Crop evolution: from genetics to genomics

John M Burke, Jutta C Burger and Mark A Chapman

The advent of the genomics age has greatly facilitated the study of crop evolution. While full-scale genome sequencing projects are underway for just a handful of crop plants, recent years have witnessed a tremendous increase in the availability of DNA sequence data for virtually all major crops. Such resources have bolstered 'traditional' genetic approaches such as QTL mapping and candidate gene-based association studies. They have also allowed us to undertake genome-wide analyses in which we simultaneously consider the importance of a large and essentially random collection of genes. These sorts of analyses promise a more or less unbiased view of the genetic basis of crop evolution and will probably result in the identification of agronomically important genes that would have otherwise been overlooked.

Addresses
University of Georgia, Department of Plant Biology, Miller Plant Sciences Building, Athens, GA 30602, USA

Corresponding author: Burke, John M (jmburke@uga.edu)

Current Opinion in Genetics & Development 2007, **17**:525–532

This review comes from a themed issue on
Genomes and evolution
Edited by Susan Wessler, Kelly Dawe and Jim Leebens-Mack

Available online 22nd October 2007

0959-437X/$ – see front matter

DOI 10.1016/j.gde.2007.09.003

Introduction

Over the years, the evolution of crop plants has been a topic of great interest to crop scientists and evolutionary biologists alike. The reasons for this are simple. From an agricultural perspective, elucidation of the genetic basis of traits that make for a desirable crop plant has the potential to facilitate ongoing plant breeding efforts [1,2]. From an evolutionary perspective, crop evolution has the potential to shed light on basic processes such as the phenotypic and genetic responses of populations to long-term directional selection, the genetic consequences of recent selective sweeps, and/or the limitations imposed by genetic architecture on the response to selection (e.g. references [3–5]). Crop evolution thus serves as a useful model for investigating the molecular basis of adaptive trait evolution [6,7].

Following their initial domestication, which involved wholesale phenotypic changes in a suite of traits collectively known as the 'domestication syndrome' (Figure 1) [8,9], virtually all major crop lineages have experienced more recent, intensive selection on a variety of agronomic traits. These include increased yield, improved nutritional value, and resistance to various abiotic and biotic stresses. Thus, crop evolution can be viewed as having occurred in two major phases: the initial period of domestication and a subsequent (and ongoing) period of improvement.

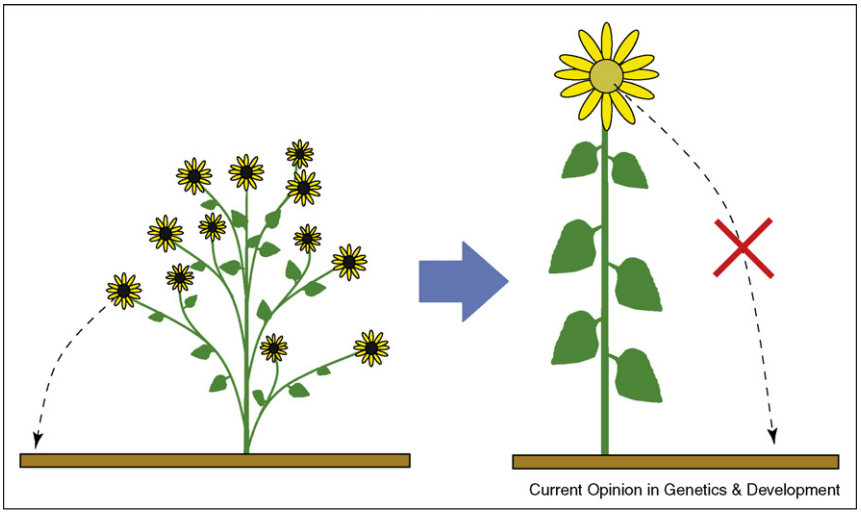
While the transformation of wild plant species into useful crops has been the subject of genetic studies dating back decades (e.g. references [10–12]), it was not until modern molecular tools became available that our understanding of the genetic basis of crop evolution really took off. In recent years, the pace of such research has accelerated in parallel with the development of genomic resources in an increasing number of crop lineages. In this review, we summarize our current knowledge of the genetics of crop evolution and discuss advances that have been made possible by the advent of the genomics age.

Crop genome resources: where are we now?

