

Opinion

The 1001 Genomes Project for *Arabidopsis thaliana*

Detlef Weigel* and Richard Mott†

Addresses: *Department of Molecular Biology, Max Planck Institute for Developmental Biology, 72076 Tübingen, Germany. †Wellcome Trust Centre for Human Genetics, Oxford OX3 7BN, UK.

Correspondence: Detlef Weigel. E-mail weigel@weigelworld.org

Published: 27 May 2009

Genome Biology 2009, **10**:107 (doi:10.1186/gb-2009-10-5-107)

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2009/10/5/107>

© 2009 BioMed Central Ltd

Abstract

We advocate here a 1001 Genomes project for *Arabidopsis thaliana*, the workhorse of plant genetics, which will provide an enormous boost for plant research with a modest financial investment.

Arabidopsis thaliana

Thale cress, *Arabidopsis thaliana*, is a member of one of the largest families of flowering plants, the Brassicaceae, to which mustards, radishes and cabbages also belong. *A. thaliana* is thought to have originated in Central Asia and spread from there throughout Eurasia. During the last glaciation, *A. thaliana* was confined to the southern limit of its range, and after the ice retreated, much of Europe was recolonized by different populations, resulting in complex admixture patterns. Today, *A. thaliana* occurs throughout the Northern Hemisphere, mostly in temperate regions, from the mountains of North Africa to the Arctic Circle (Figure 1). Like many other European plants, it has also invaded North America, most probably during historic times [1-5].

The ascendancy of *A. thaliana* to become one of the most popular species in basic plant research [6], despite its lack of economic value, is due to the favorable genetics of this plant. It has a diploid genome of only about 125 to 150 Mb distributed over five chromosomes, with fewer than 30,000 protein-coding genes. The ease with which it can be stably transformed is unsurpassed by any other multicellular organism [7]. Moreover, as flowering plants only appeared about 100 million years ago, they are all relatively closely related. Indeed, key aspects of plant physiology such as flowering are highly conserved between economically important grasses such as rice and *A. thaliana* [8].

A. thaliana was the first plant species for which a genome sequence became available. This initial sequence was from a

single inbred strain (accession), and was of very high quality, with each chromosome represented by merely two contigs, one for each arm [9]. In addition to functional analyses, the 120 Mb reference sequence of the Columbia (Col-0) accession proved to be a boon for evolutionary and ecological genetics. A particular advantage in this respect is that the species is mostly self-fertilizing, and most strains collected from the wild are homozygous throughout the genome. This distinguishes *A. thaliana* from other model organisms such as the mouse or the fruit fly. In these systems, inbred strains have been derived, but they do not represent any individuals actually found in nature.

Identifying genotypic and phenotypic variation in natural accessions

Natural *A. thaliana* accessions show tremendous genetic and phenotypic diversity [10,11] (Figure 1b). Over the past 10 years, traditional quantitative trait locus (QTL) mapping has led to the identification of sequence variants that modulate a range of physiological and developmental traits, from germination and flowering to ion content [10,11]. Prior knowledge of the biological function of the affected genes was often helpful in identifying them, but increasingly, the responsible locus is found to encode a protein without known biochemical function such as the *FRIGIDA* (*FRI*) flowering regulator or the *DELAYED GERMINATION1* (*DOG1*) gene [12-14]. Apart from alleles that alter expression levels or protein function, a surprising number of drastic mutations such as deletions and stop codons underlie

