Genome-wide mapping with biallelic markers in *Arabidopsis thaliana*

Raymond J. Cho[1*], Michael Mindrinos[2,3*], Daniel R. Richards[1], Ronald J. Sapolsky[4], Mary Anderson[5], Eliana Drenkard[3], Julia Dewdney[3], T. Lynne Reuber[3], Melanie Stammers[6], Nancy Federspiel[7], Athanasios Theologis[8], Wei-Hsien Yang[2], Earl Hubbell[9], Melinda Au[7], Edward Y. Chung[7], Deval Lashkari[10], Bertrand Lemieux[11], Caroline Dean[6], Robert J. Lipshutz[9], Frederick M. Ausubel[3], Ronald W. Davis[1,2] & Peter J. Oefner[2]

*These authors contributed equally to this work.

Single-nucleotide polymorphisms, as well as small insertions and deletions (here referred to collectively as simple nucleotide polymorphisms, or SNPs), comprise the largest set of sequence variants in most organisms[1,2]. Positional cloning based on SNPs may accelerate the identification of human disease traits and a range of biologically informative mutations[3–6]. The recent application of high-density oligonucleotide arrays to allele identification has made it feasible to genotype thousands of biallelic SNPs in a single experiment[3,7]. It has yet to be established, however, whether SNP detection using oligonucleotide arrays can be used to accelerate the mapping of traits in diploid genomes. The cruciferous weed *Arabidopsis thaliana* is an attractive model system for the construction and use of biallelic SNP maps. Although important biological processes ranging from fertilization and cell fate determination[8–11] to disease resistance[12,13] have been modelled in *A. thaliana*, identifying mutations in this organism has been impeded by the lack of a high-density genetic map consisting of easily genotyped DNA markers[14]. We report here the construction of a biallelic genetic map in *A. thaliana* with a resolution of 3.5 cM and its use in mapping *Eds16*, a gene involved in the defence response to the fungal pathogen *Erysiphe orontii.* Mapping of this trait involved the high-throughput generation of meiotic maps of F_2 individuals using high-density oligonucleotide probe array-based genotyping. We developed a software package called InterMap and used it to automatically delimit *Eds16* to a 7-cM interval on chromosome 1. These results are the first demonstration of biallelic mapping in diploid genomes and establish means for generalizing SNP-based maps to virtually any genetic organism.

Recently, a genome-wide biallelic linkage map of *Saccharomyces cerevisiae* has been constructed based on the direct hybridization of whole genomic DNA from two polymorphic strains to high-density oligonucleotide arrays sampling every known yeast ORF (ref. 4). In organisms with larger genomes, however, including *A. thaliana*, the signal intensity from hybridization of whole genomic DNA decreases, making the direct application of this approach more challenging (M.M. and E. Winzeler, unpublished

Table 1 • Characterization of SNPs identified in *A. thaliana*

Nature of SNP markers	% of AT412 SNPs	Total number
identified by DHPLC/dideoxysequencing		487
formatted on AT412 oligonucleotide array		412
successful PCR amplification	95	390
high hybridization signal following singleplex amplification	85	351
high hybridization signal following multiplex amplification	81	332
discrimination between Columbia and Landsberg *erecta* homozygotes	64	262
discrimination between homozygotes and heterozygotes	57	235
discrimination between homozygotes and mapping to unique chromosomal position	58	237

Successive tests used to select SNPs for final biallelic map. Each selection was only performed on markers passing the previous selection, unless otherwise noted. All 412 SNPs chosen for incorporation into an oligonucleotide array were PCR amplified individually from genomic DNA and examined on an agarose gel to ensure generation of a single product. We successfully amplified 390 loci, or 95% of the markers: 22 of the loci failed to amplify under these conditions, the primary cause being an inability to design optimal PCR primers within limitations on the maximum amplicon size. Of 390 markers, 351 were found to pass both the intensity and discrimination tests. From the 351 markers that passed these tests when amplified separately, 332 were found to pass the same test when amplified in multiplex PCR reactions. Based on results from array hybridization, 48 of these markers failed to show any discrimination between the Columbia and Landsberg *erecta* ecotypes. Resequencing of these markers indicated that the polymorphism had been identified correctly in 94% of cases. An additional 22 markers displayed weak discrimination between the two ecotypes, leaving 262 markers amenable for genetic mapping. From the 262 SNPs that discriminated homozygotes, 237 markers distinguished heterozygous genotypes from the two parental homozygotes genotypes. These markers were mapped unambiguously to the *A. thaliana* genome using MAPMAKER software[20] and used to generate a biallelic map.

