Available online at www.sciencedirect.com
ScienceDirect
Current Opinion in Plant Biology

Accessing genetic diversity for crop improvement

JC Glaszmann[1,2], B Kilian[3], HD Upadhyaya[4] and RK Varshney[2,4,5]

Vast germplasm collections are accessible but their use for crop improvement is limited—efficiently accessing genetic diversity is still a challenge. Molecular markers have clarified the structure of genetic diversity in a broad range of crops. Recent developments have made whole-genome surveys and gene-targeted surveys possible, shedding light on population dynamics and on the impact of selection during domestication. Thanks to this new precision, germplasm description has gained analytical power for resolving the genetic basis of trait variation and adaptation in crops such as major cereals, chickpea, grapevine, cacao, or banana. The challenge now is to finely characterize all the facets of plant behavior in carefully chosen materials. We suggest broadening the use of 'core reference sets' so as to facilitate material sharing within the scientific community.

Addresses
[1] UMR DAP, CIRAD, TA A96/03, Avenue Agropolis, 34398 Montpellier, Cedex 5, France
[2] Generation Challenge Programme (GCP), c/o CIMMYT, Int APDO Postal 6-641, 06600 Mexico, DF, Mexico
[3] Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Corrensstrasse 3, D-06466 Gatersleben, Germany
[4] International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru 502324, A.P., India
[5] School of Plant Biology (M084), Faculty of Natural and Agricultural Sciences, The University of Western Australia, 35 Stirling Highway, Crawley, WA 6009, Australia

Corresponding author: Glaszmann, JC (glaszmann@cirad.fr)

Current Opinion in Plant Biology 2010, **13**:167–173

This review comes from a themed issue on
Genome studies and molecular genetics - Plant biotechnology
Edited by Rajeev K. Varshney and Douglas R. Cook

Available online 16th February 2010

1369-5266/$ – see front matter
© 2010 Elsevier Ltd. All rights reserved.

DOI 10.1016/j.pbi.2010.01.004

Introduction

Genetic resources enable plant breeders to create novel plant gene combinations and select crop varieties more suited to the needs of diverse agricultural systems. A wealth of germplasm is accessible worldwide, with about 6 million accessions held in over 1400 gene banks [1]. Yet the collections are barely tapped (less than 1%) [2] by breeders, owing to the scarcity of information on accessions other than their taxonomic status and geographical origin.

Genome analysis tools provide access to thousands of polymorphisms, thus considerably broadening our capacity to monitor genetic diversity. Our whole approach to ecology and biological adaptation has been enriched [3,4]. *Arabidopsis thaliana* – the first plant with a sequenced genome – was used to develop and explore innovative applications including high-density array re-sequencing and genome-wide association mapping [5–7,8•]. Given their economic importance, major crops have also benefited from early investment in genomics. However, crops are not like wild plants in natural populations, that is, they have undergone and are still undergoing domestication. This is a complex anthropogenic process caused by numerous human populations with specific habits and needs [9•].

Over the past five years, an increasing number of studies have been carried out on the molecular diversity of crop plants and their wild relatives, illustrating various facets of the domestication process and suggesting ways of devising targeted approaches to access the diversity conserved in *ex situ* germplasm collections. Soon it will be possible to determine and compare the whole sequence of hundreds of accessions. We therefore advocate identification of a common set of reference materials to help **R.E.A.D.** (**R**epresent existing diversity – **E**nter the whole collection – **A**ssess phenotypic variation – **D**issect trait–gene associations) germplasm through concerted efforts within the research community.

Unraveling the drivers of crop evolution

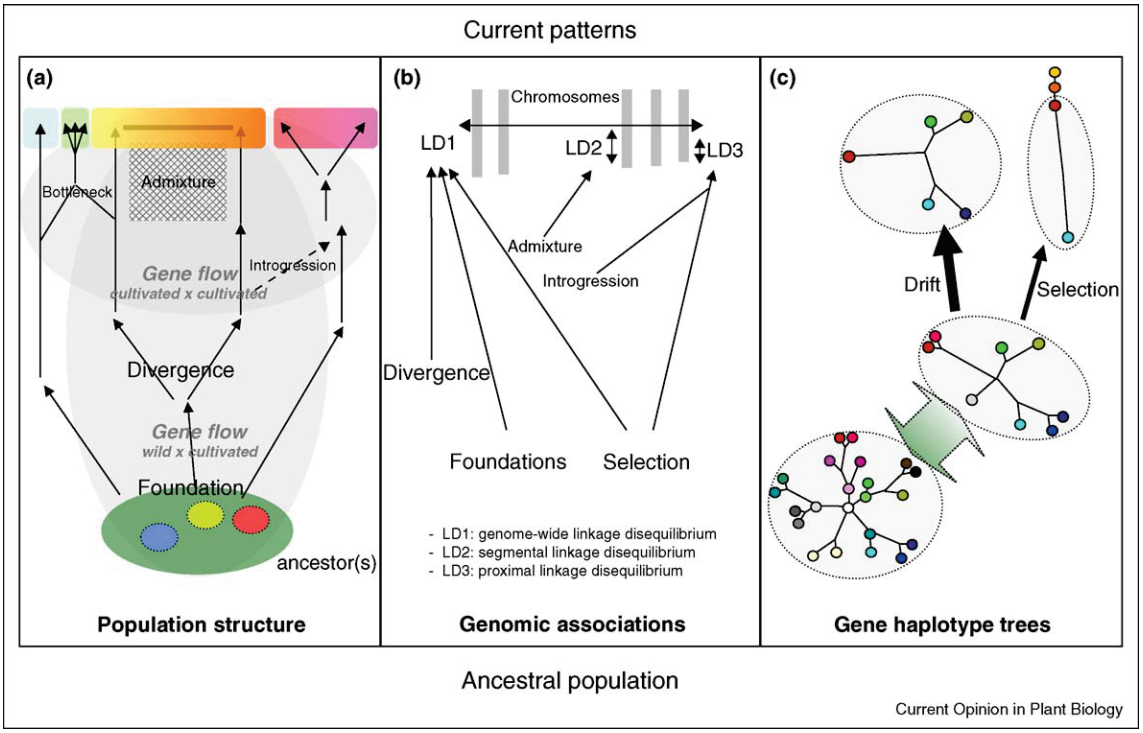
Over the past 12 000 years, humans have sampled, selected, cultivated, travelled through, and colonized new environments, thus inducing a plethora of bottlenecks, drifts, and selection. Plant breeders have accelerated the whole process by selecting preferred genotypes. Meanwhile, evolution was progressing, some genomes were being reshuffled and genes occasionally mutated. Overall, plant domestication tailored plant development and adaptation to meet the needs of human populations [10–12,13•,14–16]. Observing the concomitant modifications of the genome provides clues to the genetic bases of useful variation.

