

Plant molecular diversity and applications to genomics

Edward S Buckler IV* and Jeffrey M Thornsberry

Surveys of nucleotide diversity are beginning to show how genomes have been shaped by evolution. Nucleotide diversity is also being used to discover the function of genes through the mapping of quantitative trait loci (QTL) in structured populations, the positional cloning of strong QTL, and association mapping.

Addresses

*USDA-ARS and Department of Genetics, North Carolina State University, Raleigh, North Carolina 27695-7614, USA;
e-mail: buckler@statgen.ncsu.edu

Current Opinion in Plant Biology 2002, 5:107–111

1369-5266/02/\$ – see front matter
© 2002 Elsevier Science Ltd. All rights reserved.

Published online 30 January 2002

Abbreviations

Adh Alcohol dehydrogenase
CRY2 CRYPTOCHROME2
LD linkage disequilibrium
QTL quantitative trait loci

Introduction

Surveys of nucleotide diversity provide a snap shot of evolution at its most basic level. This nucleotide diversity reflects a rich history of selection, migration, recombination, and mating systems. Additionally, the nucleotide diversity across a genome is the source of most of the phenotypic variation.

In the past few years, there has been tremendous progress in studying diversity within plant genomes, particularly those of maize and *Arabidopsis*. In this review, we describe some of the processes that are shaping diversity within species and across their genomes, and how some of this nucleotide variation can be related to phenotypic variation.

Plant diversity

Molecular diversity has been studied in plants for about three decades. The most comprehensive early studies were done using isozymes [1], which provided many insights into population structure and breeding systems. Although these markers allowed large numbers of samples to be analyzed, comparisons of samples from different species, loci, and laboratories were problematic. More importantly, only a limited number of loci could be scored easily. In the past decade, the focus has shifted to nucleotide-level surveys of single genes from 10–20 different individuals within a species. These nucleotide studies have identified thousands of polymorphic sites that may be undergoing selection but, in comparison with the isozyme studies, these nucleotide surveys were often limited in terms of sample size. Small sample sizes may impair our ability to detect the impact of selection [2]. In the next decade, with

advances in genotyping capabilities, nucleotide surveys will surely include sufficiently large numbers of samples to allow robust analysis of population genetics.

The extent of polymorphism differs substantially between species and sampled loci. Nucleotide diversity is normally measured as the average sequence divergence between any two individuals for given locus. For example, average nucleotide diversity at any one locus ranges from less than 0.05% in some cotton loci [3] to over 5% at certain loci in *Leavenworthia stylosa* and maize [4,5]. Some of this variation in the extent of polymorphism reflects the choice of species, but major differences are also observed for random genes within a single genome. In a comprehensive study of variation within a maize chromosome, the diversity at 21 loci varied by 16-fold [6••]. The variation between loci partly reflects sampling effects, but selection and other factors also play an important role (Table 1). Until a large number of orthologous loci are sampled [7], conclusions from cross-species comparisons should be considered extremely tentative.

Although many factors influence diversity (Table 1), the neutral theory of evolution suggests that the level of polymorphism (θ) should be the product of the effective population size (N_e) and the mutation rate (μ) ($\theta = 4N_e\mu$) [8]. Unfortunately, there is little empirical proof of this simple relationship in plants. Although plant lineages differ in mutation rates [9,10], research has yet to show the connection between the mutation rate and extent of gene diversity. Proving the relationship between species population size and level of polymorphism is complicated by the need to integrate estimates of population size over evolutionary time. There has been some success in showing the effect of demographic changes in *Arabidopsis thaliana*; rapid population expansion and inbreeding have resulted in many isolated, and probably slightly deleterious, polymorphisms becoming fixed in small populations [11–13].

Background selection is likely to be one of the major factors determining nucleotide diversity [14]. In background selection, reduced diversity at neutral sites can result from selection against linked deleterious alleles that have arisen by mutation [14]. Normally, recombination breaks up chromosome regions. But, regions with low rates of recombination should experience substantial background selection, as large genomic regions are selected against whenever a linked deleterious mutation appears. In addition, a high incidence of selfing reduces the effective recombination rate, and should reduce diversity in selfing species. Background selection suggests that diversity should be shaped by recombination at the intragenomic scale and by outcrossing rate at the species level.

Table 1

Factors that impact nucleotide diversity.

