

Impacts d'événements démographiques et sélectifs sur la diversité des plantes cultivées : apports de l'analyse du polymorphisme allié à la théorie de la coalescence

Thomas BATAILLON^{(1,2)*}, Nathalie CHANTRET⁽¹⁾, Alberto CENCI⁽¹⁾,
Marie-Françoise GAUTIER⁽³⁾, Philippe JOUDRIER⁽³⁾,
Annabelle HAUDRY⁽¹⁾, Marie-Hélène MULLER⁽¹⁾, Charles PONCET⁽¹⁾,
Jean Marie PROSPERI⁽¹⁾, Joëlle RONFORT⁽¹⁾, Pierre ROUMET⁽¹⁾,
Sylvain SANTONI⁽¹⁾, Anne-Céline THUILLET⁽¹⁾, Jacques DAVID⁽¹⁾

⁽¹⁾UMR 1097 « Diversité et génomes des plantes cultivées », INRA,
34130 Maugeio, France.

⁽²⁾Centre de Bioinformatique, Université de Aarhus, Hoegh-Guldbergs Gade,
Building 1090, DK-8000 Aarhus C.

⁽³⁾UMR PIA (CIRAD-INRA-ENSA), 2, Place Viala,
34060 Montpellier Cedex, France.

* Auteur pour correspondance: T. Bataillon, BiRC, University of Aarhus,
Hoegh-Guldbergs Gade, Building 1090, DK-8000 Aarhus C.

Abstract: Impact of demography and selection on patterns of genetic diversity in wild/crop species complex: insights from the analysis of patterns of polymorphism using a coalescence framework. The comparative study of patterns of genetic diversity in crops and their wild ancestors coupled with recent advance in coalescent theory opens the way for making some historical inference on the demographic (and possibly selective) processes that have shaped genetic diversity in crops since domestication. The idea is that demographic events affect genome-wide levels of polymorphisms and the spectrum of allelic frequencies. Some forms of selection, in particular recent selective sweeps are expected to reduce diversity locally that is at the target of selection and in surrounding areas in the genome. Although a full reconstruction of the demographic history seems impossible based on patterns of polymorphism alone, contemporary polymorphism data can be used to calibrate simplified demographic models that can account for amounts and patterns of currently observable diversity. We illustrate that approach in two crop species with contrasted histories and mating systems. First, patterns of diversity are studied in the (allo)tetraploid wheat (*Triticum turgidum*) complex a species with a mixed mating system dominated by high levels of selfing. The intensity of bottle-

* Correspondance et tirés à part : tbata@daimi.au.dk

necks associated with major historical episodes of domestication and subsequent selection are quantified using microsatellite markers. Patterns of microsatellite diversity are also compared with patterns of sequence polymorphism detected at the *Gsp-1* locus. We found that domestication and subsequent selection resulted in a massive loss of genetic diversity and estimated a several fold reduction in the long term effective size realized in cultivated wheats relative to their wild ancestors (even after accounting for the probable loss of diversity in wild populations due to extensive habitat loss since Neolithic times). Next we analyze patterns of nucleotide diversity in alfalfa (*Medicago sativa*) an autotetraploid and outcrossing crop species. Sequence polymorphism among individuals spanning a wide geographic distribution was surveyed and analyzed at two intronic regions. These were located in respectively the alfalfa NADH-dependent glutamate synthase and PO146, a pollen pectate lyase-like gene. These genes were chosen while searching in GenBank a public genetic sequence database for nucleotidic sequences of *M. sativa* and not chosen with any phenotypic a priori. This allowed us to (i) quantify the extent of genetic divergence between wild and cultivated populations and thereby assess the impact of domestication on this divergence, (ii) assess the originality of the genetic makeup of wild Spanish populations of alfalfa relative to the rest of extant population throughout the species range, (iii) determine whether there is likely ongoing (recent) gene-flow among wild and cultivated populations in Spain.

domestication/ nucleotide diversity/ bottleneck/ microsatellites

Résumé. L'étude comparée du polymorphisme moléculaire des compartiments sauvages et cultivés des plantes domestiques permet d'étudier de façon indirecte les processus démographiques et sélectifs qui façonnent la diversité des plantes cultivées. Nous illustrons à travers deux exemples contrastés comment l'analyse du polymorphisme moléculaire (polymorphisme microsatellite et polymorphisme de séquence) permet dans le cadre de la théorie de la coalescence de faire un certain nombre d'inférences sur l'histoire des populations. L'intensité des goulots d'étranglement qui ont accompagné les différentes phases historiques de la domestication et de la sélection moderne sont étudiés chez le complexe des *Triticum turgidum* (blé dur). L'analyse du polymorphisme de séquence de deux régions introniques a permis, (i) d'analyser l'impact de la domestication sur la divergence sauvage/cultivé chez *Medicago sativa*, (ii) d'apprécier l'originalité du compartiment sauvage espagnol par rapport au reste de l'espèce, et (iii) de déterminer s'il existe des flux de gènes entre formes sauvage et cultivée notamment en Espagne.