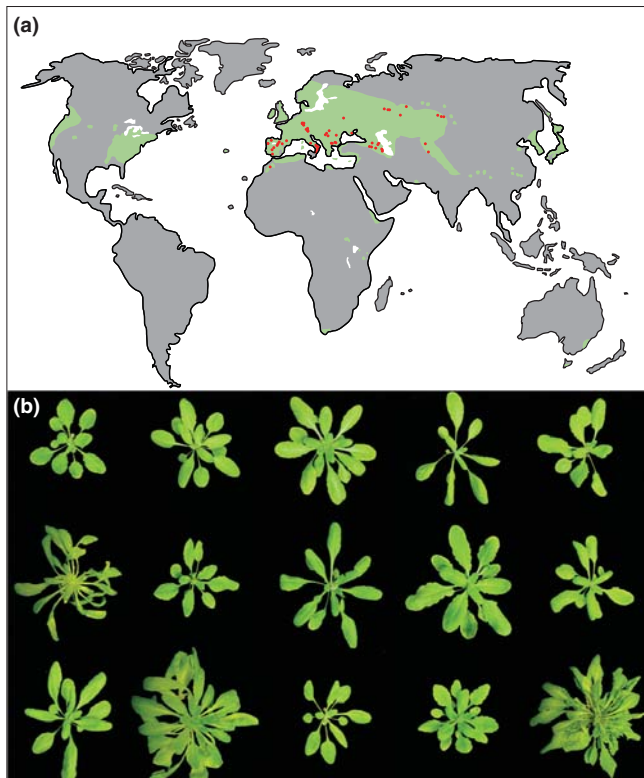


Figure 1
Intraspecific variation in *Arabidopsis thaliana*. **(a)** *A. thaliana* (area of distribution shaded in green) is found throughout the Northern Hemisphere. It is a native of Eurasia and has been introduced into North America, Australia and southern Africa. The provenances of the first 74 accessions that have been sequenced as part of the 1001 Genomes project are indicated by the red dots. **(b)** Vegetative rosettes illustrating genetically determined variation in morphology among *A. thaliana* accessions.

phenotypic variation. Some of these changes are found in many accessions (see, for example [12,15]), suggesting that they are adaptive. Nevertheless, despite some success stories, the number of known alleles responsible for phenotypic variation among accessions remains limited, mostly because fine mapping and dissection of QTLs are so tedious.

Efforts to accelerate the discovery of functionally important variants began with a large-scale study in which some 1,000 fragments across the genomes of 96 accessions gathered from all over the world were compared by dideoxy sequencing [4]. A major conclusion from this work was that there has been considerable global gene flow, so that most sequence variants are found worldwide, although genotypes are not entirely random. There is isolation by distance, and even though population structure is relatively moderate, it can easily be a confounding factor in association studies. These properties are reminiscent of what has been described for humans [16-20].

A first-generation haplotype map (HapMap) for *A. thaliana*

From this first set of 96 strains, 20 maximally diverse strains were chosen for much denser polymorphism discovery using array-based resequencing [21]. This led to the identification of about one single nucleotide polymorphism (SNP) for every 200 bp of the genome, constituting one quarter or so of all SNPs estimated to be present. In addition, regions that are missing or highly divergent in at least one accession encompass about a quarter of the reference genome [22].

The progress made with genome-wide association (GWA) mapping in humans during the past three years has been nothing but phenomenal [23], and bodes well for applying association mapping to *A. thaliana*. As in humans, linkage disequilibrium (LD), which is the basis for GWA studies, decays over about 10 kb, the equivalent of two average genes [24]. That the average LD in *Arabidopsis* is not so different from that in humans might seem surprising, given the selfing nature of *A. thaliana*, but it reflects the fact that outcrossing is not that rare, and that this species apparently has a large effective population size. A 250k SNP chip (containing 250,000 probes), corresponding to approximately one SNP every 480 bp, has been produced, and should predict some 90% of all non-singleton SNPs [24]. A collection of over 6,000 *A. thaliana* accessions, both from stock centers and recent collections (for example [25]) has been assembled, and a subset of 1,200 genetically diverse strains will be interrogated with the 250k SNP chip [26], providing a fantastic resource for GWA studies in this species.