The genomics age in plants was ushered in by the publication of the *Arabidopsis* genome sequence in 2000 [13]; however, the genome sequences of only two economically important plants (rice and poplar) [14–16] have been published since then. Full-scale genome sequencing projects are currently underway for only a small number of crop species [17•], partly because of the unwieldy size and highly repetitive nature of many crop genomes (e.g. references [18–20]). While great strides have been made in developing strategies for targeting the non-repetitive, gene-rich fractions (i.e. the so-called 'gene space') of genomes for sequencing [21–24], large-scale genome sequencing projects in crop plants are still relatively few and far between. Thus, while genome sequences have the potential to facilitate a wide range of research endeavors, ranging from gene identification to comparative analysis of genome structure and evolution, such resources simply do not exist yet for the vast majority of crops.

To the extent that gene content, order, and function are conserved across taxa, genome sequences derived from one taxon can be brought to bear on research problems in other taxa. For example, the rice genome sequence has been used to facilitate map-based cloning in wheat and barley [25]. Unfortunately, while comparative analyses have revealed extensive co-linearity between closely related taxa [26–28], the length of conserved regions decreases dramatically with increasing evolutionary distance [29–31].

Figure 1



Schematic representation of the 'domestication syndrome' in sunflower. In general terms, domestication has resulted in the production of a wide variety of crops that share a number of traits, including increased seed or fruit size, more determinate growth and/or flowering, increased apical dominance (i.e. reduced branching), suppression of natural seed dispersal, the loss of seed dormancy, and (if applicable) a loss of self-incompatibility.

Thus, the utility of any particular genome sequence is relatively limited when it comes to dissecting the genetic changes that occurred during the evolution of a more or less unrelated crop.

In contrast to full-scale genome projects, virtually all major crops have been the subject of gene discovery efforts via expressed sequence tag (EST) sequencing. Such efforts have resulted in the production of increasingly comprehensive 'gene catalogs' for the targeted species.

These EST collections represent a rich source of both molecular markers and candidate genes for downstream analyses. Moreover, depending on the sampling strategy employed, such sequence collections can themselves be directly subjected to evolutionary analyses aimed at identifying genes that experienced selection during crop evolution [32].

Genetic maps, linkage disequilibrium, and gene identification

Our first detailed insights into the genetics of crop evolution came from quantitative trait locus (QTL) mapping studies of domestication-related traits in mapping populations derived from crop × progenitor crosses [33–38]. While the increased availability of molecular markers stemming from EST and other sequencing projects has greatly facilitated QTL mapping efforts, linkage-based approaches of this sort are still relatively limited in terms of resolution, typically resulting in the localization of genes of interest to intervals that often span 5–10 (or more) centimorgans, and which may include hundreds of genes. Nonetheless, such studies have provided a great deal of insight into the genetic architecture of crop evolution. What we have learned is that, with rare exceptions, plant domestication has involved a relatively small number of genetic changes, each of which had relatively major phenotypic effects (e.g. reviewed in reference [39]; but see references [33,40•]). Such analyses have also served as a jumping off point for the positional cloning and characterization of a handful of genes underlying domestication-related traits [41–45].

A complementary approach for dissecting genetically complex traits is association mapping. Also known as linkage disequilibrium (LD) mapping, this general approach was initially developed for use in human genetics [46,47], where the production of experimental populations via controlled matings is not an option. Rather, association mapping involves correlating molecular variation with phenotypic variation in a population consisting of a diverse assemblage of individuals. Because such populations typically reflect many generations of historical recombination, LD (i.e. the non-random association of alleles between loci) is much lower than in the family-based mapping populations upon which traditional QTL approaches are based. This much lower level of LD means that association-based approaches can provide much higher levels of resolution, in some cases allowing for the mapping of functional variation to the level of one or a few genes (e.g. references [48•,49,50]). Unfortunately, this high level of resolution is somewhat of a double-edged sword, in that low LD makes it much more difficult to detect genotype–phenotype correlations in the first place. Most

association-based approaches in plants have thus relied on the *a priori* identification of candidate genes that are then tested for an association with a trait of interest, though genome-wide association analyses are possible in study systems with sufficient genomic resources [51•].