Departments of [1]Genetics and [2]Biochemistry, Stanford University School of Medicine, Stanford, California 94305, USA. [3]Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. [4]Genencor, 925 Page Mill Road, Palo Alto, California 94304, USA. [5]Nottingham Arabidopsis Stock Centre, Division of Plant Science (University Park), University of Nottingham, Nottingham NG7 2RD, UK. [6]Department of Molecular Genetics, John Innes Centre, Colney Lane, Norwich NR4 7UH, UK. [7]Stanford DNA Sequencing and Technology Center, Palo Alto, California 94304, USA. [8]Plant Gene Expression Center Genome Sequencing Lab, Albany, California 94710, USA. [9]Affymetrix, Inc., Santa Clara, California 95051, USA. [10]Incyte Pharmaceuticals, Fremont, California 94555, USA. [11]Department of Biology, York University, North York, Ontario, Canada. Correspondence should be addressed to P.J.O. (e-mail: oefner@genome.stanford.edu).

© 1999 Nature America Inc. • http://genetics.nature.com

© 1999 Nature America Inc. • http://genetics.nature.com

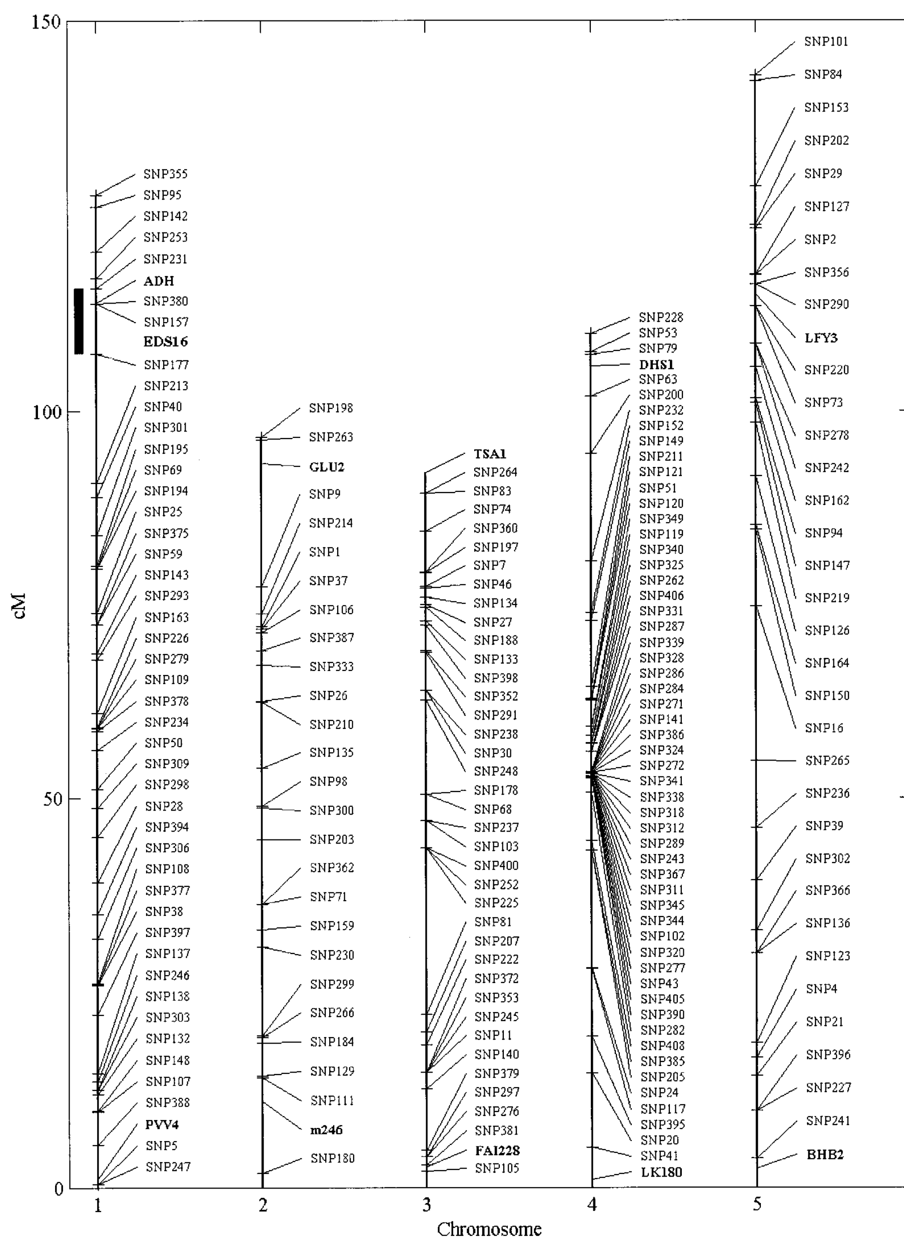

Fig. 1 Genetic positions of the 237 SNP markers used to define a biallelic SNP map for *A. thaliana*. A number of conventional markers are also shown to provide orientation on each chromosome. The name, the exact position of each SNP marker and the sequence surrounding the polymorphism are available (http://nasc.nott.ac.uk/new_ri_map.html). We identified 34 SNP markers for high-resolution recombination studies in a 150-kb region on chromosome 4, between 52 cM and 53 cM.

data). Therefore, to generate high-density biallelic maps in most eukaryotic genomes, we investigated an alternative method for identifying and assaying a large set of well-distributed SNPs.