A single genome is not enough

It is becoming increasingly clear that it is inappropriate to think about 'the' genome of a species, even though this is what the initial sequencing papers stated in their titles just a few years ago (as in "Initial sequencing and analysis of the human genome" and "The sequence of the human genome") [27,28]. The previous emphasis on relatively minor changes between individuals, such as SNPs and small indels, was largely due to the fact that sequence variation had overwhelmingly been studied by PCR-based methods or hybridization to known sequences. It is now known that *A. thaliana* accessions can vary in hundreds of genes [21,29], and similar findings have emerged for other species, including humans (for example [30,31]). Of particular importance is the observation that some genes with fundamental effects on life-history traits such as flowering are not even functional in the *A. thaliana* Col-0 reference accession [12], and thus could not have been discovered on the basis of the first genome sequence alone.

The 250k SNP genotyping effort discussed above is an important step towards identifying haplotype blocks associated with specific trait variants, but it has several limitations. First, the initial SNP discovery phase had

considerable, technology-inherent shortcomings, and only a minority of all SNPs was detected [21]. Second, these SNPs were defined in a relatively small initial sample that probably captures only a fraction of species-wide diversity. Genotyping with SNPs common in the global population will provide little information on new alleles that have arisen on the background of older haplotypes, which would be particularly relevant for studies of local populations. Third, although the impact of structural variation is unknown, it might have dramatic consequences on phenotypic diversity.

The *A. thaliana* 1001 Genomes project

Together with partners from around the world, we have initiated a project with the goal of describing the whole-genome sequence variation in 1,001 accessions of *A. thaliana* [32]. The current technological revolution in sequencing means that it is now feasible and inexpensive to sequence large numbers of genomes. Indeed, a 1000 Genomes Project for humans was announced in January 2008 [33], and the first results of this initiative are very encouraging [34,35]. It builds, in a manner similar to the *A. thaliana* project, on previous HapMap information, but because of the greater complexity and repetitiveness of human genomes, much of the initial effort for the human project will go towards comparing the feasibility of different approaches. In contrast, even short reads of the *A. thaliana* sequence, such as those produced by the first generation of Illumina's Genome Analyzer instrument, have already been proved to support not only the discovery of SNPs, but also of short to medium-size indels, including the detection of sequences not present in the reference genome [29].

We are proposing a hierarchical strategy to sequence the species-wide genome of *A. thaliana*. The first aspect of this approach is to make use of different technologies and different depths of sequencing coverage. A small number of genome sequences that approach the quality of the original Col-0 reference will be generated by exploiting mostly technologies such as Roche's 454 platform, which generates longer reads, in combination with libraries of different insert sizes, allowing long-range assembly. A much larger number of genomes will be sequenced with a less expensive technology such as Illumina's Genome Analyzer or Applied Biosystems' SOLiD and with only a single type of clone library. For this set of accessions, local haplotype similarity will be exploited in combination with information from the reference genomes to deduce the complete sequence, using methods similar those employed in inbred strains of mice [36]. The power of this approach is in the large number of accessions that can be sequenced. For example, even if a particular haplotype is only present at 1% frequency, and each of the 1,001 strains is only sequenced at 8x coverage, there would still be on average 80 reads for each site in this haplotype.

The second aspect of the hierarchical approach will be the sampling of ten individuals from ten populations each in ten geographic regions throughout Eurasia, plus at least one North African accession ($10 \times 10 \times 10 + 1$) (see Figure 1a). We expect individuals from the same region to show more extensive haplotype sharing than is observed in worldwide samples [4,24], which will be advantageous for the imputation strategy discussed above. An argument that might be raised against this approach is the strong population structure it entails, but we note that it is probably impossible to sample accessions in a manner that avoids population structure completely, and that our strategy will allow us to address questions of local adaptation, which are of great interest to evolutionary scientists. The output of the 1001 Genomes project will be a generalized genome sequence that encompasses every *A. thaliana* accession analysed as a special case. It will comprise a mosaic of variable haplotypes such that every genome can be aligned completely against it.