Global diversity patterns

Molecular characterization is now the favored means to quantify variation within large germplasm samples. New DNA sequencing and genotyping technologies provide the power to interrogate thousands to millions of diagnostic polymorphisms, across hundreds to thousands of genotypes, thus facilitating the analysis of genetic structure and providing a rationale basis to identify and select among the underlying lineages (Figure 1a). Such approaches not only resolve genetic relationships at fine scale, but they also provide important measures of genetic

Tecan Ex. 2110, Page 1 of 7
QIAGEN Sciences, LLC v. Tecan Genomics - IPR2025-00029

Figure 1



The impact of domestication and selection on genetic diversity patterns among cultivated forms. (**a**) Fundamental demographic processes contribute to patterning diversity during domestication from wild ancestors. Multiple domestications can result in separate foundations. Introgression between wild and cultivated forms is common and can result in selection of favorable wild alleles in cultivated backgrounds. Migration of cultigens with mankind typically causes drift, except for genes useful in adapting the crop to new environments. New sympatry between distinct lineages can result in recombination, from a balanced admixture to fine introgression. (**b**) These processes generate various types of linkage disequilibrium (LD), from global LD spanning the whole genome (LD1) to admixture LD, which extends to large chromosome segments when the number of generations has been limited (LD2), and to proximal LD around a gene under selection (LD3). (**c**) When studied within the specific window of high LD, haplotype networks are expected to collectively reflect the most significant lineages among domesticates and possibly new branches corresponding to novel variation arising through recent mutations. Discordance between global structure patterns and allele phylogenies are useful indicators of introgression, possibly under the action of selection.

divergence between and genetic diversity within the major genetic clusters that comprise crop germplasm. Numerous studies have been undertaken with a range of molecular marker technologies, focusing principally on nuclear markers. The precision and robustness of the patterns thus revealed now principally rest on a pertinent choice of materials, implying that a sufficient number of accessions are analyzed. DNA markers also allow access to cytoplasmic (i.e. mitochondrial or chloroplastic) variation, which is usually maternally inherited and not affected by recombination. This provides another view of genetic diversity, which is very helpful in highlighting the role of hybridization in the overall crop evolution process.

Molecular diversity studies assess all levels of genetic structure, ranging from relationships between species complex components, as illustrated by recent results on potato [17], tomato [18], wheat [19], or common bean [20], to the origin of particular genotypes. *Musa*, which encompasses banana and plantain crops, illustrates a species complex from which several very successful clonal cultivar groups have emerged, whose parentage can now be inferred through molecular markers [21].

Variations along the genome

Accurate genome coverage makes it possible to detect associations within the genome and to characterize the levels of linkage disequilibrium (LD) (Figure 1b).

The selection on 'domestication genes', while the rest of the genome is subjected to drift, can be documented through selection signatures – usually peaks of localized LD around a homogenized locus – within the cultivated gene pool. This has been most successful in maize [22,23•,24•].

The analysis of rice (*Oryza sativa* L.), whose cultivated forms are annual and predominantly self-pollinating, has led to in-depth descriptions of genetic diversity among

Tecan Ex. 2110, Page 2 of 7
QIAGEN Sciences, LLC v. Tecan Genomics - IPR2025-00029

landraces, while highlighting the impact of domestication in a highly structured species. This has prompted a number of very interesting reviews [25–28]. The best documented examples suggest that domestication has been built on diffuse selection of new alleles in different lineages, and on the mobilization within single lineages (e.g. *indica*) of domestication alleles that emerged in another lineage (e.g. *japonica*) through fine introgression (reviews [25,26,28,29,30•]). An alternative scenario would consist of a diversification essentially through the introgression of a major domestication gene into diverse wild forms [27]. The spread of domestication alleles by means of introgression could be a general phenomenon in cereal domestication [31].

Traits outside the domestication syndrome but under targeted human selection, such as fragrance in rice, are subjected to the same phenomenon [32,33]. Other illustrations of the occurrence of introgression, highlighted by single nucleotide polymorphisms (SNPs) of 20 rice cultivars for more than a hundred thousand loci [34••], include several examples of large genome segments, spanning several Mb, introgressed between varietal groups in all directions, including the group centered around the 'Aus' varieties [35]. In species with longer reproductive cycles, molecular data have revealed germplasm 'compartments' whose specific history determines important internal features such as LD. Cacao (*Theobroma cacao* L.) [36•] is an example of a fruit tree species where one of the major compartments typically displays admixture-derived LD over 15–20 cM. These examples illustrate cases where admixture and introgression are important in the domestication process and can be used for genetic analysis using extant materials.

