



HAL
open science

La diversité génétique des plantes cultivées à l'heure de l'informatique génétique à haut débit: génomique comparative de la domestication

Jacques David, Sylvain Glémin, Gautier Sarah, Benoit Nabholz, Francois Sabot, Yan Holtz, Félix Homa, Stéphanie Pointet, Sandy Contreras, Laure Sauné, et al.

► To cite this version:

Jacques David, Sylvain Glémin, Gautier Sarah, Benoit Nabholz, Francois Sabot, et al.. La diversité génétique des plantes cultivées à l'heure de l'informatique génétique à haut débit: génomique comparative de la domestication. *Sélectionneur Français*, 2016, 67, pp.3-13. hal-01607736

HAL Id: hal-01607736

<https://hal.science/hal-01607736>

Submitted on 28 May 2021

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

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



Distributed under a Creative Commons Attribution 4.0 International License

LA DIVERSITE GENETIQUE DES PLANTES CULTIVEES A L'HEURE DE L'INFORMATION GENETIQUE A HAUT DEBIT : GENOMIQUE COMPARATIVE DE LA DOMESTICATION

Jacques DAVID¹, Sylvain GLEMIN² &

Gautier Sarah¹, Benoît Nabholz², François Sabot³, Yan Holtz¹, Felix Homa¹, Stephanie Pointet, Sandy Contreras¹, Laure Sauné, Morgane Ardisson¹, Roberto Bacilieri¹, Bouchaib Khadari⁴, Claire Lanaud¹, David Pot¹, Christopher Sauvage⁵, Nora Scarcelli³, James Tregear³, Yves Vigouroux³, Nabila Yahiaoui¹, Manuel Ruiz¹, Sylvain Santoni¹, Jean-Pierre Labouisse¹, Jean-Louis Pham³

¹ UMR AGAP, Montpellier Supagro-INRA-CIRAD, ² UMR ISEM-CNRS, ³ UMR DIADE-IRD,

⁴ Conservatoire Botanique National, ⁵ UR GAFL, INRA

Email : jacques.david@supagro.fr

1 - LES RESSOURCES GENETIQUES ET L'HISTOIRE DES ESPECES CULTIVEES

Les ressources génétiques ont une acception très large. Pour le sélectionneur, elles sont vues comme la diversité nécessaire à l'introduction de la variabilité génétique indispensable pour améliorer ses variétés, leur donner une plus grande rusticité ou des tolérances à différents stress, développer des produits d'usage spécifique ; la gamme des mises en œuvre de la diversité des plantes dans les variétés est très large.

Pour autant, la diversité mesurée pour une espèce à un instant de son histoire n'est pas figée, elle résulte d'une histoire évolutive complexe, au cours de laquelle ont interagi les pratiques de culture, de sélection, d'adaptation aux territoires, le hasard, les échanges de toutes sortes et ce pendant plusieurs millénaires depuis la domestication.

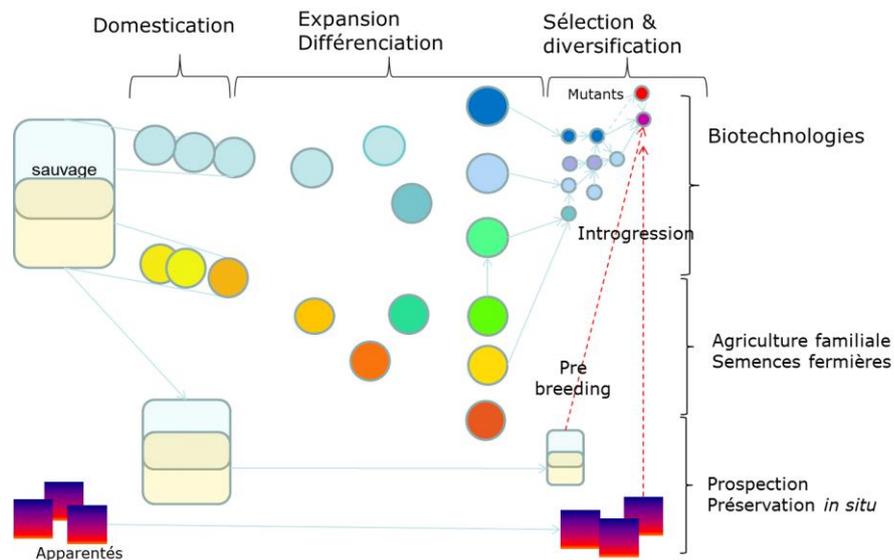


Figure 1 : Représentation schématique de l'évolution des compartiments cultivés et sauvages d'une espèce domestique. La diversité évolue en fonction des événements démographiques et sélectifs qui jalonnent l'histoire du compartiment cultivé. Ce schéma peut être modifié assez profondément en fonction de l'espèce considérée, notamment pour les espèces pérennes ou récemment domestiquées.

Nombre d’auteurs se sont penchés sur cette histoire et les relations qu’ont entretenues l’homme et les plantes cultivées, voire plus généralement les organismes domestiqués (Darwin, 1868 ; Candolle, 1883 ; Harlan, 1992). De manière générale, les ressources génétiques sont souvent caractérisées par leur éloignement du compartiment élite cultivé aujourd’hui (Buckler *et al.*, 2001). Elles sont les descendants actuels de formes plus anciennement cultivées, mais aussi de formes sauvages présentes au moment de la domestication. Un schéma très simplifié qui conviendrait à poser un cadre historique pour un grand nombre d’espèces, surtout de grande culture, est donné à la figure 1, bien qu’il existe de nombreuses déviations à cette représentation, notamment pour les espèces pérennes, ou les espèces à domestication quasi immédiate telles que des espèces polyploïdes comme le bananier (Perrier *et al.*, 2011). Un compartiment sauvage soumis à une agriculture pré-domestique (Tanno and Willcox, 2006) durant quelques siècles, dans une ou plusieurs zones d’origine, donne les premières formes cultivées qui vont ensuite se disperser et se différencier dans un espace géographique et temporel plus ou moins large selon les espèces. Certaines vont se déplacer à des latitudes homogènes comme le blé tandis que d’autres vont partir des tropiques et changer de latitude comme le maïs ou la pomme de terre.