It is instructive to compare our proposal with the 1000 Genomes effort for humans [37] and the *Drosophila* Genetic Reference Panel projects [38]. Because *A. thaliana* accessions are inbred with effectively constant genomes, and can be readily distributed as seeds, the genome sequence data we generate can be used directly in association mapping; of particular importance, the causative mutations will be observed in most cases. In contrast, the human population is not made up of highly inbred individuals, and the genetic variation discovered in 1000 humans is only a first step, yielding a deep catalog of genetic variation that allows one to infer indirectly much of the genome sequence in the samples used in association studies [33]. The *A. thaliana* 1001 Genomes project is relatively simple compared with its bigger human cousin, and much more affordable because *A. thaliana* genomes are about 20 times smaller than human genomes (40 times, if one counts both homologs in the outbred genomes of our species). Consequently, the powerful arguments that justified funding the human effort are even more persuasive in the case of *A. thaliana*. Indeed, the reasoning for the *Drosophila* Genetic Reference Panel [38] spearheaded by Trudy Mackay is very similar to that advanced for the *A. thaliana* project. Important differences are, however, that *Drosophila melanogaster* does not self-fertilize. Inbred lines therefore have to be derived by repeated brother-sister matings, and although they capture variation present in nature, wild individuals are genetically more complex. Moreover, the initial *Drosophila* 192 lines, which are the focus of this project, were collected from a single locale, in contrast to the much wider sampling for both the human and the *A. thaliana* projects.

Some of the *A. thaliana* genomes will be immediately useful, as they are from parents of recombinant inbred line populations, a widely used resource for QTL mapping in *A. thaliana* [10]. The genome sequences will provide information on

potential functional polymorphisms responsible for the identified QTL.

The main motivation for the 1001 Genomes project is, however, to enable GWA studies in this species. The seeds from the 1,001 accessions will be freely available from the *Arabidopsis* stock centers [39], and each accession can be grown and phenotyped by scientists from all over the world, in as many environments as desired. Importantly, because an unlimited supply of genetically identical individuals will be available for each accession, even subtle phenotypes and ones that are highly sensitive to the microenvironment, which is often difficult to control, can be measured with high confidence. The phenotypes will include morphological analyses, such as plant stature, growth and flowering; investigations of plant content, such as metabolites and ions; responses to the abiotic environment, such as resistance to drought or salt stress; or resistance to disease caused by a host of prokaryotic and eukaryotic pathogens, from microbes to insects and nematodes. In the last case, a particularly exciting prospect is the ability to identify plant genes that mediate the effects of individual pathogen proteins, which are normally delivered as a complex mix to the plant, as is being done in the Effectoromics project, which has the aim of “understanding host plant susceptibility and resistance by indexing and deploying obligate pathogen effectors” [40]. The value of being able to correlate many different phenotypes, including genome-wide phenotypes, has already been beautifully demonstrated for the *Drosophila* Genetic Reference Panel [41], and we expect similar dividends for the *A. thaliana* project.

We envisage that ultimately there will be web-based tools for GWA scans to identify candidate polymorphisms affecting these phenotypes in the 1,001 accessions. As part of the *Arabidopsis* 2010 Project, the US National Science Foundation is already supporting the development of web resources that will help the wider community to exploit such sequence data [42]. It goes without saying that one needs to employ appropriate statistical methods to control for population structure caused by the hierarchical choice of accessions, which might otherwise produce false-positive associations.