domestication/ diversité nucléotidique/ goulot d'étranglement/ microsatellite

1. INTRODUCTION

L'excellence agronomique des variétés élités des principales espèces cultivées a été obtenue par une sélection humaine qui a débuté au Néolithique. Il est admis que cela s'est accompagné par une réduction plus ou moins

marquée suivant les espèces de la diversité surtout lorsque cette dernière est évaluée à l'aide de marqueurs moléculaires [1]. Ainsi dans les formes plus primitives ou sauvages, la diversité moléculaire apparaît plus élevée mais il est difficile de savoir si cette variation présente un intérêt intrinsèque ou si à l'exception de quelques gènes de résistance majeurs, elle n'est qu'une information obsolète, inutile pour l'amélioration du compartiment domestique moderne. De plus, la diversité morphologique des formes cultivées est souvent, paradoxalement, bien supérieure à celle de leurs ancêtres sauvages [3].

Une façon d'étudier cette question est de s'interroger sur les effets sélectifs et démographiques qui ont façonné la diversité des plantes cultivées. Si les effets sélectifs ont été forts (domestication rapide, adaptation rapide à des stress biotiques et abiotiques) et ont porté sur un petit nombre de plantes dont la descendance a ensuite connu un progrès foudroyant, une part importante du polymorphisme adaptatif présent à d'autres locus que ceux soumis directement à l'effort de sélection a disparu du compartiment élite (cf. par exemple le cas du gène *tb1*, un gène clef de la domestication du maïs [22]). Dans ce cas, l'étude comparée des niveaux de diversité entre les compartiments sauvage, primitif et domestique à l'aide de marqueurs neutres, anonymes et répartis aléatoirement dans le génome devrait nous renseigner sur la diversité génétique « utile » présente à d'autres points du génome dans ces compartiments.

L'analyse des polymorphismes révélés entre individus à l'aide de marqueurs neutres dans des modèles de génétique des populations basés sur le principe de coalescence [9] permet d'étudier plus précisément les forces démographiques (goulot d'étranglement, flux de migration) et sélectives façonnant la diversité des espèces (fig. 1). Il n'est probablement pas possible de reconstruire en détails l'histoire « démo-évolutive » récente d'une espèce uniquement à partir de la diversité génétique observable en divers points du génome. De telles données génétiques, associées à des informations historiques, sont cependant précieuses pour étudier de façon plus quantitative :

- l'impact des épisodes de domestication ou de sélections humaines intenses sur le niveau de diversité actuel des espèces cultivées,
- l'importance des flux de gènes récents *vs.* anciens sur le niveau actuellement observable de différenciation génétique existant entre les différents compartiments d'un complexe cultivé/sauvage.

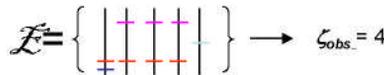
Nous présentons ici l'analyse du polymorphisme microsatellite et les premiers résultats de l'analyse du polymorphisme nucléotidique dans les compartiments sauvages et cultivés de deux espèces ayant des biologies et des histoires contrastées : le complexe des blés tétraploïdes (*Triticum turgidum*) avec un système de reproduction mixte et un fort pourcentage d'autogamie et celui de la luzerne cultivée (*Medicago sativa*) composé de plantes autotétraploïdes et allogames.

Encadré : Théorie de la coalescence des gènes & diversité neutre

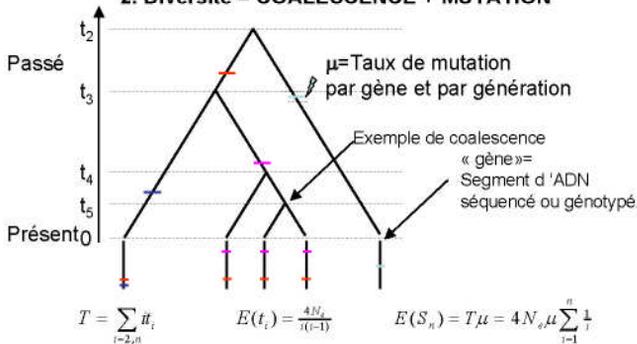
La diversité neutre observée, résumée par une statistique ζ_{obs} comme par exemple le nombre de site en ségrégation, ci après S_n , dans un échantillon \mathcal{X} de gènes (cf. 1) dépend uniquement des relations de parenté entre les gènes de \mathcal{X} , et du taux de mutation μ des gènes.

On mesure les relations de parenté entre 2 gènes en calculant le temps d'attente pour que ces gènes retrouvent leur ancêtre commun dans le passé (coalescence) (cf. 2). Dans une population de taille efficace N_e constante au cours du temps, il est possible de calculer la distribution des temps de coalescence, t_2, t_3, \dots, t_n , entre les gènes de \mathcal{X} comme une fonction de N_e (cf. Hein et al⁹ pour une introduction sur la coalescence et la présentation unifiée des démonstration mathématiques des résultats ci-dessous)

1. Diversité observée, résumée par la statistique ζ_{obs} (ici S_n)



2. Diversité = COALESCENCE + MUTATION



3. Typologie des statistiques de diversité

Espace des temps de coalescence	Statistique de diversité en fonction du modèle de mutation des locus marqueurs		
	Allozymes / RFLP Modèle Allèle Infini	Microsatellite Modèle de mutation "stepwise"	Données de Séquence Modèle Site infini Model site finis
Temps de coalescence moyen d'une paire de gènes	Hétérozygotie de Nei, H_e	H_s , ou distance $\delta\mu^2$	Hétérozygotie nucléotidique
Temps total dans l'arbre de coalescence	Variance du nombre de répétitions d'un microsatellite	Nombre d'allèles attendu dans un échantillon de n gènes	S_n

Figure 1 : Encadré illustrant le lien entre le processus de coalescence des gènes et les mesures de polymorphisme.

