also suggests that newly emerging resequencing methods should be used to create the definitive haplotype map of the mouse, including not only SNPs but also copy number variants and genome rearrangements.

COMPETING INTERESTS STATEMENT
The author declares no competing financial interests.

1. Flint, J., Valdar, W., Shifman, S. & Mott, R. *Nat. Rev. Genet.* **6**, 271–286 (2005).
2. Frazer, K.A. *et al. Nature*, advance online publication 29 July 2007 (doi:10.1038/nature06067).
3. Yang, H., Bell, T.A., Churchill, G.A. & Pardo-Manuel de Villena, F. *Nat. Genet.* **39**, 1100–1107 (2007).
4. Waterston, R.H. *et al. Nature* **420**, 520–562 (2002).
5. Mott, R., Talbot, C.J., Turri, M.G., Collins, A.C. & Flint, J. *Proc. Natl. Acad. Sci. USA* **97**, 12649–12654 (2000).
6. Wade, C.M. *et al. Nature* **420**, 574–578 (2002).
7. Yalcin, B. *et al. Proc. Natl. Acad. Sci. USA* **101**, 9734–9739 (2004).
8. Frazer, K.A. *et al. Genome Res.* **14**, 1493–1500 (2004).
9. Zhang, J. *et al. Genome Res.* **15**, 241–249 (2005).
10. Churchill, G.A. *et al. Nat. Genet.* **36**, 1133–1137 (2004).

An *Arabidopsis* haplotype map takes root

Edward Buckler & Michael Gore

The report of a haplotype map for the selfing plant *Arabidopsis thaliana* has uncovered numerous major-effect polymorphisms and rapid linkage disequilibrium decay. This work lays the foundation for genome-wide association studies at near-gene-level resolution in a model organism possessing substantial functional diversity and extensive community resources.

Systematic mutagenesis, the use of large linkage populations and positional cloning are common approaches for dissecting complex traits in plants. In human genetics, recent genome-wide association studies have mapped common complex diseases, but the plant genetics community has been awaiting a haplotype map comparable to the human HapMap to make a similar approach feasible in plants. Clark *et al.*[1] recently reported a haplotype map for the predominantly self-pollinating ('selfing') species *Arabidopsis thaliana*, and in a companion paper in this issue, Kim *et al.* (page 1151)[2] characterize linkage disequilibrium (LD) in these strains. This extensive catalog of polymorphisms[1] and detailed examination of LD[2] set the stage for high-resolution genome-wide association scans in this model selfing species.

A. thaliana diversity and LD

Laying the foundation for an *A. thaliana* haplotype map, Clark *et al.*[1] conducted a thorough array resequencing of 20 diverse *A. thaliana* genomes at single-base resolution. This provided a powerful catalog of genetic diversity, with more than 1 million SNPs and hypervariable regions (50-bp to >10-kb deletions and SNP clusters). More than 100,000 amino changes were identified, along with nearly 2,500 polymorphisms that should radically alter transcript or protein structure. Overall, they discovered an average of one polymorphism every 166 bp. For comparison, this polymorphism rate is 11 times that found in a human haplotype map study done with a comparable platform[3]. This is a tremendous amount of functional variation, which has certainly had a role in the adaptation of *A. thaliana* to environments throughout the globe.

In a companion study, Kim *et al.* characterize LD in these *A. thaliana* collections[2]. Theory suggests that selfing species should have higher levels of LD across the genome, as there are few opportunities for effective recombination[4]. In outcrossing species with large population sizes like that of maize and *Drosophila melanogaster*, rapid LD decay provides sub-gene-level mapping resolution[5]. However, the LD structure of selfing species with large population sizes has been an open question. Initial *A. thaliana* surveys, at the FRI locus, suggested extensive LD (up to 250 kb)[6], and a low-density genome scan suggested that LD decayed within 20–40 kb[7]. Kim *et al.* now show that LD decays substantially within 10 kb and sometimes within 3–4 kb for a very diverse subset of ecotypes[2]. So despite being predominantly selfing, *A. thaliana* must historically have had enough sex and recombination to break down linkage blocks. Kim *et al.* also found hotspots of LD decay that showed moderate to weak correlations with the level of polymorphism, recombination and GC content, although the mechanisms controlling regions of high and low recombination were not clear[2]. Larger sample sizes will be necessary to understand these processes, as even repeat content had little correlation, despite it being very important in other plant species. The low degree of LD found in these studies is exciting, as it suggests that diverse *A. thaliana* lines may provide near-gene-level resolution for association mapping.