Local sequence variation

One growing form of molecular characterization is allele re-sequencing in diverse materials. Local sequence variation can be finely interpreted within small genetic distance windows, where there is sequence variation but little or no confounding recombination. The order of mutation appearance can thus be inferred, while distinguishing between ancient, if not ancestral, and recent haplotypes (Figure 1c). These phylogenies can be individually affected by specific drift and selection history, but they collectively depict the structure of a crop's ancestry. They can also highlight variation emerging via positive selection.

Recent analyses of specific genes of proven or suspected function involved in flower, fruit, and seed development in tomato [37•], grapevine [38], barley [39•,40], rice [41•,42–44], and sorghum [45•] or plant adaptation to specific constraints in maize [46], rice [47••], and wheat [48•] have revealed multiple examples of mutations that may have occurred and been selected during domestication. Adaptive neo-diversity undoubtedly superimposes on ancestral diversity inherited from wild relatives.

Ecogeographical (environmental, ethnological, etc.) information concerning the materials (ideally included in the passport information in germplasm banks) is essential for locating and identifying unique variants for specific adaptation. This was recently illustrated with wheat *Pm3* alleles uncovered through the Focused Identification of Germplasm Strategy (FIGS) applied using molecular amplification from a proven disease resistance gene [49••], that is, allele mining using the known molecular structure of a locus.

With the growing body of gene function hypotheses, an increasing number of genes will be analyzed and allele phylogenies compared to the global population structure. This will shed new light on the domestication process, including the wild-to-domesticated transition and the differentiation between domestication lineages [31,50, 51], as well as on specific pressures affecting gene evolution.

Organizing access to diversity

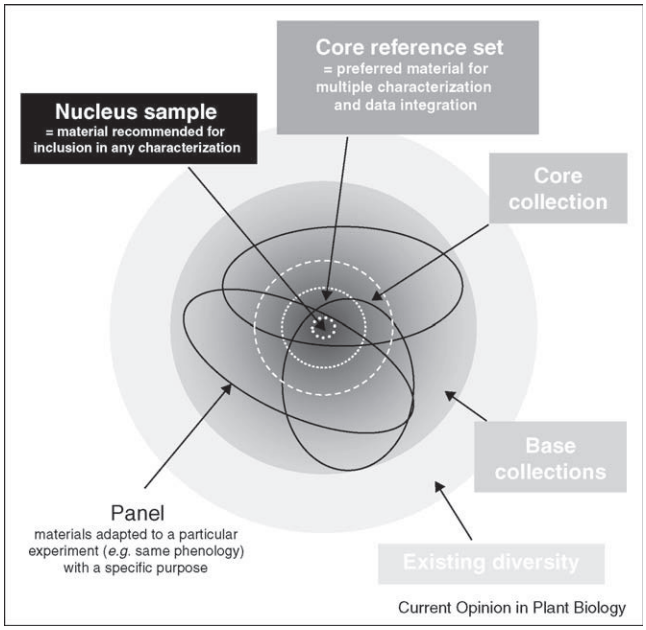
Access to genetic diversity contained in large germplasm collections continues to be a significant challenge. The core collection concept [52] was developed 25 years ago to facilitate access to the diversity available in these large collections. The idea is to identify a representative manageable sample upon which analysis will be concentrated before re-exploring broader ranging materials. The rationale underlying core collections has been thoroughly discussed [53•] and for many species has led to germplasm subsets containing 3000 accessions or more. In practice, however, core collections composed of thousands of accessions are too large for use in breeding programs, and as a consequence breeders have preferred to focus on dozens to hundreds accessions. The result has been increased incorporation of useful genetic and phenotypic diversity into cultivated material, as illustrated for example in rice [54], chickpea [55•], and groundnut [56]. 'Mini core' approaches focusing on only 1% of the collection [57,58] when whole collections are very large have been implemented for seven important crops at the International Crop Research Institute for the Semi-Arid Tropics (ICRISAT). The availability of molecular markers offers an opportunity for adjusting the size, the representativeness and the general quality of 'core' samples.

Accessible core reference sets

We suggest implementing the core collection concept through 'core reference sets'. As argued by AHD Brown, "One aim of the core is to build up a body of information on a restricted 'reference' set of lines" [53]. A crop core reference set is to be understood as *a set of genetic stocks that are representative of the genetic resources of the crop and are used by the scientific community as a reference for an integrated*

Tecan Ex. 2110, Page 3 of 7
QIAGEN Sciences, LLC v. Tecan Genomics - IPR2025-00029

Figure 2



Concepts proposed for organized access to genetic diversity. Existing diversity is the ultimate resource. The part accessible from *ex situ* collections is distributed among numerous accessions gathered in *base collections* exceeding the observation capacity of the community. The *core collection* concept can be used to focus broad surveys using molecular markers, which then provide complementary information for identifying a set of manageable size that represents the diversity thus described and can be distributed as a *core reference set*; currently, the option adopted is in the 50–500 range. This makes it a material of choice for contributing to *association panels* to assess diverse phenotypes and relate traits to genes and alleles through association studies. It should be accompanied by a *nucleus sample* that any experiment addressing and characterizing diversity could incorporate.

characterization of its biological diversity. The value of a formalized reference set will emerge from its use by the largest number of scientists. Ideally it will be adopted as a reference, and its description will capitalize on successive efforts and serve to integrate data. This community is potentially very broad, as the capacity to finely characterize materials is extremely varied and evolves with the advent of new technologies. Moreover, biologists may be interested first in a crop, but also in a trait, or a gene family, for example. However, the chance of making good biological sense of materials will certainly be greater when there is substantial data on this material to be tapped.