Ces deux événements, domestication et expansion, fondent en général la base de la diversité utilisée par la sélection actuelle. La domestication, parce qu’elle est l’étape à travers laquelle s’est transmise la diversité initialement présente dans la forme sauvage, et l’expansion, car elle a créé une large diversité de formes dans l’immense espace (en termes de durée et de superficie) de l’agriculture traditionnelle. La mise en relation du Nouveau et de l’Ancien Monde a aussi créé de grandes modifications dans les patrons de diversité. Enfin les pratiques de sélection moderne, délibérées, basées à la fois sur une exploitation efficace de la diversité intra-variétale des populations traditionnelles par l’intercroisement réfléchi et contrôlé des meilleurs parents, et sur l’expérimentation et la diffusion de variétés beaucoup plus homogènes que les formes traditionnelles vont grandement modifier les patrons de diversité des formes actuelles.

Lors de ces grandes transitions, les génomes ont subi des pressions de sélection visant à adapter les variétés aux besoins des cultures. Elles ont pu perdre nombre d’aptitudes importantes pour une plante sauvage mais gênantes en agriculture comme la dissémination spontanée, ou la dormance (Nave *et al.*, 2016; Purugganan and Fuller, 2009 ; Glemin and Bataillon 2009 ; Ross-Ibarra *et al.*, 2007) et gagner des adaptations favorables à l’intensification des pratiques, comme des statures demi-naines pour les céréales à paille ou une accumulation de résistance aux maladies. Ces pressions de sélection ont été responsables d’effets de balayage sélectif plus ou moins prononcés : la sélection des allèles favorables à un locus pouvant avoir un effet d’entraînement sur les autres locus situés à proximité, neutres ou dont la valeur sélective était moins importante localement (Smith and Haigh, 1974). La sélection interagit alors avec la dérive génétique et vient réduire localement la diversité de gènes non touchés directement par la sélection. Par exemple, en séquençant le génome de la téosinte et du maïs domestique, des premières estimations pan-génomiques des transitions démographiques lors de la domestication montrent que la réduction de la taille efficace a été très intense au départ : seuls 5 % de la population de la téosinte a été mise à contribution pour la fondation du compartiment domestique tandis qu’une rapide et intense expansion démographique du maïs a suivi, permettant d’atteindre des tailles efficaces bien supérieures à celles connues par son ancêtre téosinte (Beissinger *et al.*, 2016).

Ainsi la littérature rapporte souvent que les plantes cultivées actuelles montrent moins de diversité moléculaire que les formes ancestrales (Glemin and Bataillon, 2009) mais une compensation s’observe fréquemment entre niveau de diversité (ou originalité génétique) et performances agronomiques.

2 - UTILISATION DE LA DIVERSITE GENETIQUE POUR DISSEQUER LES CARACTERES

2.1. Les panels de génétique d’association

L’accès au génome par les outils de séquençage a relancé l’intérêt des collections de ressources

génétiques. Des panels pour mener des études de génétique d’association en ont été sortis pour un très grand nombre d’espèces (Glaszmann *et al.*, 2010). La recherche sur l’architecture génétique des caractères a été révolutionnée par cette approche et l’utilisation des techniques modernes (Rafalski, 2002 ; Kilian and Graner, 2012). Ces panels apportent plusieurs allèles possibles par locus et l’histoire de leur recombinaison est beaucoup plus riche et complexe que celle découlant d’un croisement biparental simple suivi d’une fixation. Ces panels sont ainsi devenus de formidables plateformes pour la recherche d’associations entre polymorphismes et caractères quantitatifs. La plus grande difficulté étant d’assembler un panel adapté à la question posée (Glaszmann *et al.*, 2010). Nous ne rentrerons pas ici dans le détail des possibilités permises par les panels de génétique d’association qui a fait l’objet d’un grand nombre de travaux. Cependant, si la diversité présente au sein des panels est un atout, la complexité de l’histoire et de l’origine du matériel nécessite de prendre de multiples précautions dans l’interprétation des associations trouvées positives dans la mesure où elles peuvent être le fait d’une structure cachée du matériel (Korte and Farlow, 2013).

2.2. Détection des zones chromosomiques impliquées dans l’adaptation des formes cultivées

Avec le développement de la connaissance sur les génomes, de nombreuses statistiques de diversité et de neutralité par locus peuvent être calculées et visualisées le long des chromosomes pour réaliser de véritables scans génomiques (Siol *et al.*, 2010). Les empreintes laissées dans les génomes par les épisodes sélectifs majeurs se caractérisent, parmi d’autres statistiques pouvant être calculées à partir de données de polymorphismes, par trois propriétés essentielles : des pertes de diversité locale (mesurée par la diversité de Nei ou bien la diversité nucléotidique) plus importantes que celles subies par le génome en moyenne lors des épisodes de goulots démographiques, une augmentation du déséquilibre de liaison qui signe que la rapidité de l’évènement n’a pas laissé le temps à la recombinaison de rétablir l’équilibre des fréquences haplotypiques à proximité des gènes ayant subi la sélection, et enfin, une différenciation accentuée par rapport à la forme sauvage par la mesure de critères comme le *Fst* par exemple. Sur le soja, ces trois signaux concomitants ont permis d’identifier des zones soumises à sélection de manière très nette (Li *et al.*, 2013).