The future of association mapping

These studies pave the way for developing a powerful association mapping panel for *A. thaliana*. In fact, these groups are now developing a tiling array to score an association panel of 1,000 ecotypes. One key question for this future genotyping effort is how many SNPs will be needed to provide markers in high LD with all regions of the genome. Kim *et al.* convincingly show that 200,000 tag SNPs should be sufficient to cover almost all of the genome[2]. A second matter is ascertainment bias, as these resequencing arrays were designed from a single sequenced reference genome. With resequencing arrays, multiple adjacent polymorphisms cause a decrease in hybridization signal intensity, thus preventing many of the highly polymorphic regions from being characterized at single–base pair resolution. In the current studies, this resulted in only 27% of the total polymorphisms being scored in a given line[1]. Additionally, although non-allelic structural polymorphisms are probably common[8], these have not been comprehensively catalogued or discovered by this approach, which will probably require next-generation sequencing approaches to identify these potentially important variants. These tiling arrays, when combined with the recently developed analysis methods that account for complex population structures[9,10], will allow powerful full-genome association scans.

Understanding the phenotypic importance of rare SNPs is also critical for association studies and genetic architecture analyses. In this sample of 20 *A. thaliana* ecotypes, nearly 50% of the polymorphisms were unique to an ecotype[2]. Do these rare alleles have phenotypic effects? If these rare SNPs have important phenotypic effects, then alternative mapping strategies may be needed. Currently, the phenotypic

Edward Buckler is in the US Department of Agriculture–Agricultural Research Service, Ithaca, New York 14853, USA, and the Department of Plant Breeding and Genetics, Cornell University, Ithaca, New York 14853, USA. Michael Gore is in the Department of Plant Breeding and Genetics, Cornell University, Ithaca, New York 14853, USA.
e-mail: esb33@cornell.edu

Table 1 Comparison of publicly available association mapping resources in plants and humans

	A. thaliana[7]	Maize[5]	Barley[11]	Rice[12]	Sorghum[13]	Soybean[14]	Human[15]
Silent diversity	Ecotypes: 0.7%	Wild: 2.1% Landraces: 1.4% Diverse inbreds: 1.2% Elites: 0.63%	Wild: 1.7% Landraces: 0.71% Elites: 0.47%	Wild: 0.58% *O.sativa*: 0.35%	Landraces: 0.24%	Wild: 0.28% Landraces: 0.18% Elites: 0.12%	0.05%
LD decay[a]	Ecotypes: <10 kb	Wild: <1 kb Landraces: <1 kb Diverse inbreds: 1–2 kb Elites: >100 kb	Wild: <1 kb Landraces: 80–100 kb Elites: >200 kb	Divergent haplotypes and extensive LD	Landraces: 5–50 kb	Wild: 36–77 kb Elites: >300 kb	10–100 kb
Predominant mating type	Selfing	Outcrossing	Selfing	Selfing	Selfing	Selfing	Outcrossing
Diverse association samples[b]	100 (plus 1,000 in 2008) ecotypes	281 inbreds	102 elite 1,000+ breeding lines (by 2010)	400 *O. sativa* 100 *O. rufipogon*	377 inbreds	120 wild, ancestral, elite	
Biparental populations for association[c]	5 (plus 18 in progress)	27 by 2007–2008 (5,500 lines) and several others	1 (plus 2 in progress)	Numerous separate populations	21 by 2010	Numerous separate populations	

[a]Note large variances. [b]Publicly available lines (clones) that have been genotyped with background markers and are ready for candidate gene studies and/or genome scans. [c]Generally, recombinant inbred populations derived from two diverse inbred parents. In plants, many of these populations have been created, but we highlight those being made into public resources and being genotyped for the purpose of association mapping.

importance of rare polymorphisms is unclear, but in plants, where numerous bi-parental populations can be readily constructed, we should be able to resolve this question.