This requires a collective effort from the community, as advocated by Zhu *et al.* [59] and illustrated in barley [60,61•]. Indeed, for a given crop, the most relevant base should be sampled, which generally implies more than a single collection. The Generation Challenge Programme (GCP) has devoted much support to developing such core reference sets from the findings of collective studies. Composite (i.e. derived from several collections) core collections have been analyzed with molecular markers and reduced to potential core reference sets of 50–500 accessions depending on the crop. The materials must be transformed into genetic stocks that have been purified (homogeneous/stabilized) and roughly phenotyped to facilitate practical choices for comparative phenotyping studies. Furthermore, they must be publicly, quickly, and cheaply available. This is currently the case for all resources managed by CGIAR-hosted germplasm centers, which are best positioned to deal with the pressing constraints of intellectual property legislation and quarantine regulations.

From core to global diversity

The core reference set has diverse applications (Figure 2). It provides a representation of the major components of genetic structure, which any assessment must relate to for proper interpretation. It provides a means for entering the broader collection, using accurate attached passport data to establish correlations and guide further exploration. It helps assess donors of genes and alleles, by giving clues to phenotype comparability, sample structure descriptions, meaningful checks for breeders, and known extreme phenotypes. It helps dissect the genetic control of trait variation through contributions to panels formed for association studies, thus paving the way for further targeted diversity mining.

The core reference set is a flexible concept that welcomes updates and adjustments under close monitoring by the community. It is viewed as a process facilitating a practical trade-off between the wish to always include an absolute reference (here represented as the 'nucleus sample'), the wish to cover the broadest range of materials, and the importance of adjusting the materials in relation to the practical constraints or specific purpose of the study which they are used for. The choice of the materials can be guided by the genotypic and phenotypic information already accumulated.

A better understanding of core diversity is expected to encourage the use of broader ranging germplasm derived from existing *ex situ* collections or from new *in situ* analyses. Access to rare alleles will require renewed searches in large collections. Moreover, in many cases, populations of materials are still standing in and around fields, in evolving environments, with people caring for them. The new analytical power of 'ecological' genomics can now be used for *in situ* collection of information and materials, similarly to what is currently under way for sorghum and pearl millet in Africa [62,63,64•,65,66].

Conclusion

Genome studies applied to crop germplasm shed light on the role of selection, foundations, migrations, and introgressions on population patterns, genomic associations, and genic diversity. Thanks to the sharply declining cost

Tecan Ex. 2110, Page 4 of 7
QIAGEN Sciences, LLC v. Tecan Genomics - IPR2025-00029

of genotyping technologies, it is now possible to make surveys that can be equally broad and whole-genome oriented [67•], or targeted on specific genes of suspected function. The history and diversity of crops can then be analyzed as are those of human populations [68,69•,70•]. Such new information can efficiently foster those essential interactions – pioneered by Jack Harlan [71] – with the fields of archeology and ethnobotany so as to gain greater insight into domestication, while identifying the main historical benchmarks and biological drivers. Factors limiting the practical use of germplasm have clearly become tied to their proper phenotypic assessment. The use of shared core reference sets of materials can help the research community to focus studies and be more efficient. Materials specifically adapted to local constraints and uses will not all be present in reduced samples. Renewed sampling within and outside existing collections will still be necessary. The adaptive potential of these materials can also be grasped through accurate description of their environments of origin. The availability and quality of ecogeographical/passport information will be the key to a more ecological approach to germplasm management. Together, genome studies and molecular genetics will make the future of 'germplasm science' very exciting.

Acknowledgements

The authors are thankful to the Generation Challenge Programme (JCG, HDU, RKV), the Agropolis Foundation, France (JCG), the Indian Council of Agricultural Research (ICAR), and the Department of Biotechnology, Government of India (RKV) for sponsoring research on genetic diversity in their laboratories. Thanks are also due to Andreas Graner of the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany, for useful suggestions and discussion on this manuscript.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Hammer K, Arrowsmith N, Gladis T: **Agrobiodiversity with emphasis on plant genetic resources**. *Naturwissenschaften* 2003, **90**:241-250.

2. Upadhyaya HD, Furman BJ, Dwivedi SL, Udupa SM, Gowda CLL, Baum M, Crouch JH, Buhariwalla HK, Singh S: **Development of a composite collection for mining germplasm possessing allelic variation for beneficial traits in chickpea**. *Plant Genet Resour* 2006, **4**:13-19.

3. Holderegger R, Herrmann D, Poncet B, Gugerli F, Thuiller W, Taberlet P, Gielly L, Rioux D, Brodbeck S, Aubert S, Manel S: **Land ahead: using genome scans to identify molecular markers of adaptive relevance**. *Plant Ecol Divers* 2008, **1**:273-283.

4. Avise JC: **Phylogeography: retrospect and prospect**. *J Biogeogr* 2009, **36**:3-15.

5. Clark RM, Schweikert G, Toomajian C, Ossowski S, Zeller G, Shinn P, Warthmann N, Hu TT, Fu G, Hinds DA *et al.*: **Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana***. *Science* 2007, **317**:338-342.

6. Kim S, Plagnol V, Hu TT, Toomajian C, Clark RM, Ossowski S, Ecker JR, Weigel D, Nordborg M: **Recombination and linkage disequilibrium in *Arabidopsis thaliana***. *Nat Genet* 2007, **39**:1151-1155.

7. Nordborg M, Weigel D: **Next-generation genetics in plants**. *Nature* 2008, **456**:720-723.

8. • Ossowski S, Schneeberger K, Clark RM, Lanz C, Warthmann N, Weigel D: **Sequencing of natural strains of *Arabidopsis thaliana* with short reads**. *Genome Res* 2008, **18**:2024-2033.
This comprehensive study, after analyzing 15–25-fold coverage in Illumina sequencing-by-synthesis reads for the reference accession, Col-0, and two divergent strains, Bur-0 and Tsu-1, has revealed that the nuclear genomes of natural strains of *Arabidopsis thaliana* can differ by several percent in their sequence.