On the basis of DNA sequence information obtained from the *A. thaliana* EST database and random shotgun sequencing, we amplified more than 1,800 PCR products, averaging 280 bp, from both the Columbia and Landsberg *erecta* ecotypes, naturally occurring polymorphic strains of *A. thaliana* that have been widely used for genetic mapping[14,15]. We assayed these sequences for the presence of variation by formation of Columbia-Landsberg PCR product heteroduplexes, which were then subjected to denaturing HPLC (DHPLC) analysis[16,17]. PCR products identified by DHPLC as containing a potential variation were sequenced on both strands to determine the nature of the polymorphism. Using this method, we identified 487 loci containing SNPs between the Columbia and Landsberg *erecta* ecotypes. We screened more than 500,000 kb of total DNA sequence for SNPs, representing nearly 0.5% of the *A. thaliana* genome.

To generate an oligonucleotide array specifically designed to genotype the identified *A. thaliana* SNPs, we initially screened the 487 polymorphic sequences for optimal melting temperatures and the absence of AT-rich regions in the vicinity of the polymorphic base. On the basis of this information, we formatted 412 polymorphisms into detection blocks on a high-density oligonucleotide probe array, named AT412, using the variant detector array (VDA) approach[3,5]. The AT412 array contains 1,648 VDAs, each consisting of forty-four 25-mer probes interrogating each of the two alleles and the sequence adjacent to the polymorphic site. A subset of these SNP markers can also be genotyped using co-dominant allele-specific PCR amplification and conventional gel-electrophoretic methods (E.D., M.M. and F.M.A., unpublished data). Sequence information and further details concerning these polymorphisms are available (http://genome-www.stanford.edu/Arabidopsis/maps/SNP.html).

To determine the ability of the AT412 array to genotype SNP markers, all 412 corresponding polymorphic loci were amplified from Columbia and Landsberg genomic DNA, PCR-labelled with