Factor	Correlation with diversity	Scope
Mutation rate	Positive	Often whole genome
Population size	Positive	Whole genome
Outcrossing	Positive	Whole genome
Recombination	Positive	Whole genome
Positive-trait selection	Negative	Individual genes
Line selection	Negative	Whole genome
Diversifying selection	Positive	Individual genes
Balancing selection	Positive	Individual genes
Background selection	Negative	Individual genes or whole genome
Population structure	Mixed	Whole genome
Sequencing errors	Positive	Individual genes
PCR problems	Negative	Individual genes

The first empirical demonstration of the connection between recombination and nucleotide diversity was in *Drosophila melanogaster*. In this species, recombination rates explained much of the variation in diversity [15]. In tomatoes and other *Lycopersicon spp.*, a correlation between polymorphism and crossing-over events per physical distance along the chromosome has been established, but the effect explains only a small proportion of the variation [16*,17]. A similar weak connection has also been observed in *Beta vulgaris* [18]. Tenaillon *et al.* [6**] examined the relation between recombination and nucleotide diversity in maize, only to find somewhat mixed results. The loci near the centromere, where recombination rates should be low, were only marginally less diverse than those in other regions of the chromosome. At the gene level, Tenaillon *et al.* [6**] found a strong correlation between locus recombination rates and overall levels of diversity. The pattern at the genome level may be difficult to find if maize has hotspots of genes and recombination spread throughout the genome [19].

Some data support the connection between selfing rate and level of diversity, as suggested by the background selection theories. In surveys of the *Alcohol dehydrogenase (Adh)* locus across five species [20–25], nucleotide diversity was greatest in maize, an outcrossing species, whereas selfing species often had lower levels of diversity. The various *Leavenworthia* species have different mating systems, and the expected relationship between nucleotide diversity and outcrossing rate does appear to exist in this genus [26,27]. Comparisons between self-compatible and self-incompatible *Lycopersicon* species show a strong positive connection between diversity and outcrossing rate [16*]. Comparisons of *Adh* diversity between mainly self-pollinated *Arabidopsis thaliana* and outcrossing *Arabidopsis lyrata* indicate that *A. lyrata* has greater nucleotide diversity within this locus at the population level, but lower diversity at the species level [28].

Strong selection pressure is important in decreasing the nucleotide diversity of some plant species. Most studies of

the effect of selection pressure on nucleotide diversity have focused on domesticated crops, comparing the diversity between wild relatives and cultivars. During the selection of advantageous phenotypes, some crops appear to have passed through bottlenecks that substantially reduced diversity [29]. In contrast, many of the grass domesticates have undergone rather modest decreases in diversity relative to their wild relatives [7]. In domesticated maize, the diversity is roughly 30% below that in its closest wild relative [30–33]. However, the drop in diversity can be substantially greater in some genes that were directly involved in domestication. For example, at the *teosinte branched* locus in maize nucleotide diversity is 98% below that in the closest wild relative [34]; however, this reduction does not extend across the entire gene. The maintenance of substantial diversity in the grass crops through the domestication process may reflect the importance of the grasses as subsistence crops. It is likely that grasses such as maize, wheat, barley, and rice, had large effective population sizes that met the needs of early farmers and therefore could never be severely bottlenecked. This theory may not explain nucleotide diversity in the grasses completely; manioc, a probable subsistence domesticate, exhibits a drop in diversity of roughly 75% at the *G3pdh* locus when compared with its wild relatives [35].

Balancing selection and/or frequency-dependent selection may also play an important role in increasing diversity at specific loci within a genome. In these selection regimes, selection favors the maintenance of multiple alleles with different effects over evolutionary time. Excellent evidence comes from the self-incompatibility loci. In some of these loci, the allelic diversity may date back millions of years [36]. Additionally, disease resistance genes appear to exhibit rapid adaptive evolution in their expressed regions, probably as a result of the evolutionary arms race with pathogens [37*]. Some of these loci also exhibit balancing selection, however, with high levels of diversity [37*,38]. Another example of the influence of balancing selection is found at the phosphoglucose isomerase (*PgiC*) locus in *Leavenworthia stylosa*. Some innovative tests of linkage disequilibrium suggest that some of the high level of diversity at this locus may be the product of balancing selection [4].