9. • Jarvis DI, Brown AHD, Cuong PH, Collado-Panduro L, Latournerie-Moreno L, Gyawali S, Tanto T, Sawadogo M, Mar I, Sadiki M *et al.*: **A global perspective of the richness and evenness of traditional crop-variety diversity maintained by farming communities**. *Proc Natl Acad Sci* 2008, **105**:5326-5331.
Data on the use of landraces and their number versus improved varieties were gathered from 27 crop species over 10 years in eight countries and five continents, at both farm and community levels. This study demonstrates that considerable crop diversity can be maintained on small farms in the form of traditional crop varieties adopting distinctly diverse strategies.

10. Gepts P: **Domestication as a long-term selection experiment**. *Plant Breed Rev* 2004, **24**:1-44.

11. Doebley JF, Gaut BS, Smith BD: **The molecular genetics of crop domestication**. *Cell* 2006, **127**:1309-1321.

12. Ross-Ibarra J, Morrell PL, Gaut BS: **Plant domestication, a unique opportunity to identify the genetic basis of adaptation**. *Proc Natl Acad Sci USA* 2007, **104**:8641-8648.

13. • Purugganan MD, Fuller DQ: **The nature of selection during plant domestication**. *Nature* 2009, **457**:843-848.
This paper reviews recent archeological and genetic studies that provide a clear picture of the selective pressures associated with crop origins and diversification. This combined view can increase insight into the nature of evolutionary factors during plant domestication.

14. Gregory TR: **Artificial selection and domestication: modern lessons from Darwin's enduring analogy**. *Evol Educ Outreach* 2009, **2**:5-27.

15. Glémin S, Bataillon T: **A comparative view of the evolution of grasses under domestication**. *New Phytologist* 2009, **183**:273-290.

16. Kilian B, Özkan H, Pozzi C, Salamini F: **Domestication of the Triticeae in the fertile crescent**. In *Genetics and Genomics of the Triticeae.* Edited by Feuillet C, Muehlbauer GJ. Springer; 2009:81-119.

17. Spooner DM, Nunez J, Trujillo G, del Rosario Herrera M, Guzman F, Ghislain M: **Extensive simple sequence repeat genotyping of potato landraces supports a major reevaluation of their gene pool structure and classification**. *Proc Natl Acad Sci USA* 2007, **104**:19398-19403.

18. Labate JA, Robertson LD, Baldo AM: **Multilocus sequence data reveal extensive departures from equilibrium in domesticated tomato (*Solanum lycopersicum* L.)**. *Heredity* 2009, **103**:257-267.

19. Kilian B, Özkan H, Walther A, Kohl J, Dagan T, Salamini F, Martin W: **Molecular diversity at 18 loci in 321 wild and 92 domesticate lines reveal no reduction of nucleotide diversity during *Triticum monococcum* (einkorn) domestication: implications for the origin of agriculture**. *Mol Biol Evol* 2007, **24**:2657-2668.

20. Kwak M, Gepts P: **Structure of genetic diversity in the two major gene pools of common bean (*Phaseolus vulgaris* L., *Fabaceae*)**. *Theor Appl Genet* 2009, **118**:979-992.

21. Perrier X, Bakry F, Carreel F, Jenny C, Horry JP, Lebot V, Hippolyte I: **Combining biological approaches to shed light on evolution of edible bananas**. *Ethnobot Res Appl* 2009, **7**:199-216.

22. Wright SI, Bi IV, Schroeder SG, Yamasaki M, Doebley JF, McMullen MD, Gaut BS: **The effects of artificial selection on the maize genome**. *Science* 2005, **308**:1310-1314.

23. • Burke JM, Burger JC, Chapman MA: **Crop evolution: from genetics to genomics**. *Curr Opin Genet Dev* 2007, **17**:525-532.
A well documented opinion forecasting that genome-wide analyses considering a large and essentially random collection of genes will result

Tecan Ex. 2110, Page 5 of 7
QIAGEN Sciences, LLC v. Tecan Genomics - IPR2025-00029

in the identification of agronomically important genes that would have otherwise been overlooked.

24. • Zhao Q, Thuillet AC, Uhlmann NK, Weber A, Rafalski JA, Allen SM, Tingey S, Doebley J: **The role of regulatory genes during maize domestication: evidence from nucleotide polymorphism and gene expression**. *Genetics* 2008, **178**:2133-2143.
Nucleotide diversity at 72 candidate genes in maize was investigated, highlighting 17 genes that were potential targets of selection during domestication. Domestication genes appear to be expressed on average at a significantly higher level than neutral genes in reproductive organs.

25. Kovach MJ, Sweeney MT, McCouch SR: **New insights into the history of rice domestication**. *Trends Genet* 2007, **23(11)**:578-587.

26. Sang T, Ge S: **The puzzle of rice domestication**. *J Integr Plant Biol* 2007, **49**:760-768.

27. Vaughan DA, Lu BR, Tomooka N: **The evolving story of rice evolution**. *Plant Sci* 2008, **174**:394-408.

28. Izawa T, Konishi S, Shomura A, Yano M: **DNA changes tell us about rice domestication**. *Curr Opin Plant Biol* 2009, **12**:185-192.

29. Zhang LB, Zhu Q, Wu ZQ, Ross-Ibarra J, Gaut BS, Ge S, Sang T: **Selection on grain shattering genes and rates of rice domestication**. *New Phytologist* 2009 doi: 10.1111/j.1469-8137.2009.02984.x.