© 1999 Nature America Inc. • http://genetics.nature.com

© 1999 Nature America Inc. • http://genetics.nature.com

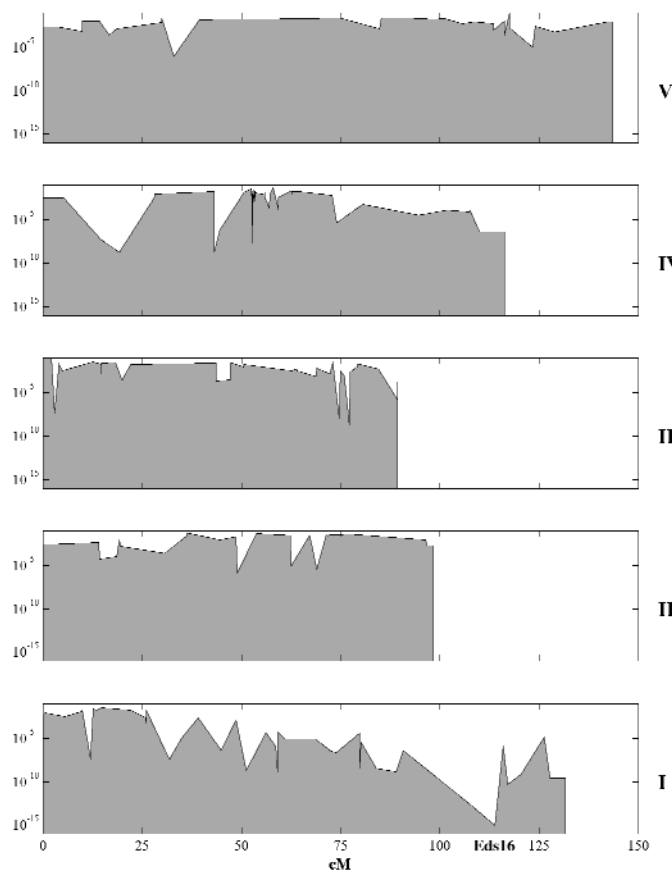

Fig. 2 Calculated probability of random segregation in the *A. thaliana* genome using 28 F_2 plants from the *Eds16-1*×Landsberg *erecta* cross that display a susceptibility phenotype to the fungal pathogen pathogen *Erysiphe orontii*. The *x* axis represents genetic position, whereas the *y* axis (log base 10) indicates the probability of random segregation. The approximate position of *Eds16*, as determined by biallelic mapping, is indicated by a filled bar.

biotin, hybridized to arrays and stained with a phycoerythrin-streptavidin conjugate. We quantitated the resulting fluorescence intensities from each array using a confocal laser scanning instrument. To facilitate SNP-based mapping using data derived from hybridization to oligonucleotide arrays, we developed a set of computer programs collectively named InterMap. The InterMap package uses an AWK script to call alleles at individual VDAs based on fluorescence intensities, assigns genotypes based on comparison with known homozygote fluorescence intensities, and determines genetic linkage by calculating the probability of observing a segregation pattern by chance at each marker. This software is freely available for download (http://genome-www.stanford.edu/Arabidopsis/maps/SNP.html).

Based on analysis using InterMap, 262 of 412 (64%) markers displayed clear discrimination between Columbia and Landsberg homozygous plants, and 235 (58%) markers discriminated against the heterozygote (Table 1). Sequences that did not result in successful discrimination on the array were dideoxysequenced a second time from both Columbia and Landsberg ecotypes. Of 48 markers that did not display discrimination, 45 (94%) were shown to have been sequenced correctly, confirming that failure of markers is attributable to lack of discrimination by the array. It has been shown that more extensive optimization of conditions and array design will increase the percentage of functional markers; however, sufficient markers displayed discrimination for the purposes of this study[18].

We constructed a linkage map of the SNP markers that discriminate between Columbia and Landsberg (Fig. 1) by obtaining segregation data for all the available markers in 68 recombinant inbred (RI) lines derived from a cross between Columbia and Landsberg *erecta*[19]. Specifically, the 262 markers displaying discrimination between homozygotes were amplified from each strain and hybridized to AT412 arrays. The resulting fluorescence data was processed using InterMap and then analysed with MAPMAKER software[20] to construct a biallelic map. We localized 237 of these SNP markers to unique chromosomal positions and integrated them into the existing *A. thaliana* RI linkage map[19,20]. Of these 237 markers, 226 also displayed reliable discrimination against the heterozygote. This integrated map is available (http://nasc.nott.ac.uk/new_ri_map.html). The average resolution of the linkage map obtained from these SNP markers is greater than 3.5 cM, and the largest gap between markers is approximately 15 cM.

To assess the feasibility of whole-genome mapping using biallelic markers, we localized the *A. thaliana* mutation *Eds16-1*, which causes enhanced susceptibility to the fungal pathogen *Erysiphe orontii*. *Eds16-1* was shown to segregate as a single recessive mendelian trait by backcrossing to the wild-type Columbia parent. For mapping purposes, *Eds16-1* was crossed to Landsberg *erecta*, and 28 F_2 plants were collected that displayed the susceptibility phenotype. For each of the 28 F_2 plants, all 412 SNP loci were amplified in multiplexed PCR reactions, labelled, hybridized to an AT412 array and stained. Fluorescence readings from these arrays were analysed using the InterMap mapping function, which automatically localized *Eds16* to a 7-cM interval on the bottom of chromosome 1 between markers SNP 177 and SNP 231 (Fig. 2). *Eds16* was independently mapped to approximately the same location using a different mapping procedure based on cleaved amplified polymorphic sequences[21] (CAPS).