More recently, maize researchers have developed an approach known as nested association mapping (NAM), which combines the strengths of linkage analyses and association mapping [52••]. In short, NAM involves the production of multiple recombinant inbred line (RIL) mapping populations derived from a set of crosses between a common parent and a diverse set of individuals (Figure 2a). The founding lineages are then subjected to high-density single-nucleotide polymorphism (SNP) genotyping or, in the extreme, full genome sequencing. The individual RILs are then genotyped for a set of 'framework' markers that are evenly distributed across the genome, and for which the common parent harbors a rare allele. The balance of the genotypic or sequence information can then be 'projected' onto these lines based on the framework marker data (i.e. an individual carrying the alleles of one parent at an adjacent pair of framework markers is inferred to be carrying the SNP variants of that parent at the intervening loci; Figure 2b). Thus, the framework markers allow for the tracking of chromosomal segments and efficient 'genotyping' of the intervening loci while the limited LD and high levels of variation within common intervals across the diverse founding lineages allow for greatly improved mapping resolution based on the inferred genotypes. Perhaps the biggest advantage of this approach is that the extrapolation of genotypic (or sequence) data from the diverse founding lineages to the multitude of RILs allows for the execution of a high-resolution, genome-wide scan for associations between molecular polymorphisms and phenotypic traits while minimizing the amount of actual genotyping that needs to be performed. The genome-wide nature of this approach also obviates the need for *a priori* candidate gene identification. Another clear advantage is that potential genetic background effects are controlled for by (1) the shuffling of the parental genomes during RIL production and (2) the joint analysis of all RILs across all crosses. Unfortunately, while NAM promises unprecedented power for the genetic dissection of complex traits, this approach clearly requires a tremendous upfront investment, and is thus not currently feasible for the majority of crop species. Moreover, it remains to be seen whether or not NAM will prove to be the best way forward in other taxa as next-generation technologies reduce the cost of genome sequencing to the point at which whole-genome association studies can be carried out with robust sample sizes.

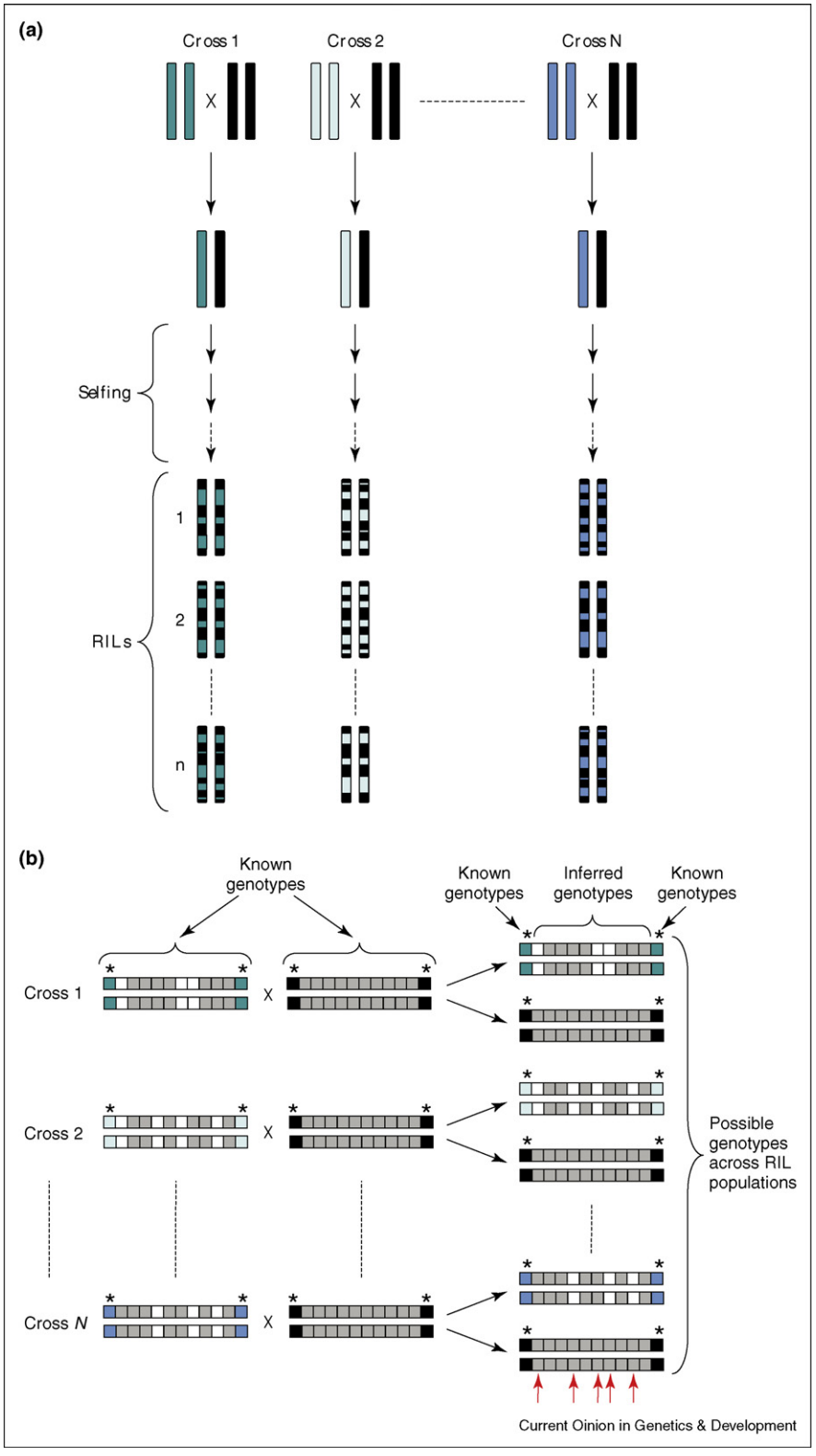
Evolutionary analyses and the signature of selection