A potential shortcoming of GWA scans is that some alleles responsible for interesting traits are strongly partitioned between different populations. They are in strong LD with many physically unlinked loci and thus difficult to pinpoint. A powerful approach to circumvent such problems of population structure is the generation of experimental populations in which members of different populations are intercrossed in a systematic way. Such a strategy, dubbed nested association mapping (NAM), has been developed for maize [43], and similar designs are being used in mice [44,45]. Corresponding efforts are under way for *A. thaliana* as well [46]. As part of the 1001 Genomes Project, the

parental accessions in these lines are already being sequenced, which will enable the reconstruction of complete haplotype maps in the hundreds of derived intercrossed lines, which need to be characterized at only a relatively modest number of informative SNPs. Association scans with this material will provide an extremely useful complement to conventional GWA. In future phenotyping projects, it might be advisable to split efforts between wild accessions and the intercrossed lines.

This leaves the question: why 1,001 genomes, and not 101 or 10,001? As with the human 1000 Genomes project, 1,001 is obviously an arbitrarily chosen number, to capture the imagination of our colleagues (and of the funding agencies). Some might argue that rather than sequencing 1,001 *A. thaliana* accessions, one should sequence, say, 200 *A. thaliana* strains and 200 rice strains. Our answer is that we see the *A. thaliana* 1001 Genomes project only as a first feasibility study, and that we are fully expecting similar projects for rice and other crops to follow soon. The dawn of a new era of plant genetics is truly upon us.

Acknowledgements

We thank our many colleagues around the world, including Joe Ecker (Salk Institute), Wolf Frommer and Len Penacchio (JGI and JBEI), Christian Hardtke (Lausanne), Jonathan Jones (Sainsbury Laboratory), Todd Michael (Waksman Institute), and Magnus Nordborg (USC/GMI), for contributing to the 1001 Genomes vision. *Arabidopsis thaliana* sequencing efforts in our labs are supported by the BBSRC (RM), BMBF (ERA-PG ARABRAS and GABI-GNADE), a Gottfried Wilhelm Leibniz Award (DFG) and the Max Planck Society (DW).

References

1. 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.
2. Hoffmann MH: **Biogeography of *Arabidopsis thaliana* (L.) Heynh. (Brassicaceae).** *J Biogeography* 2002, **29**:125-134.
3. Schmid KJ, Torjek O, Meyer R, Schmuths H, Hoffmann MH, Altmann T: **Evidence for a large-scale population structure of *Arabidopsis thaliana* from genome-wide single nucleotide polymorphism markers.** *Theor Appl Genet* 2006, **112**:1104-1114.
4. Nordborg M, Hu TT, Ishino Y, Jhaveri J, Toomajian C, Zheng H, Bakker E, Calabrese P, Gladstone J, Goyal R, Jakobsson M, Kim S, Morozov Y, Padhukasahasram B, Plagnol V, Rosenberg NA, Shah C, Wall JD, Wang J, Zhao K, Kalbfleisch T, Schulz V, Kreitman M, Bergelson J: **The pattern of polymorphism in *Arabidopsis thaliana*.** *PLoS Biol* 2005, **3**:e196.
5. François O, Blum MG, Jakobsson M, Rosenberg NA: **Demographic history of European populations of *Arabidopsis thaliana*.** *PLoS Genet* 2008, **4**:e1000075.
6. Chory J, Ecker JR, Briggs S, Caboche M, Coruzzi GM, Cook D, Dangl J, Grant S, Guerinot ML, Henikoff S, Martienssen R, Okada K, Raikhel NV, Somerville CR, Weigel D: **National Science Foundation-Sponsored Workshop Report: “The 2010 Project” functional genomics and the virtual plant. A blueprint for understanding how plants are built and how to improve them.** *Plant Physiol* 2000, **123**:423-426.
7. Somerville C, Koornneef M: **A fortunate choice: the history of *Arabidopsis* as a model plant.** *Nat Rev Genet* 2002, **3**:883-889.
8. Kobayashi Y, Weigel D: **Move on up, it's time for change—mobile signals controlling photoperiod-dependent flowering.** *Genes Dev* 2007, **21**:2371-2384.
9. The *Arabidopsis* Genome Initiative: **Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*.** *Nature* 2000, **408**:796-815.