Whole-genome biallelic mapping in the *A. thaliana* genome required less than two days following harvest of F_2 individuals. Currently, too few PCR-based markers exist in *A. thaliana* for whole-genome mapping using conventional methods, and the use of RFLPs to achieve comparable results would take several months. These results demonstrate that whole-genome mapping using array-based genotyping of SNPs can accelerate the process of positional cloning. The generation of denser biallelic maps should allow high-throughput identification of both monogenic and polygenic traits, effectively removing the rate-limiting nature of high-resolution mapping from the study of biological processes.

Methods

DHPLC variation scanning. We isolated genomic *A. thaliana* DNA from nuclei preparations through cesium chloride gradient purification as described[22]. Amplimers derived from shotgun sequence and the EST database were used to amplify loci from both Columbia and Landsberg *erecta* DNA. Standard PCR reactions were performed in a volume (20 µl) containing *A. thaliana* DNA (20 ng), primers (0.15 µM each; Gibco-BRL), AmpliTaq Gold DNA polymerase (1 U; Perkin-Elmer), dNTPs (250 µM; New England Biolabs), Tris-HCl (10 mM, pH 8.3; Gibco-BRL), KCl (50 mM) and $MgCl_2$ (2.5 mM; Sigma). Thermocycling was performed with initial denaturation/activation at 95 °C for 10 min, followed by 30 cycles of 95 °C for 30 s, 52 °C for 45 s and 72 °C for 1 min. To generate heteroduplexes for DHPLC scanning, we combined corresponding PCR products from

© 1999 Nature America Inc. • http://genetics.nature.com

each ecotype at roughly equimolar ratios based on agarose gel electrophoresis of the amplicons, denatured at 95 °C for 3 min and slowly reannealed by decreasing the temperature from 95 °C to 65 °C over 30 min, with final cooling to 6 °C. Each heteroduplex was run on an automated DHPLC instrument (Transgenomic) at 50 °C, 56 °C and 62 °C, respectively. DHPLC profiles were visually examined for the presence of two or more peaks, indicative of SNPs. With dye terminator sequencing, more than 95% of the sequences identified by DHPLC scanning were confirmed to carry mutations between the two strains.

Design and construction of the AT412 array. We synthesized high-density oligonucleotide arrays using parallel light-directed oligonucleotide synthesis as described[23,24] (Affymetrix). The AT412 genotyping array contains a detection block for each of the 412 biallelic markers. Each detection block array consists of four variant detector arrays (VDAs) corresponding to the alternative alleles: two for the forward strand sequence and two for the reverse strand sequence. For each of 11 interrogated positions on either strand (–5 to +5 relative to the polymorphic base), the VDA for each allele has a set of four 25-mer oligonucleotide probes, complementary to the SNP sequence, that are called a 'miniblock'. These four probes only differ at the central, interrogated position, which is substituted by each of the four nucleotides. Hybridization of DNA from a perfectly matching allele results in a strong signal at the corresponding probe site in the VDA.

To determine whether the chip could reliably identify each locus, we defined threshold tests based on specificity (signal strength) and sensitivity (discrimination) as described in the InterMap sections below.

Amplification of biallelic marker loci. For genotyping, we extracted genomic *A. thaliana* DNA from one or two leaves of plant material as described[22]. For each of the 412 loci, we chose PCR primers with similar calculated melting temperatures to amplify a fragment as close as possible to the polymorphic base, but not to exceed 120 bp. T7 sequence was incorporated at the 5′ end of forward primers (5′–TAATACGACTCACTATAGGGAGA...genomic seq–3′), whereas T3 sequence was incorporated at the 5′ end of reverse primers (5′–AATTAACCCTCACTAAAGGGAGA...genomic seq–3′). In this way, these 'tailed' PCR products were labelled in a secondary reaction using 5′-biotinylated T7 and T3 primers. Singleplex PCR reactions were performed in a volume (30 µl) containing *A. thaliana* DNA (30 ng), primers (0.005–0.01 µM each), AmpliTaq Gold DNA polymerase (1 U), dNTPs (250 µM), Tris-HCl (10 mM, pH 8.3), KCl (50 mM) and $MgCl_2$ (7.5 mM). Thermocycling was performed in a Perkin Elmer GeneAmp PCR System 9600, with initial denaturation/activation at 95 °C for 10 min, followed by a cycle of denaturation at 95 °C for 30 s, annealing at 63 °C for 1 min and primer extension at 72 °C for 1 min. We performed 16 additional cycles identical to the first, except the primer annealing temperature was decreased 0.5 °C in each consecutive cycle. We then performed 30 additional cycles at 95 °C for 30 s, 56 °C for 45 s and 72 °C for 1 min. A final extension reaction was performed at 72 °C for 10 min. This initial reaction was secondarily amplified using 5′-biotinylated T7 and T3 primers. The secondary PCR reaction was performed in a total volume (30 µl), with the multiplex reaction (0.6 µl), biotinylated primers (0.15 µM each), AmpliTaq Gold DNA polymerase (1 U), dNTPs (250 µM), Tris-HCl (10 mM, pH 8.3), KCl (50 mM) and $MgCl_2$ (25 mM). Thermocyling was performed with initial denaturation/activation at 95 °C for 10 min, followed by 30 cycles at 95 °C for 30 s, 52 °C for 45 s and 72 °C for 1 min. We carried out a final extension reaction at 72 °C for 10 min. We successfully amplified 390 loci, representing 95% of the biallelic markers chosen for VDA development, in singleplex reactions. Of 412 loci, 22 failed to amplify under these conditions: for most of these loci, sequence characteristics precluded the generation of optimal PCR primers within the given size limitations.