Ces « scans » génomiques donnent accès à des informations sans *a priori* sur les caractères concernés et permettent d’identifier les zones des génomes ayant connu le plus grand nombre de modifications (Vigouroux *et al.*, 2002 ; Wright *et al.*, 2005). Par exemple, sur le maïs, des estimations donnent deux à quatre pour cent des locus sous sélection lors de la transition sauvage vers domestique (Wright *et al.*, 2005). Très récemment, il a été suggéré que l’impact de ces sélections de quelques gènes à effets forts reste difficile à montrer chez le maïs (Beissinger *et al.*, 2016). D’autres travaux sur le haricot ont aussi montré que des échanges peuvent s’instaurer entre sauvage et cultivé et qu’une sélection différentielle, humaine d’un côté et naturelle de l’autre, crée des îlots de différenciation qui peuvent aider à pointer les zones dans lesquelles les gènes de domestication se trouvent (Papa and Gepts, 2003).

2.3. Interprétation du polymorphisme au niveau moléculaire

Si de nombreuses avancées ont eu lieu grâce à la production de marqueurs bon marché, nombreux et faciles à utiliser, elles sont souvent basées sur des polymorphismes déjà identifiés dans des panels de référence. L’information complète de séquence, via les méthodes modernes de séquençage dites de nouvelle génération, permet d’aller plus loin dans l’interprétation des patrons de polymorphisme entre les formes sauvages et cultivées. En accédant aux séquences codantes des gènes, il est possible de différencier la dynamique d’évolution des fréquences des mutations qui modifient les protéines (mutations non synonymes) et dont l’effet attendu est plutôt une altération de la fonction de la protéine (effet légèrement délétère), de celle des mutations qui ne modifient pas la protéine (mutations synonymes) dont on attend qu’elles soient *a priori* neutres en première approximation (Figure 2).

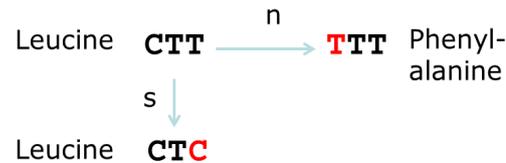


Figure 2 :Mutations synonymes (s) et non synonymes (n). Le codon initial CTT code pour une leucine. En raison du code génétique dégénéré, une mutation qui touche la troisième base du codon, (ici T mute vers C) ne change pas l'acide aminé. Cette mutation est dite synonyme (s). En revanche, un mutant C vers T à la première base du codon modifie l'acide aminé codé et donc est non synonymes (n) vis-à-vis de la protéine.

Ainsi les polymorphismes établis dans les gènes peuvent être utilisés pour calculer deux indices de polymorphisme, la diversité nucléotidique synonyme (π_s) et la diversité non synonyme (π_n). Sous l'hypothèse que la mutation fait entrer dans la population des mutations des deux formes, une façon de mesurer l'efficacité de la sélection est de comparer ces deux niveaux de polymorphisme en posant comme hypothèses que la plupart des mutations non synonymes sont légèrement délétères et que la sélection maintient leur fréquence à de faible niveau. Si la sélection est globalement attendue plus efficace dans une population que dans une autre, alors le ratio des polymorphismes π_n/π_s doit être plus faible dans la population où la sélection est la plus efficace, car elle maintient les mutations délétères à des fréquences plus faibles.

Il est évident que, de temps à autre, des mutations non synonymes peuvent se révéler avantageuses et qu'elles permettent l'adaptation. Dans ce cas, si elles échappent à la dérive lorsqu'elles sont encore rares, leur destinée n'est pas de rester polymorphes mais au contraire de se fixer rapidement en fonction de leur gain de valeur sélective. Reste alors le cas où ces mutations sont bénéfiques dans un environnement et pas dans un autre, ou celui où elles sont maintenues par des mécanismes de fréquence-dépendance. Malgré ces possibilités de sélection positive ou stabilisante, plusieurs études sont venues démontrer qu'utiliser le ratio π_n/π_s est un bon indicateur de l'efficacité de la sélection (Ellegren and Galtier, 2016).

Lorsque la sélection n'est plus assez efficace, la dérive peut entraîner l'augmentation en fréquence de mutations légèrement défavorables, voire permettre leur fixation. Ce phénomène est une des explications données de la vigueur hybride lorsque sont croisés des parents ayant divergé de la même source mais ayant accumulé des mutations différentes. Si chaque mutation a un petit effet, les effets se cumulent sur l'ensemble du génome et finissent par peser sur la valeur sélective de la population, c'est le fardeau génétique.

Cette sélection contre les allèles faiblement défavorables à effets cumulatifs dépend essentiellement de la taille efficace de la population (Ne). Plus la taille efficace est élevée, plus la sélection va pouvoir maintenir à faible fréquence des allèles à effets délétères faibles. Une autre façon de voir les choses est de regarder la valeur seuil de n pour laquelle la sélection reste efficace. Si Ne est petit, la dérive agit et fait augmenter en fréquence des mutations dont l'effet défavorable est plus grand que lorsque Ne est élevé. Conséquemment, les fréquences des allèles délétères vont pouvoir prendre des valeurs plus élevées dans les petites populations que dans les grandes, et ainsi il est attendu que π_n/π_s augmente lorsque Ne diminue.