2. PATRONS DE DIVERSITÉ AU COURS DE L'HISTOIRE DE LA DOMESTICATION CHEZ *T. TURGIDUM*: DES GOULOTS D'ÉTRANGLEMENT EMBOITÉS

2.1. Contexte de l'étude

Deux espèces majeures de blés sont cultivées dans le monde, la première est le blé tendre (*T. aestivum*) qui figure parmi les céréales les plus importantes dans l'alimentation humaine, l'autre est le blé dur (*T. turgidum* ssp *durum*), essentiellement méditerranéenne et sert exclusivement à l'alimentation humaine (pâtes alimentaires et la semoule). Le blé dur appartient à un ensemble de taxons tétraploïdes, à $2n = 4X = 28$ (génomés A et B) qui comprend la forme sauvage considérée comme ancestrale, *T. t* ssp *dicoccoides*. La culture et la domestication de cette forme ancestrale ont commencé au Proche-Orient il y a une dizaine de milliers d'années. Cela a profondément affecté la morphologie de la plante et l'a conduite aux formes modernes de blé dur par des phases dont il reste encore aujourd'hui des variétés cultivées. Ainsi *T. t.* ssp *dicoccum* est considérée comme une forme primaire, conservant des caractéristiques de grains vêtus (terme impropre, la glume n'est pas adhérente, elle est plus épaisse, indurée et forme un angle plus fermé que la forme moderne de blé dur). Dans ce taxon, le grain n'est pas facilement libérable par battage. Les premiers témoignages archéologiques de grains nus, caractéristiques des formes modernes datent d'environ 7 000 à 8 000 ans. Après cette date, la morphologie change peu si ce n'est une évolution récente vers des lignées fixées de courte stature après les années 1960 pour les pays à agriculture intensive tandis que les cultivars de grande taille sont encore largement cultivés dans les zones où l'agriculture est restée traditionnelle.

La démarche proposée ici est donc de s'appuyer sur cette connaissance taxonomique pour établir un échantillon sur lequel sera analysée l'évolution de la diversité génétique sur un ensemble de marqueurs microsatellites et de comparer cette évolution avec celle d'un gène potentiellement important dans la qualité de la semoule. Ce gène, *Gsp-1* code pour une protéine basique et riche en cystéines entrant dans la composition d'un complexe protéique appelé « Grain Softness Protein » (*GSP*) ou « friabiline ». Ce complexe protéique est lié au caractère tendre du grain de blé. Le gène est présent dans la zone distale des bras courts des chromosomes 5A et 5B, dans le locus *Ha* (pour 'Hardness'), contenant d'autres gènes impliqués dans la qualité de la farine et soumise à des remaniements lors du passage à la polyploïdie [8], [2].

2.2. Matériel végétal et méthodes

L'échantillon contient 4 groupes : l'espèce sauvage *T. dicoccoides*, l'espèce primaire *T. dicoccum* et deux groupes de blés dur, *T. durum*, l'un basé sur des cultivars traditionnels, l'autre sur des blés plus récents. Pour l'étude de la variabilité sur les microsatellites, l'échantillon comporte 314 plantes. Pour le polymorphisme de séquence du gène *Gsp-1A* et *Gsp-1B*, une série de collections emboîtées réalisées par maximisation de la diversité a été construite [4]. Pour comparer les niveaux de diversité entre les quatre groupes, un effectif efficace réalisé a été calculé pour chacun des 15 microsatellites en combinant les données de polymorphisme et des données indépendantes sur le taux de mutation des locus microsatellites estimés directement [18], [19].