We designed a multiplex PCR strategy to reduce time and labour in amplification of polymorphic loci. Multiplex reactions were identical to the singleplex reactions described above, except that on average, each multiplex reaction contained 45 discrete primer pairs. The ten multiplex reactions (30 µl) for each individual were then pooled, and this sample (76 µl) was used for array hybridization. This multiplex strategy was shown to maintain discrimination for more than 95% of markers that distinguished well between homozygotes in singleplex assay, comparable to results from the human genotyping array[3]. Critical multiplexing parameters included the identification of loci with similar amplification efficiencies (as observed with agarose gel electrophoresis), minimization of the concentration of each primer and the use of a higher magnesium concentration (7.5 mM) to match increased nucleic acid in the initial reaction.

Hybridization to arrays. The probe array was pre-wetted with 6×SSPET (0.9 M NaCl, 60 mM NaH_2PO_4, 6 mM EDTA, pH 7.4, 0.005% Triton X-100 (Sigma)) at 25 °C for 10 min, and washed once with tetramethylammonium-chloride (3 M; Sigma) for 5 min. The biotinylated sample was denatured at 95 °C for 5 min, immediately cooled on ice and brought to a 200 µl final volume with addition of hybridization cocktail (3 M tetramethylammonium chloride, 10 mM Tris-HCl (pH 8.0), 1 mM EDTA, 0.01% Triton X-100, 0.1 mg/ml herring sperm DNA (Promega) and 200 pM control oligomer). The 200 µl final volume was applied to the array at 42 °C for 12–15 h on a rotisserie at ~60 rpm.

Washing and staining of arrays. The probe array was washed on a fluidics station (FS400, Affymetrix) 10 times with 6×SSPET at 22 °C, then 3 times with 1×SSPET, followed by staining at RT with staining solution (2 mg/ml streptavidin R-phycoerythrin (Molecular Probes) and 0.5 mg/ml acetylated BSA (Sigma) in 6×SSPET). After staining, the probe array was washed 5 times again with 6×SSPET at 22 °C.

Array scanning and data analysis. We scanned the probe array on the GeneArray scanner (HP G2500A, Affymetrix) at ~40 to 80 pixels per probe feature with a 570 nm filter. Affymetrix GeneChip software was used to generate a digitized intensity table for each of the features on the chip. This signal data and positional information was then processed sequentially by software in the InterMap analysis package.

InterMap1 (AWK script for calling *A. thaliana* marker polymorphism). The InterMap software package provides three distinct SNP-based mapping applications (InterMap1, InterMap2 and InterMap3), described here in detail. As described above, for each miniblock, there is one perfectly matching probe (PM) and three mismatch probes (MM) at the central base of the 25-mer. The usefulness of a miniblock is determined both by signal intensity and by discrimination. Regardless of the genotype, we expect at least one VDA (in the case of the homozygotes) or both VDAs (in the case of a heterozygote) will have high signal and high discrimination. Therefore, a miniblock is declared 'valid' if both PM-MM for allele A or PM-MM for allele B is greater than some signal threshold value, and PM/MM for allele A or PM/MM for allele B is greater than some discrimination threshold value (in both these 'equations', MM is taken to be the maximum of the three available MM values in the relevant miniblock). The two thresholds are determined from the background noise for a given chip, which in turn is determined from the mean intensities and standard deviation of the 2% faintest intensity features on the chip. To call a genotype with confidence, a minimum fraction of miniblocks between the VDAs must be declared valid. If such a quorum exists, then we determine a $\hat{P}(A)$ value for each valid miniblock. $\hat{P}$ is the probability of A-allelic character of the genotype: for AA homozygotes, a perfect approaches 1; for BB homozygotes, the value approaches 0; for heterozygotes, the value is near 0.5. $\hat{P}(A)$ is determined as max $[0, PM_A–MM_A]/(\max[0.001, PM_A–MM_A] + \max[0.001, PM_B–MM_B]$, where the max function is the larger of the two values in the square brackets (in this way, negative and non-positive values of PM-MM are avoided). The mean is calculated for all the valid miniblocks, along with the standard deviation of that mean. This experimental value is compared with the range of known values determined for the three genotypes from reference samples. This statistical analysis has been used effectively in the genotyping of thousands of human SNPs (ref. 3). In these experiments, more than 90% of the 237 markers were called unambiguously for all of the 96 samples tested.