Dissecting diversity

Across a large genome, such as that of maize, diversity can accumulate so that 150 million sites are commonly polymorphic. A small but important proportion of these polymorphisms is responsible for the complex variation in phenotypic traits. This naturally occurring nucleotide diversity is a treasure trove for investigating and harnessing quantitative variation. To improve crops, it is essential that we sort through this diversity to find the alleles and polymorphisms that are beneficial.

The detection of nucleotide diversity by the use of polymorphic DNA markers has allowed the analysis of naturally occurring allelic variation that is responsible for

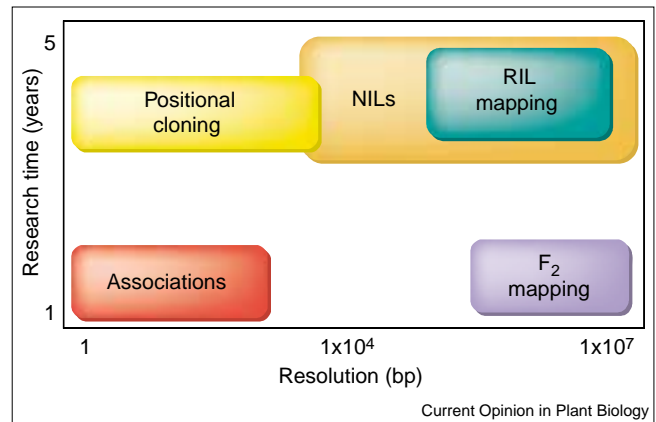
complex quantitative traits [39–41]. Initial quantitative trait loci (QTL) studies in F_2 populations and recombinant inbred lines mapped the sources of quantitative variation [42,43], but generally the resolution of these maps was limited to 5–10 cM, about 10–20 million base pairs in the case of maize. At this resolution, there are still hundreds of genes within each QTL.

Map-based strategies have been developed that can be used for the positional cloning of genes that underlie QTL (reviewed in [44]). Morphological differences between maize and its wild relative teosinte have been studied through the analysis of QTL. By combining QTL mapping, the production of near isogenic lines, and transposon tagging, one of the major QTL involved in maize domestication (i.e. *teosinte branch1*) has been cloned [45]. In tomato, two genes that underlie QTL for yield-related traits have been cloned. The gene responsible for variation in the soluble solid content of tomato fruit, identified as *Lin5*, was discovered using a map-based strategy that targeted the nucleotide diversity in wild relatives of tomato [46]. Map-based cloning and subsequent complementation tests identified a single gene, *fruit weight 2.2*, that is responsible for the variation in tomato fruit size [47]. In rice, the same strategy enabled the cloning of *Heading date1*, a major flowering-time QTL, which encodes a protein with high similarity to that encoded by the *Arabidopsis* gene *CONSTANS* [48]. A single QTL at the *Frigida* locus, which is responsible for the vernalization response of *Arabidopsis* flowering, was also cloned using a map-based strategy [49]. Most recently, QTL mapping and positional cloning were used to identify a unique allele of *CRYPTOCHROME2* (*CRY2*), which is responsible for some of the variation in *Arabidopsis* flowering time [50**].

Despite the success of these strategies, gene discovery appears to be limited to those loci that have large effects upon quantitative variation. Quantitative traits are generally the product of numerous loci with varying degrees of effect upon the observed phenotypes. Techniques are therefore needed to rapidly identify genes that play a modest role in regulating quantitative variation. Current procedures are very time consuming; in species that are limited to two growing seasons per year, it can take five years to produce the population needed for fine-scale mapping. With thousands of genes to evaluate for QTL effects, a more efficient approach is needed to complement map-based cloning. This role may be fulfilled by the application of association tests to naturally occurring populations [51].

Association approaches have been used effectively in human genetics [52,53], in which controlled breeding is not possible and large numbers of progeny are not available. In these approaches, candidate gene diversity is evaluated across natural populations, and polymorphisms that correlate with phenotypic variation are identified. The key advantages of association tests include their speed, because no mapping population need be created, and high resolution

Figure 1



Comparison of resolution and research time for various approaches to dissect quantitative variation. The research times assume the target species has only two generations per year. NIL, near-isogenic line; RIL, recombinant inbred line.

(Figure 1). The resolution of association approaches depends on the structure of linkage disequilibrium (LD) (i.e. on the correlation between polymorphic loci) within the test population. LD structure is being extensively evaluated in humans [54,55], but has received little attention in plants until recently. Surveys in maize suggest that LD structure can decay quite rapidly, within a few hundred bases in landraces and within 2000 bases in diverse breeding material [6**,56*]. Even in synthetic populations, the level of LD is modest [57]. There is new evidence, however, that this decay is much slower in elite maize germplasm (see review by Rafalski, this issue).