An alternative method for identifying genes involved in the evolution of crop plants is to perform a large-scale screen for loci that bear the so-called 'signature of selection.' While crops generally experience a major population bottleneck during their domestication, resulting in a potentially major loss of genetic diversity, these demographic effects (along with the effects of such processes as migration and inbreeding) are manifested throughout the genome (Figure 3). By contrast, selection acts in a locus-specific manner. Selective sweeps (i.e. periods of intense selection during which a favorable allele is 'swept' to fixation) should therefore dramatically reduce genetic variation in the vicinity of the target locus while having little or no effect on diversity elsewhere in the genome ([53]; but see reference [54]). In principle, this distinction between locus-specific and genome-wide effects should allow for the identification of genes that were targeted by selection during crop evolution. Of course, the ability to identify individual genes that were targeted by selection depends on the structure of LD across the genome of the focal taxon. Nonetheless, the identification of genes that experienced recent selective sweeps provides a means for identifying agronomically important genes without knowing anything about their functions and/or phenotypic effects. While this sort of work is potentially labor-intensive and needs to be coupled with downstream analyses aimed at linking genotypic changes with a particular phenotype (e.g. bioinformatic analyses, genetic mapping, and/or reverse genetic approaches), it benefits from not being influenced by pre-conceived notions about the types of genes and/or traits that are likely to be most important.

Although the potential utility of these sorts of evolutionary analyses for gene identification has long been recognized, the application of this approach has been limited by the lack of sufficient numbers of suitable molecular markers in many study systems. With the increased availability of large bodies of genomic data, however, this approach can now be directed toward analyzing patterns of allelic variation across large numbers of loci in a growing number of taxa. In the first application of this approach to crop evolution, Vigouroux *et al.* [55] identified 15 genes that were putatively selected during maize evolution, several of which co-localized with previously mapped QTLs underlying crop-related traits. In a subsequent study, Wright *et al.* [56•] analyzed SNP diversity for a large collection of genes and likewise found that candidates for selectively important genes tend to cluster near QTLs that contribute to the phenotypic differences between teosinte and maize.

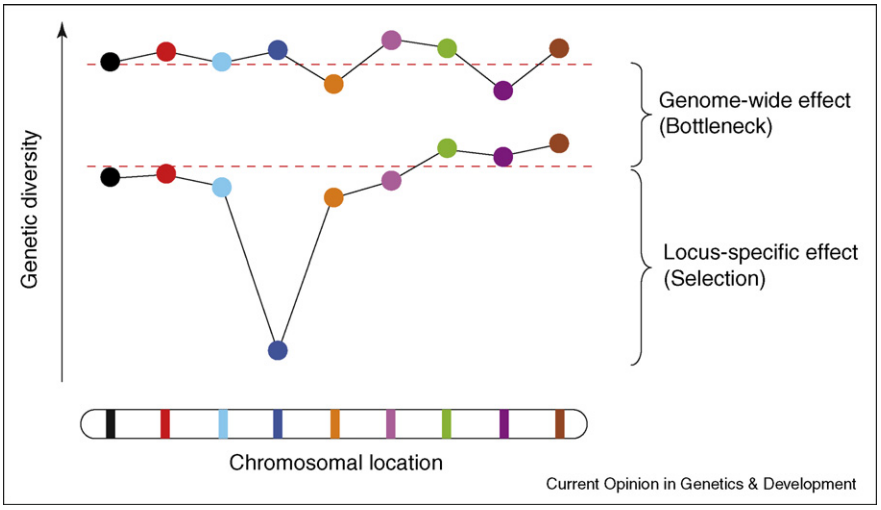
One particularly intriguing possibility is that, with the proper sampling, evolutionary analyses of this sort have the potential to distinguish between genes that experienced selection during domestication and those that experienced selection during the subsequent period of improvement (Figure 4). In the case of domestication-related genes, one would expect to see an extreme loss of