InterMap2. The $\hat{P}(A)$s of the homozygous RI lines for each marker were binned based on their median after discarding up to two outliers. The mean and variance of the $\hat{P}(A)$s in each bin was computed, and mutants were scored as heterozygous if their $\hat{P}(A)$ was further than three standard deviations away from the expected signal response of both homozygous genotypes. Markers which could not score heterozygotes were excluded from the analysis.

InterMap3. The probability of observing a segregation pattern by chance at each marker was computed by:

$$p=(1/2)^{a}\left\langle {n \atop a} \right\rangle (1/4)^{(n-a)} \left\langle {n-a \atop b} \right\rangle$$

where n is the total number of F_2 mutant genotypes, a is the number that are heterozygous and b is the number of homozygous Landsberg genotypes.

The segregation by chance analysis for trait localization was chosen because in most model organism systems, F_1 parents are essentially genetically identical. Traditional linkage analysis usually used for linking human traits to markers is based on the number of meiotic crossovers between the disease gene and the closest marker in large pedigrees. Although traditional linkage analysis could be used here, it would complicate the mapping by inclusion of the F2 progeny known to be wild type (and not scored on the chip).

Construction of the SNP map. We calculated the maps using Mapmaker Version 3 software[20]. The 237 SNP markers were placed onto the canonical Lister and Dean RI map. The map was calculated using a framework of markers for each of the linkage groups (chromosomes). We calculated two-point data for the entire collection of markers and used the results of this calculation to assign markers to chromosomes based on their linkage to the framework markers. Once the markers were linked to a specific chromosome, markers were assigned their most probable position based on three-point data. Maps were calculated using the Kosambi mapping function.

Mutagenesis and generation of *Eds* plants. The *Eds16-1* mutant (L. Reuber, J.D. and F.M.A., unpublished data) was isolated in a screen for *A. thaliana* mutants with enhanced susceptibility to the fungal pathogen *E. orontii*. Briefly, the screen was carried out by inoculating M2 plants from either fast neutron or EMS mutagenized seed pools (Lehle Seeds) with *E. orontii* conidia using a settling tower procedure[25]. Putative *eds* mutants that were more heavily infected than wild-type controls were rescored in the M3 generation. *Eds16-1* was among several mutants identified that exhibited enhanced susceptibility. Genetic analysis showed that *Eds16-1* segregated as a single recessive mendelian trait, was not allelic to described *Eds* mutants[26] and mapped to the bottom of chromosome 1 (L. Reuber, J.D. and F.M.A., unpublished data).

Acknowledgements

Helpful comments were provided by E. Winzeler, D. Preuss, M. Thomas, E. Huala, M. Cherry and D. Kim. This work was partially funded by the NIH grant 3R37 HG00198-23 and the NSF grant MCB-9729599.

Received 12 April; accepted 20 August 1999.