The primary obstacle to successful association studies in plants is the nature of population structure. The presence of subgroups with an unequal distribution of alleles within a population can result in non-functional, spurious associations [58]. In such populations, highly significant associations between a marker and a phenotype may be suggested [59], even though the marker is not physically linked to the locus responsible for the phenotypic variation. The complex breeding history of many agronomically important crops and the limited gene flow in most wild plants have created complex stratification within germplasm, which complicates association studies [60].

In recent years, a few statistical methods have been developed that use independent marker loci to detect stratified populations and to correct for them [61]. These methods work on the assumption that population structure should have similar effects upon all loci. Reich and Goldstein [62] propose scoring the association of a moderate number of unlinked genetic markers with a given phenotype, and then comparing the strength of these associations with that of the candidate gene's association. Pritchard *et al.* [63*,64**] have developed an approach that incorporates estimates of population structure directly into the association test statistic.

The Pritchard approach has been modified for use with quantitative traits and, in the first empirical application of these methods, has been used to study flowering time in maize [65**]. In this study, the polymorphisms in the maize *Dwarf8* gene were significantly associated with variation in flowering time. By accounting for population structure, false positives were reduced in number by up to 80%. Using these statistical methods in an association test allowed researchers to improve their resolution from the level of a genetic bin to an individual gene. The identified allele could be used in the molecular breeding of maize.

Conclusions

High-throughput DNA sequencing allows surveys of nucleotide diversity to be conducted for a wide range of species and loci, and evolutionary questions are starting to be addressed using this wealth of data. Until carefully designed studies of multiple orthologous loci across several species are conducted, our understanding of the processes underlying nucleotide diversity will be limited. Association tests in natural populations are providing an exciting opportunity to simultaneously use diversity to understand the function of genes and to find useful alleles for plant breeding and crop improvement. Association approaches are amenable to high-throughput genomics and could be used to characterize all of the genes in a genome.

Acknowledgements

We thank Brad Rauh for his help in researching this topic, and Sandra Andaluz, Carlyn Buckler, and Sherry Whitt for commenting on the manuscript. JMT was supported by National Science Foundation grant DBI-9872631.