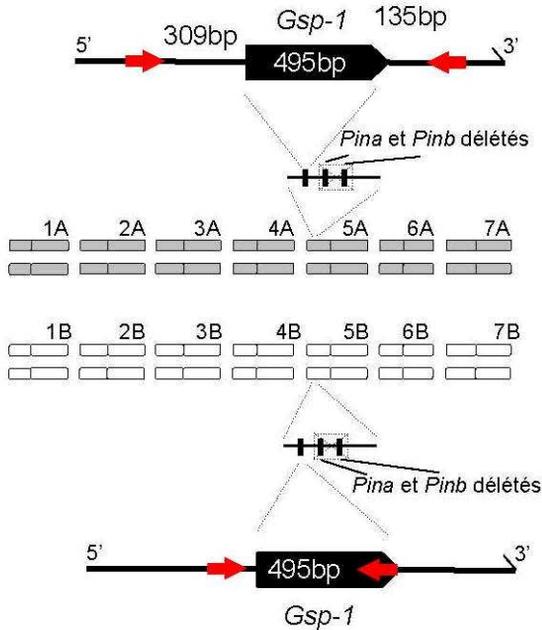


Figure 2 : Environnement génomique et structure des copies homéologues du gène *Gsp-1* chez *T. turgidum*. Le gène *Gsp-1* n'a pas d'intron et est d'une longueur de 495 bp. Les amorces utilisées pour amplifier de façon spécifique et séquencer chacune des deux copies homéologues du gène *Gsp-1* (génome A et génome B) sont représentées par des flèches rouges. NB : Les gènes de puroindolines (*Pina* et *Pinb*) sont absents dans les espèces tétraploïdes de génome AB alors qu'ils sont présents dans les espèces diploïdes progénitrices sauvages.

L'amplification des gènes *Gsp-1A* et *Gsp-1B* (fig. 2) dans les espèces polyploïdes a nécessité la mise au point d'amorces génomes spécifiques de

façon à pouvoir réaliser le séquençage des deux homéoallèles directement sur l'ADN génomique. Ces amorces ont été définies en recherchant du polymorphisme de type mutation ponctuelle ou insertion/délétion, spécifique de chaque génome (A et B). Les séquences du gène *Gsp-1* dans leur environnement génomique (extraites des séquences des BACs du locus *Ha* disponibles dans les génomes A de *T. monococcum*, A et B des blés dur et tendre [2]) ont été alignées. La spécificité de ces amorces a été testée systématiquement par PCR sur des espèces diploïdes de génome A (*T. monococcum* et *T. urartu*) et B (\approx *Ae. speltoïdes*). Les séquences (*forward* et *reverse*) ont été alignées avec le logiciel Staden Package (version 2003.0.1, <http://staden.sourceforge.net/>). Chaque site polymorphe a été vérifié sur le chromatogramme et seuls les pics sans ambiguïté ont été retenus pour définir des SNP (Single Nucleotide Polymorphism). Les estimations des paramètres de diversité sur le jeu de données ont été réalisées à l'aide du logiciel DnaSP [15] sur les parties codantes et non codantes.

2.3. Résultats

Les deux types de marqueurs donnent des résultats proches (fig. 3). La diversité du compartiment cultivé a été drastiquement réduite au cours des différents épisodes historiques qu'a connu le blé dur. Grâce aux microsatellites et à la connaissance de leur taux de mutation, il nous apparaît que l'espèce sauvage, bien que renfermant un niveau de polymorphisme important ne figure pas parmi les espèces les plus riches en diversité, son effectif génétique efficace absolu n'est pas très élevé (32 500 contre environ 900 000 pour le Maïs [23]). Cette diversité initiale, ou bien ce qu'il en reste aujourd'hui, a été divisée par plus de deux lors du passage à la forme *T. dicoccum*. Si la diversité encore présente au sein de cette forme est représentative de la diversité de ce taxon juste après la domestication, on peut dire que la domestication a créé une importante réduction de diversité mais qu'elle a été accentuée ensuite et a continué de l'être lors du passage à la sélection moderne, l'effectif efficace du compartiment élite n'étant plus que de 1 300 en moyenne (pour des résultats plus détaillés, voir [8]).

Les mêmes tendances se retrouvent au niveau du polymorphisme nucléotidique sur les deux homéologues du gène *Gsp-1*. La perte de diversité au passage vers la forme domestique apparaît modérée (la *Gsp-1B* a même plus de sites polymorphes chez *T. dicoccum* que chez *T. dicoccoides*) mais la forme moderne (nous n'avons pas fait de différences entre cultivars traditionnels et lignées modernes) n'a plus qu'un seul haplotype pour la *Gsp-1A* et deux pour la *Gsp-1B*. Sur *T. t. ssp. durum*, l'intensité du goulot n'est pas calculable sur la *Gsp-1A* (elle est infinie). Du point de vue de la valorisation des ressources génétiques, 5 haplotypes de la *Gsp-1A* chez *T. t. ssp. dicoccoides*

portent des mutations non synonymes qui changent la séquence protéique. Notons que la perte de diversité sur les marqueurs microsatellites est une moyenne sur 15 locus alors que la diversité nucléotidique est mesurée sur deux locus. L'extension de ce travail de séquençage à plusieurs locus permettra de faire une comparaison valide des deux patrons de diversité. Les introduire dans le blé dur moderne pourrait être intéressant pour voir si la variation allélique de la protéine *GSP* peut créer une diversité phénotypique sur la friabilité de la farine.