30. • Takano-Kai N, Jiang H, Kubo T, Sweeney M, Matsumoto T, Kanamori H, Padhukasahasram B, Bustamante C, Yoshimura A, Doi K, McCouch S: **Evolutionary history of *GS3*, a gene conferring grain length in rice**. *Genetics* 2009, **182**:1323-31334.
An exemplary study in rice, reporting the cloning and characterization of *GS3*, a gene controlling grain length. Haplotype analysis within and around *GS3* traced the origin of the long-grain allele to a *japonica*-like ancestor and demonstrated introgression into the *indica* gene pool, although the long-grain trait is usually considered an *indica* feature.

31. Sang T: **Genes and mutations underlying domestication transitions in grasses**. *Plant Physiol* 2009, **149**:63-70.

32. Bourgis F, Guyot R, Gherbi H, Tailliez E, Amabile I, Salse J, Lorieux M, Delseny M, Ghesquière A: **Characterization of the major fragrance gene from an aromatic *japonica* rice and analysis of its diversity in Asian cultivated rice**. *Theor Appl Genet* 2008, **117**:353-368.

33. Kovach MJ, Calingacion MN, Fitzgerald MA, McCouch SR: **The origin and evolution of fragrance in rice (*Oryza sativa* L.)**. *Proc Natl Acad Sci USA* 2009, **106**:14444-14449.

34. •• McNally KL, Childs KL, Bohnert R, Davidson RM, Zhao K, Ulat VJ, Zeller G, Clark RM, Hoen DR, Bureau TE *et al.*: **Genomewide SNP variation reveals relationships among landraces and modern varieties of rice**. *Proc Natl Acad Sci USA* 2009, **106**:12273-12278.
The authors determined SNP variation in 100 Mb of the rice genome and describe the characterization of 160 000 nonredundant SNPs distributed across the entire genome of 20 rice varieties. Typical haplotype blocks in *Indica* varieties reach 200 kb. Shared SNPs among rice groups indicate introgression caused by breeding, often bearing quantitative trait loci (QTLs) for important traits, or historical out-crossing events. The number of SNPs appears sufficient for most foreseen applications, whereas the array of rice varieties must be broadened in order to allow recognition of minority varietal groups, to offer a representation of the wild progenitors, and ultimately to provide statistical power to take full advantage of the unprecedented level of genomic resolution.

35. Glaszmann JC: **Isozymes and classification of Asian rice varieties**. *Theor Appl Genet* 1987, **74**:21-30.

36. • Marcano M, Morales S, Hoyer MT, Courtois B, Risterucci AM, Fouet O, Pugh T, Cros E, Gonzalez V, Dagert M, Lanaud C: **A genomewide admixture mapping study for yield factors and morphological traits in a cultivated cocoa (*Theobroma cacao* L.) population**. *Tree Genet Genomes* 2009, **5**:329-337.
Two-hundred and fifty-seven Criollo cocoa individuals were analyzed with 92 microsatellite markers in order to make use of the extensive LD observed for this germplasm compartment in an earlier study. This modern cultivated population exhibited a wide range of variation and significant associations between markers and yield factors were identified.

37. • Lippman ZB, Cohen O, Alvarez JP, Abu-Abied M, Pekker I, Paran I, Eshed Y, Zamir D: **The making of a compound inflorescence in tomato and related nightshades**. *PLoS Biol* 2008, **6**:2424-2435.
A clearly presented model of transient sequential expression of two genes, *compound inflorescence* (*s*) followed by *anantha* (*an*), which promotes branch termination and flower initiation in sympodial plants. Gene *s*, responsible for a major portion of inflorescence variation in domesticated tomatoes, relates to a tomato mutant described a century ago.

38. Fournier-Level A, Lacombe T, Le Cunff L, Boursiquot J-M, This P: **Evolution of the *VvMybA* gene family, the major determinant of berry colour in cultivated grapevine (*Vitis vinifera* L.)**. *Heredity* advance online publication, 18 November 2009; doi:10.1038/hdy.2009.148.

39. • Komatsuda T, Pourkheirandish M, He C, Azhaguvel P, Kanamori H, Perovic D, Stein N, Graner A, Wicker T, Tagiri A *et al.*: **Six-rowed barley originated from a mutation in a homeodomain-leucine zipper I-class homeobox gene**. *Proc Natl Acad Sci USA* 2007, **104**:1424-1429.
The authors cloned the gene responsible for a six-rowed spike in barley (*vrs1*) and demonstrated that the six-row phenotype originated several times, through independent mutations of *Vrs1*.

40. Taketa S, Amano S, Tsujino Y, Sato T, Saisho D, Kakeda K, Nomura M, Suzuki T, Matsumoto T, Sato K *et al.*: **Barley grain with adhering hulls is controlled by an ERF family transcription factor gene regulating a lipid biosynthesis pathway**. *Proc Natl Acad Sci USA* 2008, **105**:4062-4067.

41. • Wang E, Wang J, Zhu X, Hao W, Wang L, Li Q, Zhang L, He W, Lu B, Lin H *et al.*: **Control of rice grain-filling and yield by a gene with a potential signature of domestication**. *Nat Genet* 2008, **40**:1370-1374.
The paper reports the isolation and functional analysis of a gene involved in grain incomplete filling (*GIF1*) that encodes a cell-wall invertase. This gene is required for carbon partitioning during early grain filling and was probably selected during domestication.

42. Shomura A, Izawa T, Ebana K, Ebitani T, Kanegae H, Konishi S, Yano M: **Deletion in a gene associated with grain size increased yields during rice domestication**. *Nat Genet* 2008, **40**:1023-1028.

43. Weng J, Gu S, Wan X, Gao H, Guo T, Su N, Lei C, Zhang X, Cheng Z, Guo X *et al.*: **Isolation and initial characterization of *GW5*, a major QTL associated with rice grain width and weight**. *Cell Res* 2008, **18**:1199-1209.

