R	ACGGTGACCTTTGCCAGTAGAGC
LOC_Os01g70670	7067UF	ACTGCACCTCTTTGGGGGTGGTAGG	7067U1R	TGACTAAGCGGTGTTCTCATGTCCAAGG
LOC_Os01g73790	7379UF	TCTGCGAGAAGCTCATGGCCGAGAC	7379U1R	TCCTTGGGGAGGAAGCTGCCGA
LOC_Os01g74650	7465U1F	GCCCTGCTGTAGTGTCAAGGACAGG	7465U2R	CGCCAAAAGCCTTGAGTAGGACTCTCC
LOC_Os01g74650	7465U2F	TGCTTTCATTGCTGCTTCCAGGGG	7465U3R	GCTCCAATGCCTTGAATCTTATGTGGGC
LOC_Os01g74650	7465U3F	ACCTACTGAAAGTAACATACTCTCTGGTGG	7465U4R	GCAATCCTTCCTGAAGAGCCAACTGC
Crtiso	CRT3F	AGTCATACCAGACCCTTCTACG	CRT3R	GTGTCCCATCTTGTGGCATTTG
Eif4E	eIF4E-F5	CTGGACCTTCTGGTTCGACAAC	eIF4E-R5	TAGCATCCGCCCGCTACAAGCT
Eifiso4E	eIFi(so)4E-F4	TCATCCTCCCCATCCCAAAACC	eIF(iso)4E-R1	ATCTCGGTCAAACACTACAAGGC
Gsp-1	GSP-ex-F1	TGCAGATTCAGTGCACGTAACA	GSP-ex-R3	AGTACATATTATTCCATGGTCAC
matK	MATK1F	AACCCGGAACTAGTCGGATG	MATK1R	CTCAATGGTAGAGTACTCGG
Psy2	PSYLF	ACTCCTGAAAGGCGCAAAG	PSY4R	CCTCCTCCATTTATCCGTCA

Expected Size (bp) Gene Amplification Promgram

1500 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0.7°C/cvcle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 1220 94°C 4min [94°C 45sec 67°C 45sec 72°C 2.30min]12x -0,7°C/cycle [94°C 45sec 58°C 45sec 72°C 2.30min]25x 72°C 4min 15°C 1650 94°C 4min [94°C 45sec 67°C 45sec 72°C 2.30min]12x -0.7°C/cycle [94°C 45sec 58°C 45sec 72°C 2.30min]25x 72°C 4min 15°C 1400 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0,7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0,7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 1470 1010 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0.7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0,7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 750 800 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0.7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 900 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0.7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 1000 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0,7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 700 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0.7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 1300 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0,7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 2100 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0,7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0,7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 1150 800 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0,7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 1500 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0,7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 1000 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0.7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 1000 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0,7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 900 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0.7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 850 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0.7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 950 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0,7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 800 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0.7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 1200 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0,7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 550 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0,7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 800 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0,7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 700 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0,7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 1000 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0,7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 1200 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0,7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 1000 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0,7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 1000 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0.7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 980 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0.7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 1600 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0,7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 1000 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0.7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C ? 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0.7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C

Expected Size (bp) Gene Amplification Promgram

1100 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0.7°C/cvcle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C ? 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0,7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 1500 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0.7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 1380 94°C 4min [94°C 45sec 67°C 45sec 72°C 2.30min]12x -0.7°C/cycle [94°C 45sec 58°C 45sec 72°C 2.30min]25x 72°C 4min 15°C 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0,7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 750 1200 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0.7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 1800 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0,7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 1050 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0.7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 1000 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0.7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 1200 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0.7°C/cvcle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 1600 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0.7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 1550 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0,7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 1400 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0,7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0,7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 1400 1800 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0,7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 1100 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0,7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 750 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0.7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 1050 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0,7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 1000 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0.7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 1000 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0.7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 1100 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0,7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 600 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0.7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 1400 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0,7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 1150 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0,7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0,7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 1600 1000 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0,7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 1000 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0,7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0,7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 1500 900 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0,7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 1200 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0.7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 900 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0.7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 700 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0,7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 1100 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0.7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 800 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0.7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C

Expected Size (bp) Gene Amplification Promgram

1200 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0.7°C/cvcle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 1000 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0,7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 1600 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0.7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 850 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0,7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0,7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 1800 1200 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0.7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 1300 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0,7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 1000 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0.7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 1200 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0.7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 850 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0,7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 800 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0.7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 1700 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0,7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 900 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0,7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0,7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 600 500 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0,7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 700 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0,7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C ? 94°C 4min [94°C 45sec 67°C 45sec 72°C 2min30]13x -0.7°C/cycle [94°C 45sec 58°C 45sec 72°C 2min30]26x 72°C 4min 10°C 1208 94°C 4min [94°C 30sec 55°C 45sec 72°C 1min30] 35x 600 94°C 4min [94°C 30sec 55°C 45sec 72°C 2min30] 35x 850 94°C 4min [94°C 30sec 50°C 45sec 72°C 2min30] 35x 800 94°C 4min [94°C 30sec 55°C 45sec 72°C 1min30] 35x 2300 94°C 4min [94°C 30sec 55°C 45sec 72°C 1min30] 35x

1000 94°C 4min [94°C 30sec 55°C 45sec 72°C 1min30] 35x

Thermocycler	DNA quantity (ng)	Sequencing Cycle				
Eppendorf	40	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	15	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	20	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	25	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec

Thermocycler	DNA quantity (ng)	Sequencing Cycle				
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	40	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	20	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	20	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	40	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
						—

Thermocycler	DNA quantity (ng)	Sequencing Cycle				
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	30	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	15	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
IMA55Long	10	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
IMA50Long	10	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	10	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	25	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec
Eppendorf	20	92°C 5min [92°C 30sec	50°C 20sec	60°C 4min30] 60x	15°C	rampes : 1°C/sec

Influence de la domestication et du système de reproduction sur la diversité et l'évolution des gènes chez les *Triticeae*

Au sein des Triticeae, il existe des variations du système de reproduction et de l'histoire démographique liée à la domestication de certaines espèces. L'objectif de cette thèse a été d'analyser l'impact de ces variations sur différentes caractéristiques du génome, telles que les niveaux de diversité intra-spécifique et les vitesses d'évolution moléculaire. D'une part, l'impact des évènements démographiques et sélectifs liés à la domestication sur le polymorphisme de séquence du blé a été quantifié. 21 gènes ont été séquencés chez trois espèces cultivées (Triticum dicoccum, T.durum et T.aestivum) et dans l'espèce sauvage T.dicoccoides. La comparaison des indices de diversité réels à ceux d'échantillons simulés par des méthodes de coalescence a permis d'estimer les pertes de diversité liées à la domestication. Nos résultats montrent que la domestication du blé a été accompagnée d'une forte érosion de la diversité nucléotidique sur les 21 gènes étudiés. La diversité des formes cultivées est de ~3 à 6 fois inférieure à celle de la forme sauvage. D'autre part, différentes prédictions théoriques quant à l'effet du système de reproduction et de la recombinaison sur l'évolution moléculaire ont été examinées au sein des Triticeae. 52 fragments de gènes ont été séquencés chez deux espèces diploïdes autogames (T.urartu et T. monococcum) et deux diploïdes allogames (Aegilops speltoïdes et Secale cereale) afin de contraster les niveaux de divergence synonymes (d_s) et non synonymes (d_N) entre les deux régimes de reproduction, l'orge servant de groupe externe. Les taux d'évolution (d_N/d_S) et la composition en bases ont été estimés sur les différentes branches de phylogénie. Nos résultats n'indiquent pas de différence d'efficacité de la sélection purificatrice entre espèces autogames et allogames, ce qui va à l'encontre des attendus. La transition vers l'autogamie pourrait être trop récente pour permettre la détection de l'accumulation de mutations délétères chez les autogames. Cependant nous avons détecté un signal de sélection positive plus efficace chez les allogames que chez les autogames. Enfin, l'équilibre de la composition en bases G et C vers lequel tendent les gènes est apparu significativement plus élevé chez les espèces allogames que chez les autogames, ainsi que dans les zones fortement recombinantes comparativement aux des zones faiblement recombinantes. Cette évolution semble principalement due au phénomène de conversion génique biaisée, dont l'intensité est fortement liée au taux de recombinaison efficace, ce dernier étant particulièrement faible chez les autogames en raison de leur forte homozygotie.

<u>Mots clés</u>: domestication, système de reproduction, recombinaison, évolution moléculaire, polymorphisme, sélection, démographie, conversion génique biaisée, *Triticeae*.

How domestication and mating system affect diversity and molecular evolution in Triticeae species?

In Triticeae species, the span of mating systems is wide and demographic changes due to domestication are common. The aim of this work was to analyse the impact such variations on genome characteristics such as the extent of intraspecific diversity and the rate of molecular evolution. The impact of domestication on nucleotidic diversity was studied on wheat. 21 genes fragments were sequenced for three cultivated forms (Triticum dicoccum, T.durum and T.aestivum) and for the wild ancestor T.dicoccoides. Comparing real diversity indexes to those of simulated samples using a coalescence method yielded an estimation of the diversity loss associated to the domestication process. Our results show that the domestication of wheat got along with an important erosion of nucleotide gene diversity on the 21 studied locus. Nucleotidic diversity is ~3 to 6 fold smaller in cultivated forms than that in the wild related species. The second part of the study aimed at testing empirically some theoretical predictions about the role of mating system and recombination on some parameters of molecular evolution. 52 genes fragments have been sequenced for two diploid selfing species (T.urartu and T.monococcum) and two diploid outcrossers (Ae.speltoides and Secale cereale) in order to contrast the synonymous (d_s) and non synonymous (d_N) rates of divergence between the two mating systems. Barley was used as outgroup. The rate of evolution (d_N/d_S) and the base composition were computed on the branches of the phylogenetic tree. Our results do not show any evidence of a reduction in selection efficiency among selfing species as predicted theoretically. Transition towards selfing may have been too recent to allow the detection of such a signal. But others results provided arguments on a higher number of codons under positive selection among outcrossers than in selfers. At least, the equilibrium GC content towards which genes are evolving was found significantly higher in outcrossers than in selfers, and higher in highly recombining regions than in lowly recombining regions as well. Such evolutionary trend seems to be led by biased gene conversion, which intensity greatly varies with the rate of effective recombination. This rate is especially low in selfing species because of their strong homozygosity.

Key words: domestication, mating system, recombination, molecular evolution, polymorphism, selection, demography, biased gene conversion, *Triticeae*.