© 1999 Nature America Inc. • http://genetics.nature.com

1. Kruglyak, L. The use of a genetic map of biallelic markers in linkage studies. *Nature Genet.* **17**, 21–24 (1997).
2. Kwok, P.Y., Deng, Q., Zakeri, H., Taylor, S.L. & Nickerson, D.A. Increasing the information content of STS-based genome maps: identifying polymorphisms in mapped STSs. *Genomics* **31**, 123–126 (1996).
3. Wang, D.G. *et al.* Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* **280**, 1077–1082 (1998).
4. Winzeler, E.A. *et al.* Direct allelic variation scanning of the yeast genome. *Science* **281**, 1194–1197 (1998).
5. Hacia, J.G., Brody, L.C., Chee, M.S., Fodor, S.P. & Collins, F.S. Detection of heterozygous mutations in BRCA1 using high density oligonucleotide arrays and two-colour fluorescence analysis. *Nature Genet.* **14**, 441–447 (1996).
6. Buetow, K.H., Edmonson, M.N. & Cassidy, A.B. Reliable identification of large numbers of candidate SNPs from public EST data. *Nature Genet.* **21**, 323–325 (1999).
7. Winzeler, E.A. & Davis, R.W. Functional analysis of the yeast genome. *Curr. Opin. Genet. Dev.* **7**, 771–776 (1997).
8. Sawa, S., Ito, T., Shimura, Y. & Okada, K. FILAMENTOUS FLOWER controls the formation and development of Arabidopsis inflorescences and floral meristems. *Plant Cell* **11**, 69–86 (1999).
9. Preuss, D., Lemieux, B., Yen, G. & Davis, R.W. A conditional sterile mutation eliminates surface components from Arabidopsis pollen and disrupts cell signaling during fertilization. *Genes Dev.* **7**, 974–985 (1993).
10. Meyerowitz, E.M. Pattern formation in plant development: four vignettes. *Curr. Opin. Genet. Dev.* **4**, 602–608 (1994).
11. Lloyd, A.M., Schena, M., Walbot, V. & Davis, R.W. Epidermal cell fate determination in Arabidopsis: patterns defined by a steroid-inducible regulator. *Science* **266**, 436–439 (1994).
12. Mindrinos, M., Katagiri, F., Yu, G.L. & Ausubel, F.M. The A. thaliana disease resistance gene RPS2 encodes a protein containing a nucleotide-binding site and leucine-rich repeats. *Cell* **78**, 1089–1099 (1994).
13. Bent, A.F. *et al.* RPS2 of Arabidopsis thaliana: a leucine-rich repeat class of plant disease resistance genes. *Science* **265**, 1856–1860 (1994).
14. Alonso-Blanco, C. *et al.* Development of an AFLP based linkage map of Ler, Col and Cvi Arabidopsis thaliana ecotypes and construction of a Ler/Cvi recombinant inbred line population. *Plant J.* **14**, 259–271 (1998).
15. Meinke, D.W., Cherry, J.M., Dean, C., Rounsley, S.D. & Koornneef, M. Arabidopsis thaliana: a model plant for genome analysis. *Science* **282**, 662, 679–682 (1998).
16. Hoogendoorn, B. *et al.* Genotyping single nucleotide polymorphisms by primer extension and high performance liquid chromatography. *Hum. Genet.* **104**, 89–93 (1999).
17. O'Donovan, M.C. *et al.* Blind analysis of denaturing high-performance liquid chromatography as a tool for mutation detection. *Genomics* **52**, 44–49 (1998).
18. Cronin, M.T. *et al.* Cystic fibrosis mutation detection by hybridization to light-generated DNA probe arrays. *Hum. Mutat.* **7**, 244–255 (1996).
19. Liu, Y.G., Mitsukawa, N., Lister, C., Dean, C. & Whittier, R.F. Isolation and mapping of a new set of 129 RFLP markers in Arabidopsis thaliana using recombinant inbred lines. *Plant J.* **10**, 733–736 (1996).
20. Lander, E.S. *et al.* MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* **1**, 174–181 (1987).
21. Konieczny, A. & Ausubel, F.M. A procedure for mapping Arabidopsis mutations using co-dominant ecotype-specific PCR-based markers. *Plant J.* **4**, 403–410 (1993).
22. Li, J. & Chory, J. Preparation of DNA from Arabidopsis. *Methods Mol. Biol.* **82**, 55–60 (1998).
23. Lockhart, D.J. *et al.* Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nature Biotechnol.* **14**, 1675–1680 (1996).
24. Chee, M. *et al.* Accessing genetic information with high-density DNA arrays. *Science* **274**, 610–614 (1996).
25. Adam, L. & Somerville, S.C. Genetic characterization of five powdery mildew disease resistance loci in Arabidopsis thaliana. *Plant J.* **9**, 341–356 (1996).
26. Volko, S.M., Boller, T. & Ausubel, F.M. Isolation of new Arabidopsis mutants with enhanced disease susceptibility to Pseudomonas syringae by direct screening. *Genetics* **149**, 537–548 (1998).