Une transition de ce type est donc attendue lors de la domestication et dans la suite de l'histoire du matériel cultivé puisqu'elle se manifeste souvent par une réduction de l'effectif efficace au début de la domestication. Ceci aboutit à émettre l'hypothèse d'un coût de la domestication en terme d'accumulation de fardeau génétique. Il serait logique que cet effet « d'autostop génétique » de ces allèles délétères soit plus fort chez les espèces autogames qui recombinent moins (Nabholz *et al.*, 2014). Etudier ce phénomène en comparant différents couples d'espèces sauvages et cultivées

permettrait d'en donner la valeur générale et de mieux saisir l'impact des facteurs évolutifs sur le polymorphisme au-delà des gènes de domestication à fort effets, détectés par les analyses QTL ou en génétique d'association.

A l'heure actuelle, les effets de la domestication ont été évalués avec des données moléculaires sur un petit nombre d'espèces, majoritairement des espèces de grande culture. Seules quelques espèces pérennes comme le pommier (Cornille *et al.*, 2012) ont été étudiées et donnent des patrons relativement différents de ce qui a été observé sur les espèces annuelles. La longévité, la capacité de s'hybrider avec le compartiment sauvage, le faible nombre de générations séparant domestique et sauvage pourraient expliquer l'absence de goulot d'étranglement. Il est intéressant de prendre du recul et de documenter les effets de la domestication sur un nombre plus large d'espèces, si possible en suivant le même protocole et en les analysant de la même manière. C'est l'objet du projet ARCAD.

3 - ARCAD : UNE ETUDE COMPARATIVE DE LA DOMESTICATION

La communauté montpelliéraine comprend des équipes travaillant sur la diversité génétique de nombreuses plantes cultivées, méditerranéennes et tropicales. Elle possédait donc les ressources idéales pour effectuer une analyse comparative de la domestication en suivant un protocole standardisé et en utilisant les technologies de séquençage NGS. Ce projet a été soumis à la Fondation Agropolis qui l'a soutenu (<http://www.arcad-project.org/>). Comme beaucoup d'espèces ciblées ne possédaient pas de génome de référence au démarrage du projet, une approche par le RNAseq a été privilégiée pour obtenir les données sur une fraction réduite et répétable du génome. Elle permettait également en accédant à l'espace codant de pouvoir documenter les polymorphismes affectant les gènes.

Parmi les questions posées par cette étude figuraient les suivantes : la diversité du compartiment sauvage est-elle comparable entre espèces annuelles, pérennes, autogames ou allogames ? Les pertes de diversité lors du goulot de domestication varient-elle en fonction des traits d'histoire de vie des espèces (autogamie vs. allogamie, pérennité vs. annualité) ? Peut-on tester l'hypothèse du coût de la domestication au travers d'une accélération du taux de fixation des mutations non synonymes ?

3.1. Le projet

Tableau 1 : Espèces prises en compte dans le projet ARCAD, réparties selon des critères taxonomiques, de traits d'histoire de vie et de régime de reproduction. Les espèces marquées d'un * ont bénéficié de l'infrastructure ARCAD pour la production et l'analyse bio-informatique des données mais ont été financées par d'autres voies que le projet ARCAD pour la production des échantillons et le séquençage.

Espèce	Taxonomie	Statut	Régime de reproduction
Bananier	Monocot	Pérenne	Allogame
Palmier	Monocot	Pérenne	Allogame
Igname	Monocot	Annuelle	Allogame/Clonale
Engrain	Monocot	Annuelle	Autogame
Mil	Monocot	Annuelle	Allogame
Riz Africain	Monocot	Annuelle	Autogame
Sorgho	Monocot	Annuelle	Autogame
Fonio*	Monocot	Annuelle	Autogame
Cacaoyer	Dicot	Pérenne	Allogame
Caféier	Dicot	Pérenne	Allogame
Olivier*	Dicot	Pérenne	Allogame
Vigne	Dicot	Pérenne	Allogame
Tomate	Dicot	Annuelle	Autogame

Pour chacune des espèces, une dizaine de représentants sauvages et autant de cultivés ont été échantillonnés sur plusieurs tissus (fleurs et feuilles lorsque cela était possible) et autant de banques d’ADNc ont été élaborées et séquencées (Nabholz *et al.*, 2014). A ces paires de taxons étaient associées au moins deux espèces phylogénétiquement distantes afin de pouvoir disposer de groupes externes pour l’orientation des SNP en état ancestral et dérivé.

Le collectif mobilisé est large et recoupe plusieurs institutions de recherche créant de ce fait une animation scientifique et méthodologique importante. Au total, ce sont 332 banques de cDNA qui ont été produites sur 45 espèces. On a ainsi produit plus de 13 milliards de fragments de cDNA séquencés à leurs extrémités sur 100bp par la technologie Illumina (read pairs). Leur analyse a nécessité l’écriture de pipe-lines bio-informatiques spécifiques (Gautier *et al.*, sous presse ; Nabholz *et al.*, 2014).