10. Koornneef M, Alonso-Blanco C, Vreugdenhil D: **Naturally occurring genetic variation in *Arabidopsis thaliana***. *Annu Rev Plant Biol* 2004, **55**:141-172.
11. Mitchell-Olds T, Schmitt J: **Genetic mechanisms and evolutionary significance of natural variation in *Arabidopsis***. *Nature* 2006, **441**:947-952.
12. 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.
13. Baxter I, Muthukumar B, Park HC, Buchner P, Lahner B, Danku J, Zhao K, Lee J, Hawkesford MJ, Guerinot ML, Salt DE: **Variation in molybdenum content across broadly distributed populations of *Arabidopsis thaliana* is controlled by a mitochondrial molybdenum transporter (*MOT1*)**. *PLoS Genet* 2008, **4**:e1000004.
14. Bentsink L, Jowett J, Hanhart CJ, Koornneef M: **Cloning of *DOGI*, a quantitative trait locus controlling seed dormancy in *Arabidopsis***. *Proc Natl Acad Sci USA* 2006, **103**:17042-17047.
15. Lempe J, Balasubramanian S, Sureshkumar S, Singh A, Schmid M, Weigel D: **Diversity of flowering responses in wild *Arabidopsis thaliana* strains**. *PLoS Genet* 2005, **1**:109-116.
16. Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovskiy LA, Feldman MW: **Genetic structure of human populations**. *Science* 2002, **298**:2381-2385.
17. Hinds DA, Stuve LL, Nilsen GB, Halperin E, Eskin E, Ballinger DG, Frazer KA, Cox DR: **Whole-genome patterns of common DNA variation in three human populations**. *Science* 2005, **307**:1072-1079.
18. The International HapMap Consortium: **A haplotype map of the human genome**. *Nature* 2005, **437**:1299-1320.
19. Conrad DF, Jakobsson M, Coop G, Wen X, Wall JD, Rosenberg NA, Pritchard JK: **A worldwide survey of haplotype variation and linkage disequilibrium in the human genome**. *Nat Genet* 2006, **38**:1251-1260.
20. The International HapMap Consortium: **A second generation human haplotype map of over 3.1 million SNPs**. *Nature* 2007, **449**:851-861.
21. Clark RM, Schweikert G, Toomajian C, Ossowski S, Zeller G, Shinn P, Warthmann N, Hu TT, Fu G, Hinds DA, Chen H, Frazer KA, Huson DH, Schölkopf B, Nordborg M, Rätsch G, Ecker JR, Weigel D: **Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana***. *Science* 2007, **317**:338-342.
22. Zeller G, Clark RM, Schneeberger K, Bohlen A, Weigel D, Rätsch G: **Detecting polymorphic regions in the *Arabidopsis thaliana* genome with resequencing microarrays**. *Genome Res* 2008, **18**:918-929.
23. McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP, Hirschhorn JN: **Genome-wide association studies for complex traits: consensus, uncertainty and challenges**. *Nat Rev Genet* 2008, **9**:356-369.
24. 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.
25. Beck JB, Schmuths H, Schaal BA: **Native range genetic variation in *Arabidopsis thaliana* is strongly geographically structured and reflects Pleistocene glacial dynamics**. *Mol Ecol* 2008, **17**:902-915.
26. **Genome wide association mapping in *Arabidopsis thaliana* (NIH R01 GM073822)** [<http://naturalvariation.org/hapmap>]
27. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, et al.: **The sequence of the human genome**. *Science* 2001, **291**:1304-1351.
28. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, et al.: **Initial sequencing and analysis of the human genome**. *Nature* 2001, **409**:860-921.