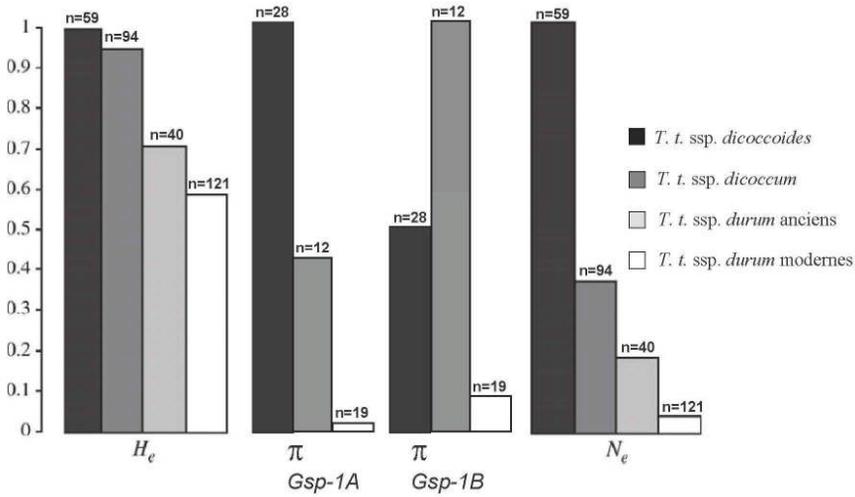


Figure 3 : Comparaison des patrons de diversité dans différents compartiments du complexe de *Triticum turgidum*. Les hétérozygoties de Nei (H_e) de chaque compartiment, moyennes sur 15 locus microsatellites sont exprimées en valeur relative par rapport à la diversité moyenne observée chez le groupe « dicoccoides ». Au milieu les patrons de diversité nucléotidique chez les deux copies homéologues du gène *Gsp-1* (copie A et B) sont comparées en utilisant l'hétérozygotie moyenne des sites (π). A droite les tailles efficaces (N_e) relatives, estimées en utilisant le polymorphisme microsatellites de chaque compartiment sont données. Pour l'analyse du polymorphisme des *Gsp-1*, les blés durs anciens et modernes sont regroupés.

3. PATRONS DE DIVERSITÉ ET DIVERGENCE DANS LES COMPARTIMENTS CULTIVÉS ET SAUVAGES DE *MEDICAGO SATIVA*.

3.1. Contexte de l'étude

Medicago sativa est une espèce allogame et pérenne, appartenant à un complexe d'espèces très représenté et diversifié en Europe et sur le pourtour méditerranéen. Les données historiques relatives à sa domestication et à l'expansion de sa culture sont peu nombreuses. La région présumée de domestication s'étend depuis l'Est de la Turquie jusqu'à l'Asie centrale, et des restes de graines permettent de penser que la domestication remonte à plus de 7500 ans av JC [6]. Son introduction en Europe résulte de différents courants de diffusion. En particulier, la péninsule ibérique semble avoir été une zone de convergence de différentes voies de diffusion de la luzerne cultivée [14]. Cette région présente une autre particularité : elle constitue à l'heure actuelle la seule zone, hormis le centre de domestication de l'espèce, où coexistent des formes sauvage et cultivée de l'espèce. Ces deux formes étant totalement inter-fertiles, ces zones de contact sont des zones potentielles de flux de gènes et d'introggression sauvage/cultivé [10], [11], [13]. Dans le cas de la luzerne, la divergence actuelle entre formes sauvage et cultivée doit donc être considérée comme la résultante de deux processus : la divergence induite par la domestication, d'une part, et des événements plus récents voire récurrents, de flux géniques sauvage/cultivé d'autre part. L'étude de l'histoire évolutive de cette espèce nécessite donc une approche et des outils statistiques permettant de distinguer l'effet « ancien » de la domestication de celui des flux géniques « actuels » entre formes sauvage et cultivée [21].

Dans ce contexte, l'étude avait pour objectifs : (i) de préciser notre connaissance de l'histoire évolutive de *M. sativa* et notamment, de quantifier l'impact de la domestication sur la divergence sauvage/cultivé chez cette espèce, (ii) d'apprécier l'originalité du compartiment sauvage espagnol par rapport au reste de l'espèce, et (iii) de confirmer l'existence de flux de gènes entre formes sauvage et cultivée notamment en Espagne.

3.2. Matériel et méthodes

Afin de représenter au mieux le compartiment cultivé, nous avons rassemblé 72 populations cultivées sur l'ensemble du vieux continent, en privilégiant les populations de pays afin de limiter les risques de confusion liés à la sélection moderne et à la mondialisation des échanges. Le compartiment sauvage était représenté par des populations naturelles de *Medicago sativa* ssp.

sativa (tétraploïde) collectées en Espagne, et dans la zone présumée d'origine de l'espèce (du Proche-Orient jusqu'en Asie centrale), et des populations naturelles de la sous-espèce diploïde supposée progénitrice de la forme cultivée *M.s. ssp coerulea*. Chaque population a été représentée par un individu choisi au hasard au sein du lot de graines disponible. Quatre autres espèces du genre *Medicago* (*M. falcata*, *M. glomerata*, *M. tunetana*, *M. marina*) ont été utilisées comme groupe externe. Deux fragments introniques d'environ 600 paires de bases ont été séquencés. Ces fragments, définis à partir des données disponibles dans les banques de gènes correspondent, l'un à un homologue d'un gène de pectate lyase [24] (ci-après *Pect*), l'autre à un gène de glutamate synthase NADH-dépendant [20] (ci-après *Glu*). Afin d'obtenir la séquence d'un seul des quatre allèles présents au sein des individus, un clonage a été réalisé pour chacun des fragments après l'amplification PCR [12].