Figure 2



Schematic diagram of the nested association mapping (NAM) design. **(a)** A set of *N* recombinant inbred line (RIL) populations, each consisting of *n* individuals, is produced by crossing a diverse set of parents (indicated by colored chromosomes) against a common parental line (black chromosomes). The *N* founding lineages and the common parent are then subjected to high-density genotyping or, in the extreme, full genome sequencing. The RILs are then genotyped for a set of 'framework' markers at which the common parent harbors a rare allele, thereby allowing

Figure 3



The identification of loci under selection during crop evolution using a genomic screen approach. The colored bars along the 'chromosome' at the bottom represent loci, whereas the colored circles indicate the level of genetic diversity at each locus across wild (above) and crop (below) populations. The genome-wide effect of the population bottleneck associated with domestication is evident for all loci, whereas the loss of diversity due to selection is evident only for the blue locus.

Figure 4

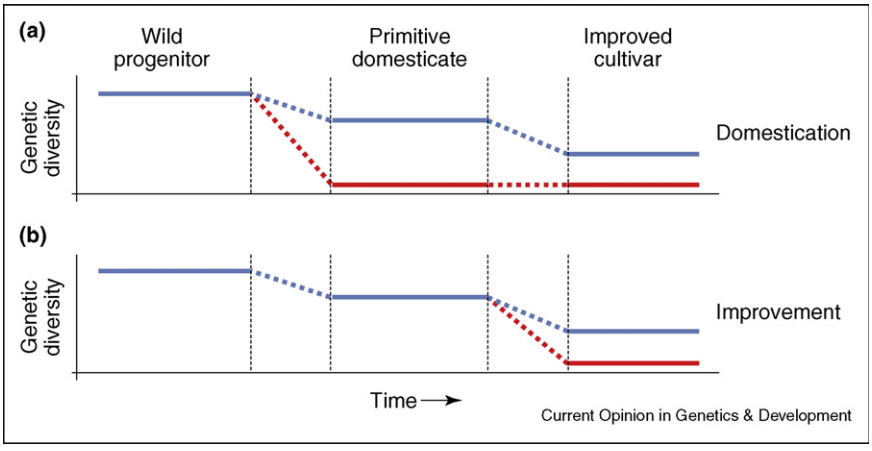


Illustration of the expected loss of genetic diversity at neutral and selected loci during crop evolution. The blue line represents the neutral expectation, which exhibits decreases corresponding to the domestication bottleneck as well as possible bottlenecks during the subsequent period of breeding and improvement. By contrast, the red line depicts the relative change in genetic diversity at loci subjected to strong and consistent directional selection during domestication **(a)** or improvement **(b)**.

diversity (as compared to neutral loci) in even the most primitive cultivars (Figure 4a). In the case of improvement-related genes, the situation may be somewhat more complex. In general, improvement-related genes would be expected to show a similar, selectively induced loss of diversity across the primitive-improved transition (Figure 4b). In the case of genes involved in crop diversification and/or adaptation to local growing environments, however, diversifying selection could result in the fixation of different alleles in different lineages,

(Figure 2 Legend Continued) for the tracking of chromosomal segments. Panel **(b)** depicts a single genomic segment flanked by a pair of framework markers (indicated with asterisks). Once the framework marker genotypes are known, full genotypic information for the segment that they flank can be 'projected' from the parental lines onto each individual RIL. Thus, the full genotypes of each RIL can be inferred with minimal genotyping. Polymorphic sites, such as those indicated by the red arrow for these three crosses, can then be tested for associations with phenotypic traits of interest. See text for further details. (After reference [52••].)

thereby preventing their loss. As such, these genes might retain diversity across lines, resulting in their being missed by approaches aimed at identifying selectively important genes based on an overall loss of diversity.

To date, two studies have successfully used population genetic data to make inferences about the timing of selection on specific genomic regions during crop evolution. In one case, SSR data were used to identify a region of the sunflower genome that was the target of one or more selective sweeps during the post-domestication era, presumably as a result of selection on seed oil-related characters [57•]. In the other case, Yamasaki *et al.* [58•] used a large sequence dataset to identify 8 genes that showed evidence of selection during the domestication of maize and another 10 genes that showed evidence of selection during improvement, with 4 of each type showing significant evidence of selection in two different analyses. This distinction between domestication-related and improvement-related genes is far more than an academic curiosity, as knowledge of when a particular gene experienced selection can guide practical efforts aimed at the discovery of novel alleles for use in modern breeding programs. More specifically, improvement-related genes will still be segregating for functional variation in landraces, whereas for domestication-related genes one would probably have to look to the wild progenitor (or other relatives) for new alleles.