44. Konishi S, Ebana K, Izawa T: **Inference of the *japonica* rice domestication process from the distribution of six functional nucleotide polymorphisms of domestication-related genes in various landraces and modern cultivars**. *Plant Cell Physiol* 2008, **49**:1283-1293.

45. • de Alencar Figueiredo LF, Calatayud C, Dupuits C, Billot C, Rami JF, Brunel D, Perrier X, Courtois B, Deu M, Glaszmann JC: **Phylogeographic evidence of crop neodiversity in sorghum**. *Genetics* 2008, **179**:997-1008.
Sequence diversity was surveyed with respect to six candidate genes for grain quality among 194 sorghum accessions representative of cultivated species. Two of these genes, in particular *opaque-2*, displayed novel variations outside the area of origin of cultivated sorghum. These are derived from post-domestication mutations, suggesting that neo-diversity contributed to new adaptations for human uses.

46. Camus-Kulandaivelu L, Chevin LM, Tollon-Cordet C, Charcosset A, Manicacci D, Tenaillon MI: **Patterns of molecular evolution associated with two selective sweeps in the *Tb1*–*Dwarf8* region in maize**. *Genetics* 2008, **180**:1107-1121.

47. •• Fukao T, Harris T, Bailey-Serres J: **Evolutionary analysis of the *Sub1* gene cluster that confers submergence tolerance to domesticated rice**. *Ann Bot* 2009, **103**:143-150.
A comprehensive analysis of the variation in the *Sub1* gene cluster identified by positional cloning in rice, which is now massively transferred through marker-assisted selection to other varietal backgrounds. This is a small gene cluster that the authors resolve through phylogenetic analysis before looking at sequence variation between wild rice and several varietal groups. They relate the original distribution of the submergence tolerance haplotypes, localized between Orissa state, India, and Sri Lanka, to past human relationships between Eastern India and Sri Lanka, supported by biological similarities for human leucocyte antigens (HLA).

48. • Yahiaoui N, Brunner S, Keller B: **Rapid generation of new powdery mildew resistance genes after wheat domestication**. *Plant J* 2006, **47**:85-98.

Tecan Ex. 2110, Page 6 of 7
QIAGEN Sciences, LLC v. Tecan Genomics - IPR2025-00029

The paper reports that *Pm3* resistance alleles from bread wheat were all derived from the susceptible allele *Pm3CS* in agricultural ecosystems after the formation of hexaploid wheat.

49. •• Bhullar NK, Street K, Mackay M, Yahiaoui N, Keller B: **Unlocking wheat genetic resources for the molecular identification of previously undescribed functional alleles at the *Pm3* resistance locus**. *Proc Natl Acad Sci USA* 2009, **106**:9519-9524.

A large-scale allele-mining project through molecular characterization of wheat gene bank accessions. The authors first selected a set of 1320 bread wheat landraces from a database of 16 089 accessions, using the focused identification of germplasm strategy, which led to selecting accessions from 323 sites likely to be exposed to high selection pressure for powdery mildew. Using the known molecular structure of the *Pm3* locus, it was possible to isolate seven new resistances, doubling the known functional allelic diversity at this locus.

50. Allaby RG, Fuller DQ, Brown TA: **The genetic expectations of a protracted model for the origins of domesticated crops**. *Proc Natl Acad Sci USA* 2008, **105**:13982-13986.

51. Caicedo AL, Williamson SH, Hernandez RD, Boyko A, Fledel-Alon A, York TL, Polato NR, Olsen KM, Nielsen R, McCouch SR *et al.*: **Genome-wide patterns of nucleotide polymorphism in domesticated rice**. *PLoS Genet* 2007, **3**:e163.

52. Frankel OH: **Genetic perspective of germplasm conservation**. In *Genetic Manipulations: Impact on Man and Society.* Edited by Arber W, Llimensee K, Peacock WJ, Stralinger P. Cambridge University Press; 1984:161-170.

53. • Brown AHD: **The case for core collections**. In *The Use of Plant Genetic Resources.* Edited by Brown AHD, Frankel OH, Marshall DR, Williams JT. Cambridge, UK: Cambridge University Press; 1989:135-156.

A very thoughtful dissertation on the definition of core collections, their utility, constitution, and use, founded on population genetics methods.

54. Glaszmann JC, Mew T, Hibino H, Kim CK, Vergel de Dios-Mew TI, Vera Cruz CM, Notteghem JL, Bonman JM: **Molecular variation as a diverse source of disease resistance in cultivated rice**. In In *Rice Genetics III: Proceedings.* Edited by Khush GS. *Rice Genetics III: Proceedings* IRRI; 1995:460-465.

55. • Upadhyaya HD, Dwivedi SL, Baum M, Varshney RK, Udupa SM, Gowda CLL, Hoisington D, Singh S: **Genetic structure, diversity and allelic richness in composite collection and reference set in chickpea (*Cicer arietinum* L.)**. *BMC Plant Biol* 2008, **8**:106.

This study reflects the process followed in multi-partner projects supported by the Generation Challenge Programme. A composite collection of 2915 accessions, consisting of accessions from ICRISAT and ICARDA (International Center for Agricultural Research in the Dry Areas) was surveyed for 48 SSR loci, revealing a total of 1683 alleles. A reference set of 300 accessions was identified, making use of the existing ICRISAT mini core collection and the SSR data; it captured 1315 (78%) of the initial 1683 alleles, essentially leaving out some rare alleles. The reference set represents all major branches of the neighbor-joining tree that summarizes the variation in the composite collection.

56. Upadhyaya HD, Reddy LJ, Gowda CLL, Singh S: **Phenotypic diversity in cold-tolerant peanut (*Arachis hypogaea* L.) germplasm**. *Euphytica* 2009, **165**:279-291.