Ces fragments ont été alignés, soit sur les génomes de référence lorsqu’ils étaient disponibles, soit sur des transcriptomes assemblés *de novo* (Gautier *et al.*, *in press*). Le pipe-line informatique a permis de produire les résultats figurant dans le tableau 2. La stratégie RNAseq s’est révélée pertinente avec les progrès réalisés ces dernières années sur les NGS. Plusieurs milliers de contigs ont pu être reconstruits pour chacune des espèces et la diversité a pu être ainsi caractérisée pour l’ensemble des espèces

3.2. Polymorphismes détectés

Tableau 2 : Polymorphismes détectés au sein des différentes espèces du dispositif ARCAD dans leurs compartiments sauvage et domestique. (N : effectif ; # Contigs : nombre de contigs sur lesquels les reads ont pu être cartographiés et des polymorphismes détectés ; SNP : Single Nucleotid polymorphism).

Espèce	Compartiment sauvage			Compartiment cultivé		
	N	# Contig	SNP	N.	# Contigs	SNP
<i>Coffea canephora</i>	12	14 581	148 508	12	15 722	49 790
<i>Dioscorea rotundata</i>	10	15 981	139 368	9	16 708	80 335
<i>Eleais guineensis</i>	10	16 975	54 238	10	17 220	51 722
<i>Musa acuminata</i>	10	21 422	286 784	10	277 146	274 179
<i>Olea europa</i>	10	19 549	209 402	10	21 168	168 372
<i>Oryza glaberrima</i>	9	12 169	23 578	7	12 169	7 597
<i>Pennisetum glaucum</i>	10	15 714	143 297	10	17 586	94 058
<i>Solanum lycopersicum</i>	10	18 887	73 778	10	16 396	40 189
<i>Sorghum bicolor</i>	10	11 397	46 298	10	13 231	41 266
<i>Theobroma cacao</i>	10	13 901	83 613	10	13 899	76 114
<i>Triticum monococcum</i>	10	5 847	11 148	10	14 066	20 776
<i>Vitis sylvestris</i>	12	14 397	133 697	12	16 222	195 603

Parmi les espèces dont les représentants sauvages étaient les plus polymorphes figurent le bananier diploïde (*Musa acuminata*) et l’olivier (*Olea europa*) tandis que le riz africain (*Oryza glaberrima*) et l’engrain (*Triticum monococcum*) ont les plus faibles nombres de SNP identifiés. Cela peut pour partie être aussi du au moment auquel ont été produites les données, les plus tardives ayant bénéficié des progrès des séquenceurs et ayant donc bénéficié d’une couverture plus importante.

3.3. Taille efficace et fardeau de mutation chez les formes sauvages

Comme nous l’avons introduit plus haut, des populations de grande taille efficace devraient avoir une meilleure efficacité vis-à-vis de la sélection et notamment vis-à-vis de l’élimination des mutations ayant des effets défavorables. Il est par ailleurs attendu que dans de grandes populations, la diversité nucléotidique synonyme soit plus élevée que dans de petites populations. Si le taux de mutation par paire de base est comparable entre espèces, alors comparer les taux de diversité revient à comparer les tailles efficaces. Ainsi, il est possible de tester l’attendu théorique que les populations à fort taux de polymorphisme (grand N_e) aient aussi des ratios π_n/π_s plus faibles (sélection efficace) que des populations à faible N_e .

Pour l’ensemble des espèces sauvages d’ARCAD la relation est donnée figure 3. La relation négative entre les deux paramètres apparaît très nettement entre π_s et le ratio π_n/π_s . Ainsi, le niveau et la composition du polymorphisme de départ des différentes espèces n’est pas comparable. Dans les espèces pour lesquelles il y a peu de polymorphisme, il est probable qu’une partie importante des variants non synonymes soient défavorables et qu’ils aient augmenté en raison de la dérive, au moins dans les conditions écologiques de la forme sauvage, un allèle délétère dans la forme sauvage pouvant devenir favorable dans la forme cultivée. La situation particulière du blé diploïde sauvage pourrait être due à une couverture moins bonne des transcriptomes ou bien à une évolution récente de sa taille efficace.