Conclusions

Recent years have witnessed a dramatic increase in the availability of genomic sequence data for a growing list of crop plants. Such resources have allowed us to move from targeted analyses of major-effect genes with an obvious role in the wild-crop transformation to genome-wide analyses that allow us to simultaneously consider a large and arbitrarily chosen collection of genes. Whether such analyses take the form of genome-wide association mapping, nested association mapping, or genomic screens for the signature of selection, they have the potential to revolutionize our understanding of the genetic basis of crop evolution. Indeed, rather than limiting ourselves to the analysis of genes that *should* be important (as is the case with *a priori* identification of candidate genes), the analysis of an essentially random set of genes results in a more or less unbiased view of the genetic basis of crop domestication and improvement. This sort of work thus has the potential to result in the identification of agronomically important genes that otherwise might have been overlooked. As new and more efficient sequencing and genotyping technologies come on line, the potential of such approaches will only increase.

Acknowledgements

The writing of this paper was supported partly by grants from the National Science Foundation Plant Genome Research Program (DBI-0332411) and the United States Department of Agriculture (03-35300-13104 and 06-39454-17637).

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. Dudley JW: **Molecular markers in plant improvement: manipulation of genes affecting quantitative traits**. *Crop Sci* 1993, **33**:660-668.

2. Tanksley SD, McCouch SR: **Seed banks and molecular maps: unlocking genetic potential from the wild**. *Science* 1997, **277**:1063-1066.

3. Kalisz S, Kramer EM: **Variation and constraint in plant evolution and development.** *Heredity* (2007), doi:10.1038/sj.hdy.6800939.

4. Laurie CC, Chasalow SD, LeDeaux JR, McCarroll R, Bush D, Hauge B, Lai C, Clark D, Rocheford T, Dudley JW: **The genetic architecture of response to long-term artificial selection for oil concentration in the maize kernel**. *Genetics* 2004, **168**:2141-2155.

5. Wang R-L, Stec A, Hey J, Lukens L, Doebley J: **The limits of selection during maize domestication**. *Nature* 1999, **398**:236-239.

6. Hancock JF: **Contributions of domesticated plant studies to our understanding of plant evolution**. *Ann Bot* 2005, **96**:953-963.

7. 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.

8. Hammer K: **Das Domestikationssyndrom**. *Genet Resour Crop Evol* 1984, **32**:11-34.

9. Harlan JR: *Crops and Man*. Madison, WI: American Society of Agronomy; 1992.

10. Beadle GW: **Teosinte and the origin of maize**. *J Hered* 1939, **30**:245-247.

11. Harlan JR, De Wet JMJ, Price EG: **Comparative evolution of cereals**. *Evolution* 1973, **27**:311-325.

12. Langham DG: **The inheritance of intergeneric differences in *Zea-Euchlaena* hybrids**. *Genetics* 1940, **25**:88-107.

13. The Arabidopsis Genome Initiative: **Analysis of the genome sequence of the flowering plant** *Arabidopsis thaliana*. *Nature* 2000, **408**:796-815.

14. Goff SA, Ricke D, Lan T-H, Presting G, Wang R-L, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H *et al.*: **A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*)**. *Science* 2002, **296**:92-100.

15. Tuskan GA, DiFazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A *et al.*: **The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray)**. *Science* 2006, **313**:1596-1604.

16. Yu J, Hu S, Wang J, Wong GK-S, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X *et al.*: **A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*)**. *Science* 2002, **296**:79-92.

17. • Paterson AH: **Leafing through the genomes of our major crop plants: strategies for capturing unique information**. *Nat Rev Genet* 2006, **7**:174-184.
The author provides a thoughtful discussion of genome sequencing strategies as they relate to crop plants, as well as a discussion of factors to consider when selecting future targets for sequencing.

18. Arumuganathan K, Earle ED: **Nuclear DNA content of some important plant species**. *Plant Mol Biol Rep* 1991, **9**:208-218.

19. Li W, Zhang P, Fellers JP, Friebe B, Gill BS: **Sequence composition, organization, and evolution of the core Triticeae genome**. *Plant J* 2004, **40**:500-511.

20. Meyers BC, Tingey SV, Morgante M: **Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome**. *Genome* 2001, **11**:1660-1676.

21. Peterson DG, Schulze SR, Sciara EB, Lee SA, Bowers JE, Nagel A, Jiang N, Tibbitts DC, Wessler SR, Paterson AH: **Integration of Cot analysis, DNA cloning, and high-throughput sequencing facilitates genome characterization and gene discovery**. *Genome Res* 2002, **12**:795-807.