3.3. Résultats : Polymorphisme de séquence et généalogie de gènes

Comme attendu, la forme cultivée présente moins de diversité que la forme sauvage de l'aire d'origine. La réduction de diversité observée sur les deux locus est cependant très différente : en utilisant l'indice de diversité θ_w (basé sur le nombre de sites polymorphes), la réduction de diversité de la forme cultivée par rapport à la forme sauvage est de 28 % et 51 % pour les gènes *Glu* ($\theta_{w,dom} = 0,0194$, $\theta_{w,sauv} = 0,0271$) et *Pect* ($\theta_{w,dom} = 0,0183$, $\theta_{w,sauv} = 0,0375$) respectivement. Les séquences obtenues dans les différents taxons (sauvage, cultivé) et groupes géographiques ne forment pas de clades différenciés. Néanmoins, quatre grands groupes de séquences, montrant une différenciation significative [12] apparaissent sur le gène *Pect*. L'un d'entre eux contient de nombreux représentants des populations naturelles collectées en Espagne et peu de séquences issues de plantes cultivées (fig. 4). Par contre, aucune association entre proximité génétique et statut sauvage/cultivé n'est observée pour ce gène. Pour le gène *Glu*, deux grands groupes de séquences apparaissent mais sans aucune association avec l'origine géographique des individus ni leur statut cultivé/sauvage ou taxonomique (Données non présentées).

Ces résultats suggèrent que les principales divergences observées entre séquences ont précédé la domestication et les séparations taxonomiques. Une autre explication pourrait être que les flux géniques sauvage/cultivé ou entre sous-espèces sont importants et ont effacé les divergences ancestrales entre compartiments. Cette faible divergence sauvage-cultivé ne nous permet pas de localiser plus précisément la zone de domestication de l'espèce. La seule situation révélant une divergence plus élevée est celle qui compare la forme sauvage espagnole au reste de l'espèce (forme cultivée et /ou sauvage de l'aire d'origine). Ce résultat peut être interprété comme le résultat d'une divergence associée à un refuge glaciaire. La péninsule ibérique semble

en effet avoir été une zone refuge pour de nombreuses espèces au cours des dernières phases glaciaires [5], [16].

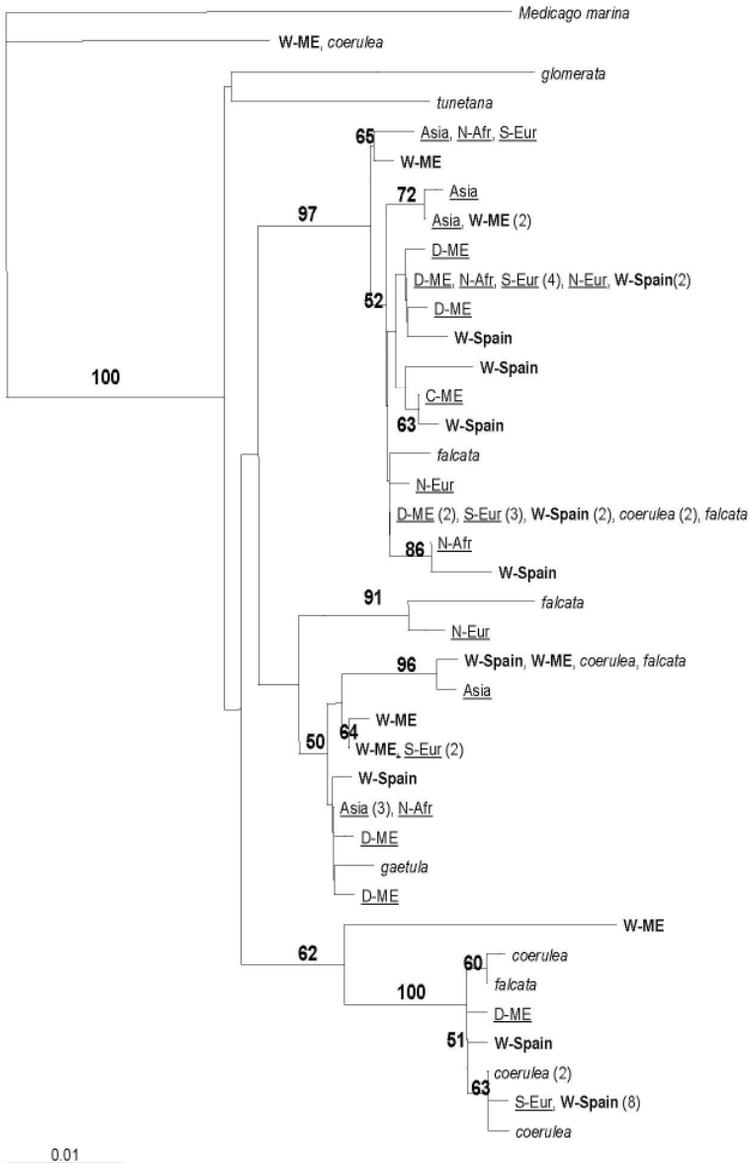


Figure 4 : Arbre (obtenu par la méthode dite du « neighbor-joining ») des différents haplotypes identifiés pour la région *Pect*. Les valeurs de bootstrap sont données en pourcentage lorsque la valeur excède 50 %. A côté de chaque haplotype est indiqué dans quel(s) groupe(s) l'haplotype a été identifié (le nombre entre parenthèse précise combien d'individus de ce groupe ont présenté cet haplotype (si ce nombre excède

