Genet. Res., Camb. (2006), **87**, *pp.* 75–85. © 2006 Cambridge University Press
doi:10.1017/S0016672306008020 Printed in the United Kingdom

Comparative gene mapping in *Arabidopsis lyrata* chromosomes 1 and 2 and the corresponding *A. thaliana* chromosome 1: recombination rates, rearrangements and centromere location

BENGT HANSSON[1]†, AKIRA KAWABE[1], SONJA PREUSS[1], HELMI KUITTINEN[2] AND DEBORAH CHARLESWORTH[1]*
[1]*Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, West Mains Road, Edinburgh EH9 3JT, UK*
[2] *Department of Biology, PL 3000, FIN-90014 University of Oulu, Finland*

(*Received 7 November 2005 and in revised form 23 December 2006*)

Summary

To add detail to the genetic map of *Arabidopsis lyrata*, and compare it with that of *A. thaliana*, we have developed many additional markers in the *A. lyrata* linkage groups, LG1 and LG2, corresponding to *A. thaliana* chromosome 1. We used a newly developed method for marker development for single nucleotide polymorphisms present in gene sequences, plus length differences, to map genes in an *A. lyrata* family, including variants in several genes close to the *A. thaliana* centromere 1, providing the first data on the location of an *A. lyrata* centromere; we discuss the implications for the evolution of chromosome 1 of *A. thaliana*. With our larger marker density, large rearrangements between the two *Arabidopsis* species are excluded, except for a large inversion on LG2. This was previously known in *Capsella*; its presence in *A. lyrata* suggests that, like most other rearrangements, it probably arose in the *A. thaliana* lineage. Knowing that marker orders are similar, we can now compare homologous, non-rearranged map distances to test the prediction of more frequent crossing-over in the more inbreeding species. Our results support the previous conclusion of similar distances in the two species for *A. lyrata* LG1 markers. For LG2 markers, the distances were consistently, but non-significantly, larger in *A. lyrata*. Given the two species' large DNA content difference, the similarity of map lengths, particularly for LG1, suggests that crossing-over is more frequent across comparable physical distances in the inbreeder, *A. thaliana*, as predicted.

1. Introduction

Genetic mapping is of evident importance in any species of ecological or genetic interest, and molecular markers are providing the ability to map ever more species. The availability of the complete genome sequence of *Arabidopsis thaliana* now makes it possible to begin studies of genome evolution in the genus *Arabidopsis*. It is well known that *A. thaliana* is a plant with a small DNA content (Arumuganathan & Earle, 1991), with a haploid genome of about 0·16 picograms (Bennett *et al.*, 2003), or about 157 Mbp, and it is likely that this is due to loss of DNA since the split from related species, since the close relative, *A. lyrata*, has a DNA content about 50% higher (Johnston *et al.*, 2005). Most likely, the weedy life-history of *A. thaliana* has selected for this reduction, but other factors may have contributed. For instance, transposable elements (TEs) may have come to lower equilibrium abundances, in response to the species' high selfing rate, as predicted from theoretical models of TE abundance (Charlesworth & Charlesworth, 1995; Wright *et al.*, 2001; Wright & Schoen, 1999; Devos *et al.*, 2002). In addition, loss of genomic DNA may have occurred in chromosome rearrangements.

It is clearly of great interest to know how these changes in DNA content evolved, the time periods involved, and how they affect recombination rates. Recombination rate differences may evolve quite quickly. Recombination hotspots differ between humans and chimpanzees (Winckler *et al.*, 2005), and genetic map distances differ between *Drosophila*

* Corresponding author. Tel: +44 131 6505751. Fax: +44 131 6506564. e-mail: deborah.charlesworth@ed.ac.uk
† Present address: Department of Animal Ecology, Lund University, Ecology Building, 22362 Lund, Sweden.

species (e.g. True *et al*., 1996). Even strains of the same species may differ. Within *Arabidopsis thaliana*, crossover frequencies differ between strains (Sanchez-Moran *et al*., 2002). Some general patterns have also been observed cytologically in recombination rates, particularly the greater chiasma frequency of inbreeding species compared with outcrossing relatives (Arroyo, 1975; Garber, 1956; Zarchi *et al*., 1972; Grant, 1958; Rees, 1961). There are some reasons for expecting this difference, since some models that select for recombination lead to higher recombination when the population is inbreeding (Charlesworth *et al*., 1977, 1979; Hedrick *et al*., 1978). Alternatively, large diversity (i.e. greater sequence differences between alleles) in outcrossers has been suggested to inhibit recombination (Borts & Haber, 1987), though the divergence will generally not be enough for this to be likely, and this interpretation is also contradicted by evidence that duplicate genes (paralogues) within yeast genomes can continue to undergo gene conversion until their sequences have diverged considerably (Gao & Innan, 2004).