22. Peterson DG, Wessler SR, Paterson AH: **Efficient capture of unique sequences from eukaryotic genomes**. *Trends Genet* 2002, **18**:547-550.

23. Rabionwicz PD, Schutz K, Dedhia N, Yordan C, Parnell LD, Stein L, McCombie WR, Martienssen RA: **Differential methylation of genes and retrotransposons facilitates shotgun sequencing of the maize genome**. *Nat Genet* 1999, **23**:305-308.

24. Whitelaw CA, Barbazuk WB, Pertea G, Chan A, Cheung F, Lee Y, Zheng L, van Heeringen S, Karamycheva S, Bennetzen JL *et al.*: **Enrichment of gene-coding sequences in maize by genome filtration**. *Science* 2003, **302**:2118-2120.

25. Yan L, Fu D, Li C, Blechl A, Tranquilli G, Bonafede M, Sanchez A, Valarik M, Yasuda S, Dubcovsky J: **The wheat and barley vernalization gene is an orthologue of FT**. *Proc Natl Acad Sci USA* 2006, **103**:19581-19586.

26. Bennetzen JL: **Comparative sequence analysis of plant nuclear genomes: microcolinearity and its many exceptions**. *Plant Cell* 2000, **12**:1021-1029.

27. Parkin IA, Lydiate DJ, Trick M: **Assessing the level of colinearity between *Arabidopsis thaliana* and *Brassica napus* for *A. thaliana* chromosome 5**. *Genome* 2002, **45**:356-366.

28. Paterson AH, Bowers JE, Burow MD, Draye X, Elsik CG, Jiang CX, Katsar CS, Lan TH, Lin YR, Ming R *et al.*: **Comparative genomics of plant chromosomes**. *Plant Cell* 2000, **12**:1523-1540.

29. Liu H, Sachidanandam R, Stein L: **Comparative genomics between rice and *Arabidopsis* shows scant colinearity in gene order**. *Genome Res* 2001, **11**:2020-2026.

30. Martin WJ, McCallum J, Shigyo M, Jakse J, Kuhl JC, Yamane N, Pither-Joyce M, Gokce AF, Sink KC, Town CD *et al.*: **Genetic mapping of expressed sequences in onion and in silico comparisons with rice show scant colinearity**. *Mol Genet Genomics* 2005, **274**:197-204.

31. Timms L, Jimenez R, Chase M, Lavelle D, McHale L, Kozik A, Lai Z, Heesacker A, Knapp S, Rieseberg L *et al.*: **Analyses of synteny between *Arabidopsis thaliana* and species in the Asteraceae reveal a complex network of small syntenic segments and major chromosomal rearrangements**. *Genetics* 2006, **173**:2227-2235.

32. Church S, Livingstone KD, Lai Z, Kozik A, Knapp SJ, Michelmore RW, Rieseberg LH: **Using variable rate models to identify genes under selection in sequence pairs: their validity and limitations for EST sequences**. *J Mol Evol* 2007, **64**:171-180.

33. Burke JM, Tang S, Knapp SJ, Rieseberg LH: **Genetic analysis of sunflower domestication**. *Genetics* 2002, **161**:1257-1267.

34. Doebley J, Stec A: **Genetic analysis of the morphological differences between maize and teosinte**. *Genetics* 1991, **129**:285-295.

35. Doebley J, Stec A, Wendel J, Edwards M: **Genetic and morphological analysis of a maizeteosinte F2 population: implications for the origin of maize**. *Proc Natl Acad Sci USA* 1990, **87**:9888-9892.

36. Doebley JF, Stec A: **Inheritance of morphological differences between maize and teosinte: comparison of results for two F_2 populations**. *Genetics* 1993, **134**:559-570.

37. Koinange EMK, Singh SP, Gepts P: **Genetic control of the domestication syndrome in common bean**. *Crop Sci* 1996, **36**:1037-1045.

38. Xiong LX, Liu KD, Dai XK, Xu CG, Zhang QF: **Identification of genetic factors controlling domestication-related traits of rice using an F_2 population of a cross between *Oryza sativa* and *O. rufipogon***. *Theor Appl Genet* 1999, **98**:243-251.

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

40. • Wills DM, Burke JM: **QTL analysis of the early domestication of sunflower**. *Genetics* 2007, **176**:2589-2599.
This paper, along with reference [33], serves as a counterpoint to the widely held view that domestication is driven by a relatively small number of genetic changes, each of major effect.