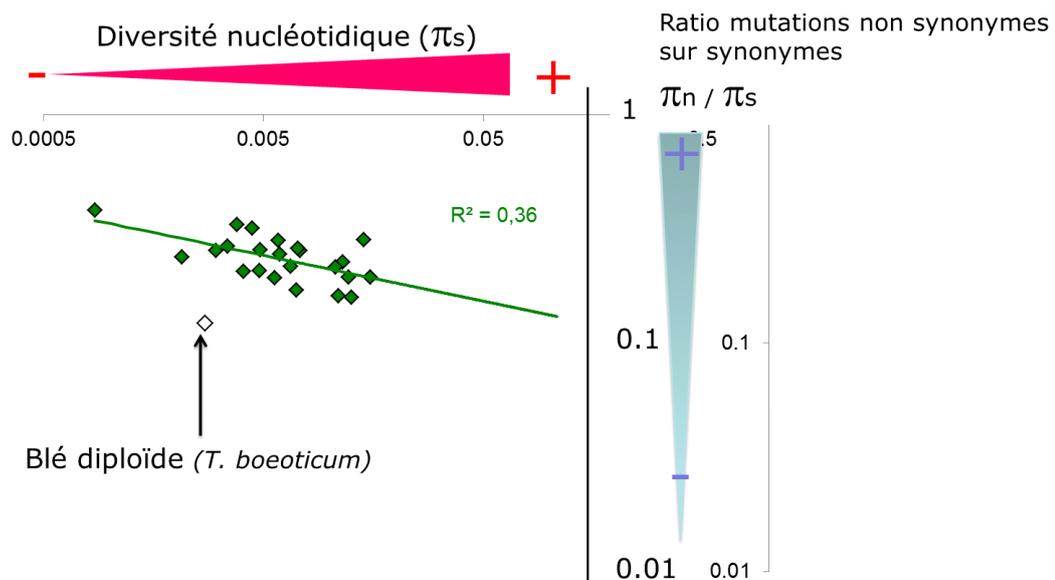


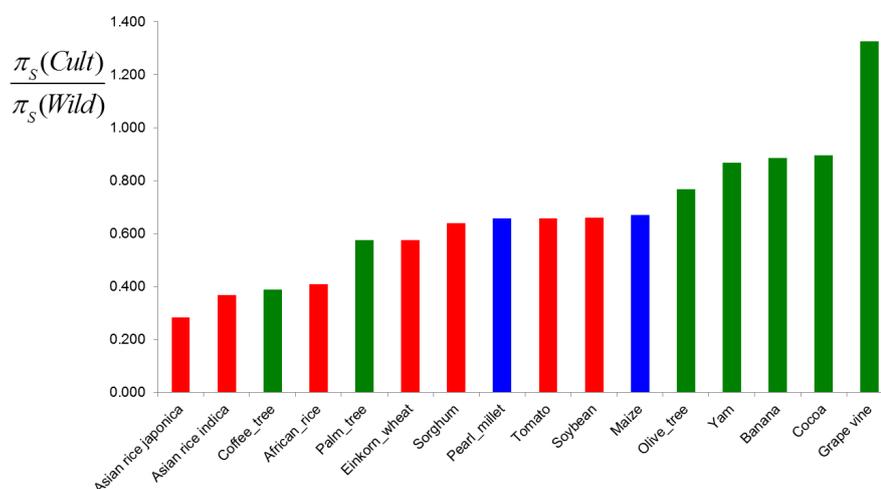
Figure 3 : Impact de la taille efficace (approximée par le taux de variation nucléotidique synonyme π_s) sur l’efficacité de la sélection des variants défavorables (approximé par le ratio π_n/π_s). Chaque point est un échantillon du dispositif Arcad.

3.4. Goulots d’étranglement

Les pertes de diversité lors du passage de la forme sauvage à la forme cultivée sont très variables (Figure 4). Elles sont de l’ordre de 60 à 70 % pour les riz asiatiques (*Oryza sativa ssp. japonica* et *indica*, *Oryza glaberrima*) et pour le café diploïde (*Coffea canephora*). Elles sont très réduites pour le cacaoyer (*Theobroma cacao*), le bananier diploïde (*Musa acuminata*), pour l’igname (*Dioscorea rotundata*) et l’olivier pour lesquels la perte de diversité est de 20 %. Les autres espèces se situent dans une zone intermédiaire de 40 % de perte environ. La vigne a un statut particulier. Pour notre

échantillon, la forme cultivée est plus diverse que les lambrusques. Cela interroge sur la représentativité de notre échantillon ou sur les processus qui ont mené à la création du pool domestique. L’Est du bassin méditerranéen et le Caucase n’ont probablement pas été bien échantillonnés ou il est aussi possible que des introgressions secondaires à la domestication aient eu lieu avec une espèce non encore identifiée (R. Bacilieri, com. pers.).

A l’instar de ce qui avait déjà été décrit chez le pommier (Cornille *et al.*, 2012) il est remarquable de constater que les espèces pérennes ont globalement moins perdu de diversité que les espèces annuelles si on excepte les situations particulières du café et du palmier à huile. Les formes domestiques de ces deux espèces sont probablement plus récentes en comparaison des autres espèces. Il est possible que pour ces espèces la domestication soit plutôt à assimiler à des épisodes de sélection consciente assez drastique, consistant à un échantillonnage conscient de plantes rapidement isolées génétiquement de leur compartiment sauvage et ne pouvant bénéficier de réintroduction régulière de diversité via des événements d’interfécondations.



* (Caicedo et al. 2007), ** (Lam et al. 2010), *** (Wright et al. 2005)

Figure 4 : Perte de diversité de plusieurs plantes domestiques lors de la transition sauvage-cultivé mesurée dans le dispositif Arcad. Les données du maïs, du soja et des riz asiatiques prises dans la littérature ont été rajoutées. L’intensité des goulots d’étranglement est mesurée par le ratio de la diversité nucléotidique synonyme de la forme cultivée rapportée à celle du compartiment sauvage. En rouge figurent les espèces annuelles, en vert les espèces pérennes.

3.5. Coût de la domestication

Pour les espèces pour lesquelles le goulot est fort, nous avons cherché à savoir si le taux de fixation de mutations non synonymes avait été accéléré dans le compartiment domestique en comparaison du compartiment sauvage. Sur le riz africain, pour lequel la diversité nucléotidique synonyme ($\pi_s = 0.0006$ per site) apparaît une des plus faible jamais rencontrée, le goulot a été extrêmement fort (Nabholz *et al.*, 2014). A partir des séquences des deux espèces externes (*Oryza meridionalis* et *Oryza sativa*), il a pu être prédit si les allèles observés chez le sauvage ou le cultivé avait un statut ancestral (car identique à celui des espèces externes) ou dérivé (car divergent de celui des espèces externes). En cumulant l’information sur l’ensemble des sites polymorphes, deux taux de divergence ont été calculés pour chacune des deux formes de riz africain, cultivé et sauvage : le taux non-synonyme (Dn) et le taux synonymes (Ds). Leur ratio Dn/Ds peut donner une indication sur la vitesse de fixation des mutations synonymes et non synonymes dans le compartiment cultivé comparativement au compartiment sauvage. Cela a été fait pour deux fractions de gènes, faiblement ou

fortement exprimés. Les résultats (Figure 5) montrent clairement que le taux de fixation de mutation non synonymes rapporté au taux de fixation synonymes (Dn/Ds) est beaucoup plus élevé dans la forme cultivée que dans la forme sauvage et ce notamment pour les gènes fortement exprimés. L’hypothèse d’un coût de domestication est donc très fortement soutenue par les données.