Functional diversity toolboxes

Aided by the *A. thaliana* haplotype map, genome-wide association mapping in plants is rapidly becoming a reality. The decreasing costs of genotyping, as well as the relatively low cost of creating inbred lines, sharing seed and evaluation of thousands or millions of genotypes in multiple environments, are easing this transition. In several plant species, diverse germplasm or genotypic panels are being established for whole-genome association mapping (**Table 1**), which facilitates the scoring of the same genotypes by an entire community of researchers. The focus on shared community panels will also allow integration of studies and wider analysis of these data sets by researchers across fields. Overall, this should facilitate mapping functional variation and allow for a deeper understanding of genetic architecture and mechanisms of adapatation. In crops, this knowledge will be applied to improving sustainability, nutritional value and yield stability in dynamic environments, as well as improving global carbon balance.

COMPETING INTERESTS STATEMENT
The authors declare no competing financial interests.

1. Clark, R.M. *et al. Science* **317**, 338–342 (2007).
2. Kim, S. *et al. Nat. Genet.* **39**, 1151–1155 (2007).
3. Hinds, D.A. *et al. Science* **307**, 1072–1079 (2005).
4. Nordborg, M. *Genetics* **154**, 923–929 (2000).
5. Yu, J. & Buckler, E.S. *Curr. Opin. Biotechnol.* **17**, 155–160 (2006).
6. Nordborg, M. *et al. Nat. Genet.* **30**, 190–193 (2002).
7. Nordborg, M. *et al. PLoS Biol.* **3**, 1289–1299 (2005).
8. Fu, H. & Dooner, H.K. *Proc. Natl. Acad. Sci. USA* **99**, 9573–9578 (2002).
9. Yu, J. *et al. Nat. Genet.* **38**, 203–208 (2006).
10. Zhao, K. *et al. PLoS Genet. [online]* **3**, e4 (2007) (doi: 10.1371/journal.pgen.0030004).
11. Caldwell, K.S., Russell, J., Langridge, P. & Powell, W. *Genetics* **172**, 557–567 (2006).
12. Tang, T. *et al. PLoS Genet. [online]* **2**, e199 (2006) (doi:10.1371/journal.pgen.0020199).
13. Hamblin, M.T. *et al. Genetics* **173**, 953–964 (2006).
14. Hyten, D.L. *et al. Genetics* **175**, 1937–1944 (2007).
15. International HapMap Consortium. *Nature* **437**, 1299–1320 (2005).

Cancer drugs to treat birth defects

Andrew O M Wilkie

Identical mutations of the same genes can lead either to congenital malformations or to cancer, depending on their cellular and temporal context. The demonstration of activated RAS-ERK signaling in a mouse model of Apert syndrome suggests that drugs designed to inhibit this pathway in cancer may also delay the progression of several serious pediatric syndromes.

Until a few years ago, the title above might have seemed to belong only in the headlines of the tabloid press. However, as the genes mutated in birth defects and cancer have been identified, and the details of how these mutations disturb the regulation of biochemical pathways have been explained, a remarkable convergence in their underlying cellular mechanisms has been uncovered. This is well illustrated in a study by Vivek Shukla, Xavier Coumoul and colleagues[1] on page 1145 of this issue. In a mouse model of Apert syndrome in which affected pups normally die within a few weeks with craniofacial malformations, injection of the pregnant mother with a specific signaling inhibitor enables the mutant

Andrew O.M. Wilkie is at the Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK.
e-mail: awilkie@hammer.imm.ox.ac.uk