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. Hamrick JL, Godt MJW: **Allozyme diversity in plant species**. In *Plant Population Genetics, Breeding, and Genetic Resources*. Edited by Brown AHD, Clegg MT, Kahler AL, Weir BS. Sunderland, MA: Sinauer Associates Inc.; 1990:43-63.
 2. Simonsen KL, Churchill GA, Aquadro CF: **Properties of statistical tests of neutrality for DNA polymorphism data**. *Genetics* 1995, **141**:413-429.
 3. Small RL, Ryburn JA, Wendel JF: **Low levels of nucleotide diversity at homoeologous *Adh* loci in allotetraploid cotton (*Gossypium* L.)**. *Mol Biol Evol* 1999, **16**:491-501.
 4. Filatov DA, Charlesworth D: **DNA polymorphism, haplotype structure and balancing selection in the *Leavenworthia PgiC* locus**. *Genetics* 1999, **153**:1423-1434.
 5. Henry AM, Damerval C: **High rates of polymorphism and recombination at the *Opaque-2* locus in cultivated maize**. *Mol Genet* 1997, **256**:147-157.
 6. Tenailon MI, Sawkins MC, Long AD, Gaut RL, Doebley JF, Gaut BS: **Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.)**. *Proc Natl Acad Sci USA* 2001, **98**:9161-9166.
- The most complete single survey of diversity for any plant is presented. The authors relate nucleotide diversity to chromosome structure, recombination, linkage disequilibrium and various types of selection. Population effects in the production of breeding germplasm in maize are also explored.
7. Buckler ES IV, Thornsberry JM, Kresovich S: **Molecular diversity, structure and domestication of grasses**. *Genet Res* 2001, **77**:213-218.
 8. Kimura M: **The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations**. *Genetics* 1969, **61**:893-903.
 9. Muse SV, Gaut BS: **Comparing patterns of nucleotide substitution rates among chloroplast loci using the relative ratio test**. *Genetics* 1997, **146**:393-399.
 10. Muse SV: **Examining rates and patterns of nucleotide substitution in plants**. *Plant Mol Biol* 2000, **42**:25-43.
 11. Purugganan MD, Suddith JI: **Molecular population genetics of floral homeotic loci: departures from the equilibrium-neutral model at the *APETALA3* and *PISTILLATA* genes of *Arabidopsis thaliana***. *Genetics* 1999, **151**:839-848.
 12. Kuittinen H, Aguade M: **Nucleotide variation at the *CHALCONE ISOMERASE* locus in *Arabidopsis thaliana***. *Genetics* 2000, **155**:863-872.
 13. Kawabe A, Yamane K, Miyashita NT: **DNA polymorphism at the cytosolic phosphoglucose isomerase (*PgiC*) locus of the wild plant *Arabidopsis thaliana***. *Genetics* 2000, **156**:1339-1347.
 14. Charlesworth B, Morgan MT, Charlesworth D: **The effect of deleterious mutations on neutral molecular variation**. *Genetics* 1993, **134**:1289-1303.
 15. Begun DJ, Aquadro CF: **Levels of naturally occurring DNA polymorphism correlate with recombination rate in *D. melanogaster***. *Nature* 1992, **356**:519-520.
 16. Baudry E, Kerdelhue C, Innan H, Stephan W: **Species and recombination effects on DNA variability in the tomato genus**. *Genetics* 2001, **158**:1725-1735.
- A nice examination of the effects of both inbreeding and recombination on diversity. The authors found that mating system had a highly significant effect on polymorphism, whereas recombination had only a weak influence. Self-compatible species had much lower levels of diversity than did self-incompatible species.
17. Stephan W, Langley CH: **DNA polymorphism in *Lycopersicon* and crossing-over per physical length**. *Genetics* 1998, **150**:1585-1593.
 18. Kraft T, Sall T, Magnusson-Rading I, Nilsson NO, Hallden C: **Positive correlation between recombination rates and levels of genetic variation in natural populations of sea beet (*Beta vulgaris* subsp. *maritima*)**. *Genetics* 1998, **150**:1239-1244.
 19. Fu H, Park W, Yan X, Zheng Z, Shen B, Dooner HK: **The highly recombinogenic *bz* locus lies in an unusually gene-rich region of the maize genome**. *Proc Natl Acad Sci USA* 2001, **98**:8903-8908.
 20. Gaut BS, Clegg MT: **Molecular evolution of the *Adh1* locus in the genus *Zea***. *Proc Natl Acad Sci USA* 1993, **90**:5095-5099.
 21. Gaut BS, Clegg MT: **Nucleotide polymorphism in the *Adh1* locus of pearl millet (*Pennisetum glaucum*) (Poaceae)**. *Genetics* 1993, **135**:1091-1097.
 22. Innan H, Tajima F, Terauchi R, Miyashita NT: **Intragenic recombination in the *Adh* locus of the wild plant *Arabidopsis thaliana***. *Genetics* 1996, **143**:1761-1770.
 23. Miyashita N, Innan H, Terauchi R: **Intra- and interspecific variation of the alcohol dehydrogenase locus region in wild plants *Arabis gemmifera* and *Arabidopsis thaliana***. *Mol Biol Evol* 1996, **13**:433-436.
 24. Cummings MP, Clegg MT: **Nucleotide sequence diversity at the alcohol dehydrogenase 1 locus in wild barley (*Hordeum vulgare* ssp. *spontaneum*): an evaluation of the background selection hypothesis**. *Proc Natl Acad Sci USA* 1998, **95**:5637-5642.
 25. Miyashita NT: **DNA variation in the 5' upstream region of the *Adh* locus of the wild plants *Arabidopsis thaliana* and *Arabis gemmifera***. *Mol Biol Evol* 2001, **18**:164-171.
 26. Liu F, Charlesworth D, Kreitman M: **The effect of mating system differences on nucleotide diversity at the phosphoglucose isomerase locus in the plant genus *Leavenworthia***. *Genetics* 1999, **151**:343-357.
 27. Liu F, Zhang L, Charlesworth D: **Genetic diversity in *Leavenworthia* populations with different inbreeding levels**. *Proc R Soc London B Biol Sci* 1998, **265**:293-301.