29. 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.
30. Korbel JO, Urban AE, Affourtit JP, Godwin B, Grubert F, Simons JF, Kim PM, Palejev D, Carriero NJ, Du L, Taillon BE, Chen Z, Tanzer A, Saunders AC, Chi J, Yang F, Carter NP, Hurler ME, Weissman SM, Harkins TT, Gerstein MB, Egholm M, Snyder M: **Paired-end mapping reveals extensive structural variation in the human genome**. *Science* 2007, **318**:420-426.
31. Sebat J: **Major changes in our DNA lead to major changes in our thinking**. *Nat Genet* 2007, **39**:S3-5.
32. **1001genomes.org** [<http://www.1001genomes.org>]
33. Kaiser J: **DNA sequencing. A plan to capture human diversity in 1000 genomes**. *Science* 2008, **319**:395.
34. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR, Boutell JM, Bryant J, Carter RJ, Keira Cheetham R, Cox AJ, Ellis DJ, Flatbush MR, Gormley NA, Humphray SJ, Irving LJ, Karbelashvili MS, Kirk SM, Li H, Liu X, Maisinger KS, Murray LJ, Obradovic B, Ost T, Parkinson ML, Pratt MR, et al.: **Accurate whole human genome sequencing using reversible terminator chemistry**. *Nature* 2008, **456**:53-59.
35. Wang J, Wang W, Li R, Li Y, Tian G, Goodman L, Fan W, Zhang J, Li J, Zhang J, Guo Y, Feng B, Li H, Lu Y, Fang X, Liang H, Du Z, Li D, Zhao Y, Hu Y, Yang Z, Zheng H, Hellmann I, Inouye M, Pool J, Yi X, Zhao J, Duan J, Zhou Y, et al.: **The diploid genome sequence of an Asian individual**. *Nature* 2008, **456**:60-65.
36. Szatkiewicz JP, Beane GL, Ding Y, Hutchins L, Pardo-Manuel de Villena F, Churchill GA: **An imputed genotype resource for the laboratory mouse**. *Mamm Genome* 2008, **19**:199-208.
37. **1000 genomes** [<http://www.1000genomes.org>]
38. **The Drosophila Genetic Resource Panel (DRGP)** [<http://tinyurl.com/192flies>]
39. **TAIR** [<http://www.arabidopsis.org>]
40. **Effectoromics** [<http://tinyurl.com/effectoromics>]
41. Ayroles JF, Carbone MA, Stone EA, Jordan KVV, Lyman RF, Magwire MM, Rollmann SM, Duncan LH, Lawrence F, Anholt RR, Mackay TF: **Systems genetics of complex traits in *Drosophila melanogaster***. *Nat Genet* 2009, **41**:299-307.
42. **Collaborative research award: An Arabidopsis Polymorphism Database** [<http://tinyurl.com/nsf0723510>]
43. Yu J, Holland JB, McMullen MD, Buckler ES: **Genetic design and statistical power of nested association mapping in maize**. *Genetics* 2008, **178**:539-551.
44. Valdar W, Solberg LC, Gauguier D, Burnett S, Klenerman P, Cookson WO, Taylor MS, Rawlins JN, Mott R, Flint J: **Genome-wide genetic association of complex traits in heterogeneous stock mice**. *Nat Genet* 2006, **38**:879-887.
45. Churchill GA, Airey DC, Allayee H, Angel JM, Attie AD, Beatty J, Beavis WD, Belknap JK, Bennett B, Berrettini W, Bleich A, Bogue M, Broman KW, Buck KJ, Buckler E, Burmeister M, Chesler EJ, Cheverud JM, Clapcote S, Cook MN, Cox RD, Crabbe JC, Crusio WE, Darvasi A, Deschepper CF, Doerge RW, Farber CR, Forejt J, Gaile D, Garlow SJ, et al.: **The Collaborative Cross, a community resource for the genetic analysis of complex traits**. *Nat Genet* 2004, **36**:1133-1137.
46. **Mapping complex traits in Recombinant Inbred lines of heterogeneous stocks of *A. thaliana*** [<http://tinyurl.com/kover>]