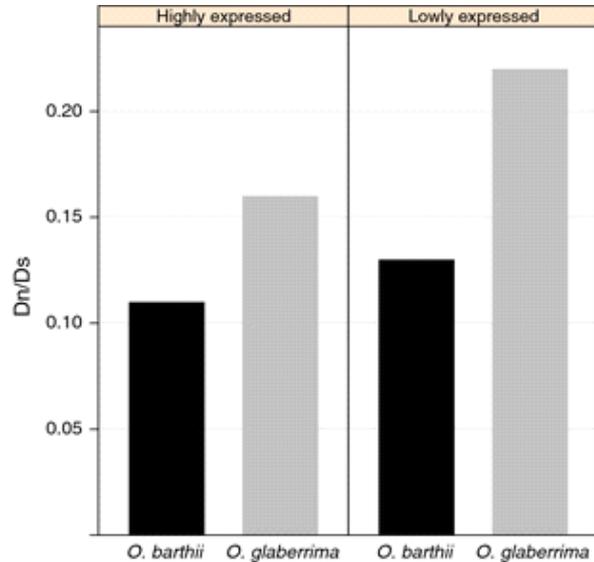


Figure 5 : Illustration du coût de la domestication chez le riz africain, *Oryza glaberrima*. Sur la forme sauvage (*O. barthii*), le taux de fixation de mutation non synonymes est très inférieur à celui observé sur la forme cultivée (d’après Nabholz *et al.*, 2014).

Nous avons effectué ce même calcul pour six autres espèces : trois annuelles (sorgho, blé et mil) et trois pérennes (bananier, olivier, cacaoyer). L’hypothèse de coût de la domestication est significative sur le blé et le mil mais n’est pas décelable sur les trois pérennes.

4 - CONCLUSIONS ET PERSPECTIVES

L’étude comparée de la domestication apporte de précieux éléments pour la compréhension de la dynamique de la diversité des plantes cultivées. Suivant leurs traits d’histoire de vie, l’ancienneté de leur domestication, leur régime de reproduction... la diversité génétique est répartie différemment et l’appel aux ressources des plantes sauvages doit être discutée au cas par cas, notamment lorsque la diversité du compartiment domestique n’a pas été réduite de manière drastique. La domestication ne découle pas d’un modèle unique et n’a pas eu les mêmes conséquences, la signification des polymorphismes n’est pas identique. Les ratios des mutations synonymes et non synonymes montrent que la sélection joue un rôle important dans leur répartition en interaction avec la dérive génétique.

La démarche décrite pour le projet ARCAD s’est révélée efficace, l’utilisation du RNAseq pour accéder à une sous partie du génome, même pour des espèces complexes, est envisageable et ARCAD a produit de nombreuses ressources bio-informatiques pour pouvoir utiliser relativement facilement ce type de données. Les données ARCAD sont publiques ; elles constituent d’importantes ressources pour l’utilisation des ressources génétiques comme les bases de données de polymorphismes ou les transcriptomes de référence assemblés *de novo*. Elles peuvent d’ores et déjà servir à établir des outils de génotypage à haut débit. Il est possible de procéder à l’exploration des séries alléliques de gènes candidats ou à la caractérisation de l’évolution des familles multigéniques.

Les données très riches sont encore à exploiter pour analyser les conditions dans lesquelles il sera possible de détecter les empreintes de sélection ; nos résultats sur le riz africain suggèrent qu’un trop fort goulot d’étranglement gêne considérablement leur détection.

Pour les sélectionneurs, il est important de prendre en compte l’hypothèse du coût de la domestication sur leur matériel et donc de s’intéresser à la nature des polymorphismes qui sont détectés et utilisés. L’apport de l’évolution moléculaire est précieux et devrait être beaucoup plus systématiquement pris en compte.

L’étude de l’évolution des patrons d’expression des gènes est également permise par les données RNAseq et les données s’accumulent pour suggérer que la diversité des patrons d’expression a elle aussi été touchée par les phénomènes de sélection pour les formes domestiques (Bellucci *et al.*, 2014).

“Journée ASF du 4 février 2016”
“ Apport des marqueurs moléculaires et de la génomique à l’utilisation
des ressources génétiques pour l’amélioration des plantes ”