To our knowledge, no previous studies have tested whether the expected chiasma frequency differences between inbreeders and outcrossers are reflected in recombination distances estimated genetically. The genus *Arabidopsis* offers an excellent opportunity to begin to test this. Its chromosome evolution is also of interest, since *A. thaliana*, with $2n=2x=10$ chromosomes, has several chromosome fusions relative to other related species, with $2n=2x=16$ (Lysak *et al*., 2003). Closely related species are known, including the outcrossing species *A. lyrata*, which is estimated, using a molecular clock, to have split from *A. thaliana* roughly 5 MYA ago. The self-compatible species, *Capsella rubella*, is more distantly related, but still split recently enough to be a suitable outgroup for inferring the directions of evolutionary changes (Koch *et al*., 2000).

Fusions are known to be common in inbreeders (Lande, 1984; Charlesworth, 1992), and two sparse maps of *A. lyrata* (Kuittinen *et al*., 2004; Yogeeswaran *et al*., 2005; Koch & Kiefer, 2005) deduced that several fusions have occurred in *A. thaliana*, accounting for this species' five chromosomes, versus the eight chromosomes of *A. lyrata* and other closely related species. One fusion joins the lower arm of two ancestral acrocentric chromosomes (corresponding to the *A. lyrata* chromosomes 1 and 2), forming *A. thaliana* chromosome 1, and two others formed the long arms of the *A. thaliana* chromosomes 2 and 5. Reciprocal translocations also occurred between chromosome arms. Large tracts of genome should nevertheless be syntenic, unless rearrangements within the major chromosome arm regions have followed these events. Most chromosome arms indeed have the same gene content, in the same order, in the mapped *Arabidopsis* and related species, but marker densities are low and rearrangements in some regions may be missed (Kuittinen *et al*., 2004; Yogeeswaran *et al*., 2005; Koch & Kiefer, 2005; Boivin *et al*., 2004).

To understand genome evolution and its consequences for sequence diversity, data are needed about both genetic and physical maps. These are available for *A. thaliana*, from the genome sequence together with mapping studies (Lister & Dean, 1993; Copenhaver *et al*., 1997; Alonso-Blanco *et al*., 1998), but detailed comparisons with *A. lyrata* will require mapping of densely spaced markers, and physical mapping information (which should become available from the forthcoming genome sequencing of this species). Here, we describe mapping of 35 additional markers in the *A. lyrata* chromosomes corresponding to linkage groups 1 and 2 (LG1 and LG2 of Kuittinen *et al*., 2004), which are jointly homologous to *A. thaliana* chromosome 1 (Kuittinen *et al*., 2004; Yogeeswaran *et al*., 2005; Koch & Kiefer, 2005; Boivin *et al*., 2004). Most of our markers were developed using a new approach named 'ASP' (see description below), which uses allele-specific primers based on coding sequences of identified genes in *A. thaliana* (Hansson & Kawabe, 2005). Using genic markers allows comparisons between the genetic maps, which is necessary for studying chromosome evolution.

In combination with the previously mapped markers in Kuittinen *et al*. (2004), we now have 55 markers on *A. lyrata* LG1 and LG2 markers. The orthologous *A. thaliana* genes are distributed across the whole of chromosome 1. Our mapping confirms the general conclusions of the previous maps that marker orders are similar in the two *Arabidopsis* species, with the exception of a few genes with two copies in *A. lyrata* but only one in *A. thaliana*, and a large inversion on LG2. Our greater marker density excludes all except small rearrangements, so that we can now compare map distances without the danger that large rearrangements might be present in the regions compared. Our study supports in more detail the previous conclusion (Kuittinen *et al*., 2004) that distances of the markers on *A. lyrata* LG1, and perhaps also LG2, are similar between the two species. Finally, the similarity of gene orders in the two species suggests that it may be possible to identify the centromere locations in *A. lyrata*, and we also describe evidence for putative centromere locations in one of the two linkage groups studied, and discuss the implications for the evolution of chromosome 1 of *A. thaliana*.