28. Savolainen O, Langley CH, Lazzaro BP, Freville H: Contrasting patterns of nucleotide polymorphism at the alcohol dehydrogenase locus in the outcrossing *Arabidopsis lyrata* and the selfing *Arabidopsis thaliana*. *Mol Biol Evol* 2000, 17:645-655.
29. Doebley J: **Molecular systematics and crop evolution**. In *Molecular Systematics of Plants*. Edited by Soltis DE, Soltis PS, Doyle JJ. New York: Chapman & Hall; 1992:202-222.
30. Eyre-Walker A, Gaut RL, Hilton H, Feldman DL, Gaut BS: Investigation of the bottleneck leading to the domestication of maize. *Proc Natl Acad Sci USA* 1998, 95:4441-4446.
31. White SE, Doebley JF: The molecular evolution of *terminal ear1*, a regulatory gene in the genus *Zea*. *Genetics* 1999, 153:1455-1462.
32. Hilton H, Gaut BS: Speciation and domestication in maize and its wild relatives: evidence from the *globulin-1* gene. *Genetics* 1998, 150:863-872.
33. Goloubinoff P, Pääbo S, Wilson AC: Evolution of maize inferred from sequence diversity of an *Adh2* gene segment from archaeological specimens. *Proc Natl Acad Sci USA* 1993, 90:1997-2001.
34. Wang RL, Stec A, Hey J, Lukens L, Doebley J: The limits of selection during maize domestication. *Nature* 1999, 398:236-239.
35. Olsen KM, Schaal BA: Evidence on the origin of cassava: phylogeography of *Manihot esculenta*. *Proc Natl Acad Sci USA* 1999, 96:5586-5591.
36. Richman AD, Uyenoyama MK, Kohn JR: Allelic diversity and gene genealogy at the self-incompatibility locus in the solanaceae. *Science* 1996, 273:1212-1216.
37. Bergelson J, Kreitman M, Stahl EA, Tian D: **Evolutionary dynamics of plant R-genes**. *Science* 2001, 292:2281-2285. This study examines the evidence for the 'arms race' hypothesis for several disease resistance genes. Contrary to a simple arms race hypothesis, the authors find substantial diversity for many resistance genes. They develop an innovative hypothesis to explain how divergent alleles could be maintained for millions of years.
38. Stahl EA, Dwyer G, Mauricio R, Kreitman M, Bergelson J: Dynamics of disease resistance polymorphism at the *Rpm1* locus of *Arabidopsis*. *Nature* 1999, 400:667-671.
39. Yano M, Sasaki T: Genetic and molecular dissection of quantitative traits in rice. *Plant Mol Biol* 1997, 35:145-153.
40. Tanksley SD: **Mapping polygenes**. *Annu Rev Genet* 1993, 27:205-233.
41. Paterson AH: **Molecular dissection of quantitative traits: progress and prospects**. *Genome Res* 1995, 5:321-333.
42. Burr B, Burr FA, Thompson KH, Albertson MC, Stuber CW: **Gene-mapping with recombinant inbreds in maize**. *Genetics* 1988, 118:519-526.
43. Edwards MD, Stuber CW, Wendel JF: **Molecular-marker-facilitated investigations of quantitative-trait loci in maize**. I. Numbers, genomic distribution and types of gene action. *Genetics* 1987, 116:113-125.
44. Yano M: **Genetic and molecular dissection of naturally occurring variation**. *Curr Opin Plant Biol* 2001, 4:130-135.
45. Doebley J, Stec A, Hubbard L: **The evolution of apical dominance in maize**. *Nature* 1997, 386:485-488.
46. Fridman E, Pleban T, Zamir D: **A recombination hotspot delimits a wild-species quantitative trait locus for tomato sugar content to 484 bp within an invertase gene**. *Proc Natl Acad Sci USA* 2000, 97:4718-4723.
47. Fray A, Nesbitt TC, Grandillo S, van der Knaap E, Cong B, Liu JP, 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.
48. Yano M, Katayose Y, Ashikari M, Yamanouchi U, Monna L, Fuse T, Baba T, Yamamoto K, Umehara Y, Nagamura Y *et al.*: ***Hd1*, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the *Arabidopsis* flowering time gene *CONSTANS***. *Plant Cell* 2000, 12:2473-2483.
49. Johanson U, West J, Lister C, Michaels S, Amasino R, Dean C: **Molecular analysis of *FRIGIDA*, a major determinant of natural variation in *Arabidopsis* flowering time**. *Science* 2000, 290:344-347.