1). C/D: Domestique, ME: Moyen Orient ; Spain: Espagne ; N-Afr: Afrique du nord ; S-Eur: Europe du sud ; N-Eur: Europe du Nord ; Asia: Asie. Les individus représentant le compartiment cultivé sont soulignés ; les formes sauvages de *M. sativa* ssp *sativa* sont en gras ; et les groupes extérieurs au complexe sauvage/cultivé sont en italique.

4. DISCUSSION

4.1. Applications des modèles de coalescence chez les plantes domestiques

Pour repérer et quantifier des goulots d'étranglement qui ont eu lieu dans l'histoire d'une population ou d'une espèce, différents modèles et tests fondés sur la théorie de la coalescence ont été proposés. Les événements démographiques passés laissent une empreinte dans le polymorphisme de séquence : *i.e.* un goulot d'étranglement suivi d'une expansion démographique crée un excès de mutations rares [7], [17] (fig. 1). Appliqués à l'étude de la domestication, ces tests n'utilisent donc que l'information contenue dans l'échantillon représentant le compartiment cultivé. Appliqués à l'échantillon représentant la forme cultivée de *Medicago sativa*, ils ne révèlent pas de signature nette d'un goulot d'étranglement (malgré une tendance vers ce type de scénario). Dans le cas d'une plante cultivée, on peut aussi utiliser l'information fournie par la forme sauvage de l'espèce. En faisant l'hypothèse que le compartiment sauvage est le descendant direct de la population progénitrice du compartiment cultivé, on peut modéliser le processus de domestication comme la divergence de deux populations depuis une population ancestrale, l'une ayant subi un goulot d'étranglement au moment de la divergence (fig. 5). Sous ce scénario, un moyen de détecter si le compartiment cultivé a subi une réduction de taille par rapport au compartiment sauvage est de comparer les diversités observées actuellement dans les deux compartiments et de confronter l'écart de diversité observé à celui attendu sous l'hypothèse nulle d'absence de goulot. En suivant cette démarche, chez le blé et la luzerne cultivée, nous avons pu montrer que la domestication et les phases ultérieures de sélection ont été associées à un goulot d'étranglement [18], [12].

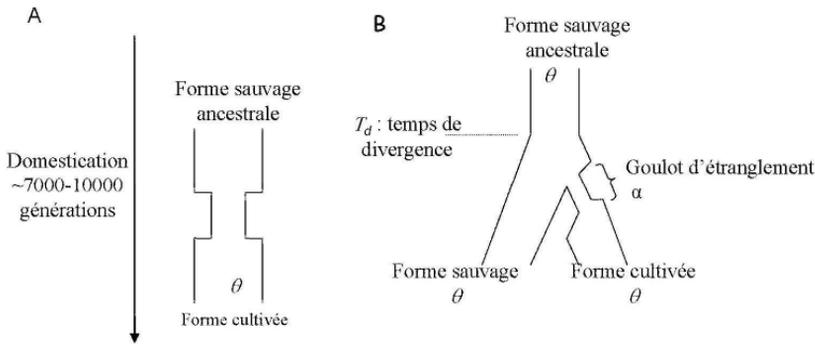


Figure 5: Représentation schématique des deux scénarios considérés pour étudier l'histoire évolutive de la luzerne cultivée. (A) Scénario représentant uniquement le goulot d'étranglement lié à la domestication (B) Scénario prenant en compte l'information obtenue sur le compartiment sauvage actuel. On considère ici que le compartiment sauvage est le descendant direct de la population sauvage dans laquelle la forme cultivée a été échantillonnée. θ taux de mutation efficace α intensité du goulot d'étranglement.

4.2. Vers l'analyse des patrons de polymorphismes au niveau génomique

Bien qu'un goulot d'étranglement ait pu être mis en évidence sur la luzerne, la divergence observée reste faible et donne peu de puissance pour distinguer l'effet de la divergence de celui des flux de gènes selon la méthode proposée par Wakeley [21]. Même la comparaison formes sauvages espagnoles/formes cultivées (qui présente la divergence maximale) ne permet pas de conclure à l'existence de flux de gènes. L'utilisation d'un nombre plus élevé de locus pourrait permettre de résoudre ce problème. Sur le blé les patrons de polymorphisme obtenus sur les deux gènes restent disparates et demandent à être confirmés sur un plus grand nombre de locus. Enfin considérer un grand nombre de locus simultanément permet de repérer des zones du génome ayant un patron de polymorphismes « atypique » suggérant l'action locale de la sélection [23]. Une telle démarche permet de compléter l'approche QTL classique pour repérer les segments génomiques impliqués dans la domestication/adaptation des plantes cultivées.