41. Doebley JF, Stec A, Hubbard L: **The evolution of apical dominance in maize**. *Nature* 1997, **386**:485-488.

42. Frary A, Nesbitt TC, Grandillo S, van der Knaap E, Cong B, Liu J, Meller J, Elber R, Alpert KB, Tanksley SD: ***fw2.2*: a quantitative trait locus key to the evolution of tomato fruit size**. *Science* 2000, **289**:85-88.

43. Konishi S, Izawa T, Lin SY, Ebana K, Fukuta Y, Sasaki T, Yano M: **An SNP caused loss of seed shattering during rice domestication**. *Science* 2006, **312**:1392-1396.

44. Li C, Zhou A, Sang T: **Rice domestication by reducing shattering**. *Science* 2006, **311**:1936-1939.

45. Wang H, Nussbaum-Wagler T, Li B, Zhao Q, Vigouroux Y, Faller M, Bomblies K, Lukens L, Doebley JF: **The origin of naked grains of maize**. *Nature* 2005, **436**:714-719.

46. Lander ES, Schork NJ: **Genetic dissection of complex traits**. *Science* 1994, **265**:2037-2048.

47. Ricsch N, Merikangas K: **The future of genetic studies of complex human diseases**. *Science* 1996, **273**:1516-1517.

48. • Olsen KM, Caicedo AL, Polato N, McClung A, McCouch SR, Purugganan MD: **Selection under domestication: evidence for a sweep in the rice *Waxy* genomic region**. *Genetics* 2006, **173**:975-983.
This paper reports an association between amylose content and alleles at the *waxy* locus in rice and provides an evolutionary analysis of the genomic region harboring this gene. Particularly notable is the conclusion that selective pressures associated with crop domestication can exceed by one to two orders of magnitude those observed for genes under even strong selection in natural systems.

49. Palaisa K, Morgante M, Tingey S, Rafalski A: **Long-range patterns of diversity and linkage disequilibrium surrounding the maize *Y1* gene are indicative of an asymmetric selective sweep**. *Proc Natl Acad Sci USA* 2004, **101**:9885-9890.

50. Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler ES: ***Dwarf8* polymorphisms associate with variation in flowering time**. *Nature Genet* 2001, **28**:286-289.

51. • Aranzana MJ, Kim S, Zhao K, Bakker E, Horton M, Jakob K, Lister C, Molitor J, Shindo C, Tang C *et al.*: **Genome-wide association mapping in *Arabidopsis* identifies previously known flowering time and pathogen resistance genes**. *PLoS Genet* 2005, **1**:e60 doi: 10.1371/journal.pgen.0010060.
This paper reports the results of the first genome-wide association analysis ever performed in plants. Although this approach suffers from a high false-positive rate, it obviates the need for *a priori* identification of candidate genes.

52. •• Yu J, Holland JB, McMullen MD, Buckler ES: **Genetic Design and Statistical Power of Nested Association Mapping in Maize.** *Genetics*, in press.
The authors of this paper describe the genetic and statistical properties of the nested association mapping (NAM) strategy that is currently being implemented in maize. This approach combines the strength of linkage and association mapping, and thus has the potential to revolutionize the genetic analysis of complex traits.

53. Maynard-Smith J, Haigh J: **The hitch-hiking effect of a favorable gene**. *Genet Res* 1974, **23**:23-35.

54. Innan H, Kim Y: **Pattern of polymorphism after strong artificial selection in a domestication event**. *Proc Natl Acad Sci USA* 2004, **101**:10667-10672.

55. Vigouroux Y, Jaqueth JS, Matsuoka Y, Smith OS, Beavis WD, Smith JSC, Doebley J: **Rate and pattern of mutation at microsatellite loci in maize**. *Mol Biol Evol* 2002, **19**:1251-1260.

56. • 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.
This paper reports the first large-scale analysis of SNP diversity for evidence of selection during the evolution of a crop plant. The authors

concluded that 2–4% of genes experienced selection during the domestication and/or improvement of maize.

57. Burke JM, Knapp SJ, Rieseberg LH: **Genetic consequences of**
• **selection during the evolution of cultivated sunflower**. *Genetics* 2005, **171**:1933-1940.
This paper and reference [58•] were the first to use population genetic data to elucidate the timing of selection on specific genomic regions during crop evolution, providing proof-of-concept for the idea that it should be possible to distinguish genes that experienced selection during domestication from those that were under selection during improvement.

58. Yamasaki M, Tenaillon MI, Bi IV, Schroeder SG, Sanchez-
• Villeda H, Doebley JF, Gaut BS, McMullen MD: **A large-scale screen for artificial selection in maize identifies candidate agronomic loci for domestication and crop improvement**. *Plant Cell* 2005, **17**:2859-2872.
See annotation for reference [57•].