57. Upadhyaya HD, Ortiz R: **A mini core subset for capturing diversity and promoting utilization of chickpea genetic resources in crop improvement**. *Theor Appl Genet* 2001, **102**:1292-1298.

58. Upadhyaya HD, Pundir RPS, Dwivedi SL, Gowda CLL, Gopal Reddy V, Singh S: **Developing a mini core collection of sorghum [*Sorghum bicolor* (L.) Moench] for diversified utilization of germplasm**. *Crop Sci* 2009, **49**:1769-1780.

59. Zhu C, Gore M, Buckler ES, Yu J: **Status and prospects of association mapping in plants**. *Plant Genome* 2008, **1**:5-20.

60. Druka A, Druka I, Centeno AG, Li H, Sun Z, Thomas WTB, Bonar N, Steffenson BJ, Ullrich SE, Kleinhofs A *et al.*: **Towards systems genetic analyses in barley: integration of phenotypic, expression and genotype data into genenetwork**. *BMC Genet* 2008, **9**:73.

61. • Waugh R, Jannink JL, Muehlbauer GJ, Ramsay L: **The emergence of whole genome association scans in barley**. *Curr Opin Plant Biol* 2009, **12**:218-222.

This review presents the current state and the potential of whole-genome association mapping in barley. This crop is the focus of substantial research and stimulates forefront innovative developments.

62. Deu M, Sagnard F, Chantereau J, Calatayud C, Hérault D, Mariac C, Pham JL, Vigouroux Y, Kapran I, Traore PS *et al.*: **Niger-wide assessment of *in situ* sorghum genetic diversity with microsatellite markers**. *Theor Appl Genet* 2008, **116**:903-913.

63. Sagnard F, Barnaud A, Deu M, Barro C, Luce C, Billot C, Rami JF, Bouchet S, Dembélé D, Pomiès V *et al.*: **Synthèse analyse multiéchelle de la diversité génétique des sorghos: compréhension des processus évolutifs pour la conservation *in situ***. *Cah Agric* 2008, **17**:114-121.

64. • Barnaud A, Deu M, Garine E, Chantereau J, Justin B, Koïda OE, McKey D, Joly HI: **A weed–crop complex in Sorghum: the dynamics of genetic diversity in a traditional farming system**. *Am J Bot* 2009, **96**:1869-1879.

An interesting picture of microevolution in an agricultural environment. The authors developed a multidisciplinary approach, involving both biologists and social scientists, to investigate the dynamics of genetic diversity of a sorghum weed–crop complex in a village of Duupa farmers in northern Cameroon. Morphological and SSR data are congruent with farmers' taxonomy and confirm the introgressed status of intermediate weedy types. Farmers actively select against these morphotypes, but several practices unconsciously favor gene flow.

65. Bezançon G, Pham JL, Deu M, Vigouroux Y, Sagnard F, Mariac C, Kapran I, Mamadou A, Gérard B, Ndjeunga J, Chantereau J: **Changes in the diversity and geographic distribution of cultivated millet (Pennisetum glaucum (L.) R. Br.) and sorghum (Sorghum bicolor (L.) Moench) varieties in Niger between 1976 and 2003**. *Genet Resour Crop Evol* 2009, **56**:223-236.

66. Saïdou AA, Mariac C, Luong V, Pham JL, Bezançon G, Yves Vigouroux Y: **Association studies identify natural variation at *PHYC* linked to flowering time and morphological variation in pearl millet**. *Genetics* 2009, **182**:899-910.

67. • Varshney RK, Nayak SN, May GD, Jackson SA: **Next generation sequencing technologies and their implications for crop genetics and breeding**. *Trends Biotechnol* 2009, **27**:522-530.

A timely evaluation of a major trend in biology that will impact plant research. Next-generation sequencing technologies make it possible to sequence hundreds or even thousands of related genomes to sample genetic diversity within and between germplasm pools. Several important areas are outlined, which will enhance progress in crop genetics and breeding, leading to crop improvement.

68. Auton A, Byrc K, Boyko AR, Lohmueller KE, Novembre J, Reynolds A, Indap A, Wright MH, Degenhardt J, Gutenkunst RN *et al.*: **Global distribution of genomic diversity underscores rich complex history of continental human populations**. *Genome Res* 2009, **19**:795-803.

69. • Pickrell JK, Coop G, Novembre J, Kudaravalli S, Li J, Absher D, Srinivasan B, Barsh GS, Myers RM, Feldman MW, Prichard JK: **Signals of recent positive selection in a worldwide sample of human populations**. *Genome Res* 2009, **19**:826-837.

This paper reports an extension of genome-wide scans for recent positive selection in humans, leading to new insights into mechanisms underlying the extensive phenotypic diversity in our species. The geographic distributions of known selective sweeps are refined, several examples of previously unrecognized candidate selection targets are presented, and several examples of local adaptation between geographically close populations are highlighted.

70. • Hellenthal G, Auton A, Falush D: **Inferring human colonization history using a copying model**. *PLoS Genet* 2008, **4**:e1000078.

A statistical approach is presented that uses SNP data to identify sharing of chromosomal segments between populations and to reconstruct a detailed colonization scenario. It is applied to a broad range of human populations and reveals novel details on genetic interactions among populations. This will inspire similar studies in crops, which should disclose parallel patterns given the close interaction between humans and crops.

71. Qualset CO: **Jack R. Harlan (1917–1988): Plant explorer, Archaeobotanist, Geneticist, and Plant Breeder**. http://harlanii.ucdavis.edu/harlan_files/harlan.pdf.

Tecan Ex. 2110, Page 7 of 7
QIAGEN Sciences, LLC v. Tecan Genomics - IPR2025-00029