50. El-Assal S, Alonso-Blanco C, Peeters A, Raz V, Koornneef M: **A QTL for flowering time in *Arabidopsis* reveals a novel allele of *CRY2***. *Nat Genet* 2001, 29:435-440. The work described in this paper is a tour-de-force that takes full advantage of the tools available to the *Arabidopsis* community. Use of QTL mapping and a positional cloning strategy allowed the identification of a novel allele of *CRY2* that is responsible for a flowering time QTL. Unlike the phenotypic variation of other genes underlying QTL, variation due to the *CRY2-Cvi* allele is the product of an amino-acid substitution. This was confirmed by transformation.
51. Risch NJ: **Searching for genetic determinants in the new millennium**. *Nature* 2000, 405:847-856.
52. Corder EH, Saunders AM, Risch NJ, Strittmatter WJ, Schmechel DE, Gaskell PC, Rimmer JB, Locke PA, Conneally PM, Schmechel KE *et al.*: **Protective effect of apolipoprotein E type 2 allele for late-onset Alzheimer disease**. *Nat Genet* 1994, 7:180-184.
53. Templeton AR: **A cladistic-analysis of phenotypic associations with haplotypes inferred from restriction-endonuclease mapping or DNA-sequencing**. 5. Analysis of case-control sampling designs – Alzheimer's-disease and the apoprotein-E locus. *Genetics* 1995, 140:403-409.
54. Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, Kouyoumjian R, Farhadian SF, Ward R *et al.*: **Linkage disequilibrium in the human genome**. *Nature* 2001, 411:199-204.
55. Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES: **High-resolution haplotype structure in the human genome**. *Nat Genet* 2001, 29:229-232.
56. Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, Kresovich S, Goodman MM, Buckler ES: **Structure of linkage disequilibrium and phenotypic associations in the maize genome**. *Proc Natl Acad Sci USA* 2001, 98:11479-11484. This study examines the structure of disequilibrium and evaluates the potential high resolution of association approaches in maize. However, this work also highlights some of the problems that may be created by population structure.
57. Labate JA, Lamkey KR, Lee M, Woodman W: **Hardy-Weinberg and linkage equilibrium estimates in the BSSS and BSCB1 random mated populations**. *Maydica* 2000, 45:243-256.
58. Knowler WC, Williams RC, Pettitt DJ, Steinberg AG: **Gm^{3:5:13:14} and Type 2 diabetes mellitus: an association in American Indians with genetic admixture**. *Am J Hum Genet* 1988, 43:520-526.
59. Pritchard JK, Rosenberg NA: **Use of unlinked genetic markers to detect population stratification in association studies**. *Am J Hum Genet* 1999, 65:220-228.
60. Sharbel TF, Haubold B, Mitchell-Olds T: **Genetic isolation by distance in *Arabidopsis thaliana*: biogeography and postglacial colonization of Europe**. *Mol Ecol* 2000, 9:2109-2118.
61. Pritchard JK, Rosenberg NA: **Use of unlinked genetic markers to detect population stratification in association studies**. *Am J Hum Genet* 1999, 65:220-228.
62. Reich DE, Goldstein DB: **Detecting association in a case-control study while correcting for population stratification**. *Genet Epidemiol* 2001, 20:4-16.
63. Pritchard JK, Stephens M, Donnelly P: **Inference of population structure using multilocus genotype data**. *Genetics* 2000, 155:945-959. The authors describe a useful approach for determining substructure within a population. This approach utilizes a set of markers, distributed throughout the genome, to determine the interrelatedness of members of a population. These estimates can then be used in association tests.
64. Pritchard JK, Stephens M, Rosenberg NA, Donnelly P: **Association mapping in structured populations**. *Am J Hum Genet* 2000, 67:170-181. The unequal distribution of alleles within subgroups of a population may lead to spurious associations. Building upon previous work by Pritchard and Rosenberg, the authors were able to reduce the number of false-positive associations through the use of a matrix that described the population structure.
65. Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler ES: ***Dwarf8* polymorphisms associate with variation in flowering time**. *Nat Genet* 2001, 28:286-289. The first to use of estimates of population structure to reduce spurious associations. By incorporating estimates of population structure, the number of false positives was reduced by up to 80%. A suite of polymorphisms was identified in the maize gene *Dwarf8*, which is associated with variation in flowering time.