REMERCIEMENTS

Nous remercions le BRG pour son soutien financier à nos programmes de recherche.

RÉFÉRENCES

- [1] Buckler E.S.t., J.M. Thornsberry and S. Kresovich, Molecular diversity, structure and domestication of grasses, *Genet Res* 77 (2001) 213-8.
- [2] Chantret, N., et al., Molecular basis of evolutionary events that shaped the hardness locus in diploid and polyploid wheat species (*Triticum* and *Aegilops*), *Plant Cell* 17 (2005) 1033-45.
- [3] Darwin, C., *The Variation of Animals and Plant under Domestication* (Second edition). ed. N.Y.D.A. Co. 1883 Electronic Scholarly Publishing <http://www.esp.org/books/darwin/variation/facsimile/title3.html>.
- [4] David, J.L., et al. Impact of demographic and selective events on the current genetic diversity of the *Triticum turgidum* complex. in: Tenth international wheat genetics symposium. 2003. Paestum, Italy.
- [5] Dumolin-Lapègue, S., et al., Phylogeographic structure of white oaks throughout the European continent, *Genetics* 146 (1997) 1475-1487.
- [6] Flannery, K.V., Origins and ecological effects of early domestication in Iran and the Near East, in *The domestication and exploitation of plants and animals*, P.J. Ucko and G.W. Dimbleby, Editors. 1969, Gerald Duckworth & co Aldine Publishing Company: Chicago. p. 73-100.
- [7] Galtier, N., F. Depaulis, and N.H. Barton, Detecting bottlenecks and selective sweeps from DNA sequence polymorphism, *Genetics* 155 (2000) 981-987.
- [8] Gautier, M.F., et al., *Triticum aestivum* puroindolines, two basic cystine-rich seed proteins: cDNA sequence analysis and developmental gene expression, *Plant Mol. Biol.* 25 (1994) 43-57.
- [9] Hein, J., M.H. Schierup, and C. Wiuf, *Gene Genealogies, Variation and Evolution: A Primer in Coalescent Theory*. 2005, Oxford: Oxford University Press.
- [10] Jenczewski, E., J.M. Prosperi, and J. Ronfort, Differentiation between natural and cultivated populations of *Medicago sativa* (*Leguminosae*) from Spain: analysis with random amplified polymorphic DNA (RAPD) markers and comparison to allozymes, *Molecular Ecology* 8 (1999) 1317-1330.
- [11] Jenczewski, E., J.M. Prosperi, and J. Ronfort, Evidence for gene flow between wild and cultivated *Medicago sativa* (*Leguminosae*) based on allozyme markers and quantitative traits, *American Journal of Botany* 86 (1999) 677-687.
- [12] Muller, M.-H., et al., Domestication history in the *Medicago sativa* species complex: inferences from nuclear sequence polymorphism, *Molecular Ecology* in press (2006).
- [13] Muller, M.-H., et al., How mitochondrial DNA diversity can help to understand the dynamics of wild-cultivated complexes. The case of *Medicago sativa* in Spain, *Molecular Ecology* 10 (2001) 2753-2763.
- [14] Prosperi, J.M., P. Guy, and F. Balfourier, Ressources génétiques des plantes fourragères et à gazon, (1995). Pages 2
- [15] Rozas, J. and R. Rozas, DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis, *Bioinformatics* 15 (1999) 174-5.
- [16] Taberlet, P., et al., Comparative phylogeography and postglacial colonization routes in Europe, *Molecular Ecology* 7 (1998) 453-464.

- [17] Tajima, F., Statistical method for testing the neutral mutation hypothesis by DNA polymorphism, *Genetics* 123 (1989) 585-595.
- [18] Thuillet, A.C., et al., Estimation of long-term effective population sizes through the history of durum wheat using microsatellite data, *Genetics* 169 (2005) 1589-99.
- [19] Thuillet, A.C., et al., Direct estimation of mutation rate for 10 microsatellite loci in durum wheat, *Triticum turgidum* (L.) Thell. ssp durum desf, *Mol Biol Evol* 19 (2002) 122-5.
- [20] Vance, C.P., et al., Alfalfa NADH-dependent glutamate synthase:structure of the gene and importance in symbiotic N₂ fixation, *The Plant Journal* 8 (1995) 345-358.
- [21] Wakeley, J., The variance of pairwise nucleotide differences in two populations with migration, *Theoretical Population Biology* 49 (1996) 39-57.
- [22] Wang, R.L., et al., The limits of selection during maize domestication, *Nature* 398 (1999) 236-9.
- [23] Wright, S.I., et al., The effects of artificial selection on the maize genome, *Science* 308 (2005) 1310-4.
- [24] Wu, Y., et al., PO146, a new member of pollen pectate lyase-like gene family from alfalfa, *Plant Molecular Biology* 32 (1996) 1037-1042.