REFERENCES BIBLIOGRAPHIQUES

- Beissinger TM, Wang L, Crosby K, Durvasula A, Hufford MB, Ross-Ibarra J (2016) Recent demography drives changes in linked selection across the maize genome. *Nature plants* 2:16084. doi:10.1038/nplants.2016.84
- Bellucci E, Bitocchi E, Ferrarini A, Benazzo A, Biagetti E, Klie S, Minio A, Rau D, Rodriguez M, Panziera A, Venturini L, Attene G, Albertini E, Jackson SA, Nanni L, Fernie AR, Nikoloski Z, Bertorelle G, Delledonne M, Papa R (2014) Decreased Nucleotide and Expression Diversity and Modified Coexpression Patterns Characterize Domestication in the Common Bean. *The Plant cell* 26 (5):1901-1912. doi:10.1105/tpc.114.124040
- Buckler ESt, Thornsberry JM, Kresovich S (2001) Molecular diversity, structure and domestication of grasses. *Genetical research* 77 (3):213-218
- Candolle Ad (1883) *Origine des plantes cultivées*. Bibliothèque scientifique internationale., vol XLIII. G. Baillièrre et cie, Paris,
- Cornille A, Gladieux P, Smulders MJ, Roldan-Ruiz I, Laurens F, Le Cam B, Nersesyan A, Clavel J, Olonova M, Feugey L, Gabrielyan I, Zhang XG, Tenaillon MI, Giraud T (2012) New insight into the history of domesticated apple: secondary contribution of the European wild apple to the genome of cultivated varieties. *PLoS genetics* 8 (5):e1002703. doi:10.1371/journal.pgen.1002703
- Darwin C (1868) *The variation of animals and plants under domestication*. Authorized edn. O. Judd & company, New York,
- Ellegren H, Galtier N (2016) Determinants of genetic diversity. *Nature reviews Genetics* 17 (7):422-433. doi:10.1038/nrg.2016.58
- Glaszmann JC, Kilian B, Upadhyaya HD, Varshney RK (2010) Accessing genetic diversity for crop improvement. *Current opinion in plant biology* 13 (2):167-173. doi:10.1016/j.pbi.2010.01.004
- Glemin S, Bataillon T (2009) A comparative view of the evolution of grasses under domestication. *The New phytologist* 183 (2):273-290. doi:10.1111/j.1469-8137.2009.02884.x
- Harlan JR (1992) *Crops & man*. 2nd edn. American Society of Agronomy : Crop Science Society of America, Madison, Wis., USA
- Kilian B, Graner A (2012) NGS technologies for analyzing germplasm diversity in genebanks. *Briefings in functional genomics* 11 (1):38-50. doi:10.1093/bfpg/elr046
- Korte A, Farlow A (2013) The advantages and limitations of trait analysis with GWAS: a review. *Plant methods* 9:29. doi:10.1186/1746-4811-9-29
- Li YH, Zhao SC, Ma JX, Li D, Yan L, Li J, Qi XT, Guo XS, Zhang L, He WM, Chang RZ, Liang QS, Guo Y, Ye C, Wang XB, Tao Y, Guan RX, Wang JY, Liu YL, Jin LG, Zhang XQ, Liu ZX, Zhang LJ, Chen J, Wang KJ, Nielsen R, Li RQ, Chen PY, Li WB, Reif JC, Purugganan M, Wang J, Zhang MC, Wang J, Qiu LJ (2013) Molecular footprints of domestication and improvement in soybean revealed by whole genome re-sequencing. *BMC genomics* 14:579. doi:10.1186/1471-2164-14-579

- Nabholz B, Sarah G, Sabot F, Ruiz M, Adam H, Nidelet S, Ghesquiere A, Santoni S, David J, Glemin S (2014) Transcriptome population genomics reveals severe bottleneck and domestication cost in the African rice (*Oryza glaberrima*). *Molecular ecology* 23 (9):2210-2227. doi:10.1111/mec.12738
- Nave M, Avni R, Ben-Zvi B, Hale I, Distelfeld A (2016) QTLs for uniform grain dimensions and germination selected during wheat domestication are co-located on chromosome 4B. *Theoretical and applied genetics* 129 (7):1303-1315. doi:10.1007/s00122-016-2704-4
- Papa R, Gepts P (2003) Asymmetry of gene flow and differential geographical structure of molecular diversity in wild and domesticated common bean (*Phaseolus vulgaris* L.) from Mesoamerica. *Theoretical and applied genetics* 106 (2):239-250. doi:10.1007/s00122-002-1085-z
- Perrier X, De Langhe E, Donohue M, Lentfer C, Vrydaghs L, Bakry F, Carreel F, Hippolyte I, Horry JP, Jenny C, Lebot V, Risterucci AM, Tomekpe K, Doutrelepont H, Ball T, Manwaring J, de Maret P, Denham T (2011) Multidisciplinary perspectives on banana (*Musa* spp.) domestication. *Proceedings of the National Academy of Sciences of the United States of America* 108 (28):11311-11318. doi:10.1073/pnas.1102001108
- Purugganan MD, Fuller DQ (2009) The nature of selection during plant domestication. *Nature* 457 (7231):843-848. doi:10.1038/nature07895
- Rafalski A (2002) Applications of single nucleotide polymorphisms in crop genetics. *Current opinion in plant biology* 5 (2):94-100
- Ross-Ibarra J, Morrell PL, Gaut BS (2007) Plant domestication, a unique opportunity to identify the genetic basis of adaptation. *Proceedings of the National Academy of Sciences of the United States of America* 104 Suppl 1:8641-8648. doi:10.1073/pnas.0700643104
- Siol M, Wright SI, Barrett SC (2010) The population genomics of plant adaptation. *The New phytologist* 188 (2):313-332. doi:10.1111/j.1469-8137.2010.03401.x
- Smith JM, Haigh J (1974) The hitch-hiking effect of a favourable gene. *Genetical research* 23 (1):23-35
- Tanno K, Willcox G (2006) How fast was wild wheat domesticated? *Science* 311 (5769):1886. doi:10.1126/science.1124635
- Vigouroux Y, McMullen M, Hittinger CT, Houchins K, Schulz L, Kresovich S, Matsuoka Y, Doebley J (2002) Identifying genes of agronomic importance in maize by screening microsatellites for evidence of selection during domestication. *Proceedings of the National Academy of Sciences of the United States of America* 99 (15):9650-9655. doi:10.1073/pnas.112324299
- 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 (5726):1310-1314. doi:10.1126/science.1107891