2. Materials and methods

(i) *Materials*

DNA samples were used from the *A. lyrata* subsp. *petraea* mapping family of Kuittinen *et al*. (2004). In

https://doi.org/10.1017/S0016672306008020 Published online by Cambridge University Press

what follows, we refer to the species as *A. lyrata*. The parents of the F2 family used for mapping were plants from two natural populations: Karhumäki (Russia) and Mjällom (Sweden) (for more information on the populations see van Treuren *et al*., 1997). One of the Fl parents, KM6-5 was available, together with 99 F2 progeny. Samples of the inbreeding species, *Arabis glabra*, were collected in Shirahama, Japan by A.K.

(ii) *Markers*

We used a set of PCR primers designed for a diversity survey of loci on chromosome 1 of *A. thaliana* (Wright, 2003). We mapped 35 loci in addition to the 20 mapped by Kuittinen *et al*. (2004). The loci were first selected out of all annotated chromosome 1 genes on the basis that they are apparently single copy (based on giving single hits in BLASTn), and had an exon of at least 800 bp (see Wright, 2003).

Additional primers were designed for loci in the centromere regions of the *A. thaliana* chromosome 1, in order to attempt to identify the centromere regions of the *A. lyrata* linkage groups 1 and/or 2. The CEN1 region of *A. thaliana* is located at about 15 Mb (Hosouchi *et al*., 2002). The closest non-transposon genes to CEN1 are located at about 14·65 Mb and 15·44 Mb in the short and long arms, respectively. Because there are several sequences highly similar to these genes in the *A. thaliana* genome, we chose slightly more distant single-copy genes for our analyses (At1g36160, 36310, 36370, 36730, 37130 in the short arm, and At1g42470, 42990 and 43980 in the long arm). In the *A. thaliana* map these genes are located between 0·5 and 1·2 Mb from the core centromere region; except for At1g37130, they are outside the reported non-recombining regions (Copenhaver *et al*., 1999), in regions that are still transposon-rich and have a low recombination rate in *A. thaliana*. All these genes map to *A. lyrata* LG1 (see Supplementary Table 1).

The 55 markers are distributed over the whole of *A. thaliana* chromosome 1, and the proximate markers (At1g01030 and 78850) were 0·1 and 0·8 Mbp, respectively, from the tips of chromosome 1. Thus, 97% of the physical map of *A. thaliana* chromosome 1 (30·4 Mbp) is now covered.

To obtain markers for mapping, a set of four to eight plants from the mapping family (one parent plus between three and seven F2 progeny) were sequenced for each gene chosen, and the sequences examined for polymorphisms suitable for scoring in the F2. If variants were present, we designed genotyping approaches (see below). In total, 13 single-nucleotide polymorphisms (SNPs) were genotyped by PCR-RFLP, and 16 SNPs with allele-specific PCR with primer-induced fragment-length variation (Hansson & Kawabe, 2005). Four loci had length variants (indels) and these were screened by standard fragment-length analysis. Nine loci segregate in one Fl parent only (in the ab × aa or aa × ab configuration), and the remaining 26 loci segregate in the ab × ab configuration.

(iii) *PCR-RFLP*

For the loci typed by PCR-RFLP (see Supplementary Table 1, http://www.journals.cambridge.org/jid_grh), PCR was initially performed using the primers used for the search for polymorphic sites. Reactions were done in 25 μl reaction volumes, including about 2–10 ng of total DNA, 1·25 pmol of primer, 0·25 mM dNTPs, and 1 × *Taq* polymerase buffer, at an annealing temperature of 55 °C for 30 cycles. If PCR failed with these conditions, the reaction was repeated with higher template DNA concentrations and/or 50 °C annealing temperature. Fifteen microlitres of PCR products was then digested in 20 μl volume reaction containing 2–4 units of restriction enzymes chosen on the basis of the sequences. After incubation at the recommended temperature for between 2 h to overnight, 5 μl aliquots of the digested products were separated in agarose gels (1% or 2% depending on the DNA fragment sizes). If PCR amplification was weak, the remaining 15 μl of the digested PCR products was re-analysed in the same manner.

For most loci, typing was straightforward, but a few loci presented difficulties, or produced unexpected results. New primers, in addition to the initial ones described above, were needed for two loci (see Supplementary Table 1). Unexpected segregation patterns for locus At1g37130 were caused by a polymorphism at the primer annealing site, causing weak amplification of some alleles. This was confirmed by sequencing, and new primers were designed in regions monomorphic in the three sequence types detected, and the F2 plants were genotyped by PCR-RFLP.

(iv) *PCR-based screening of SNPs and indels*

SNPs which did not coincide with affordable restriction endonuclease recognition sites were genotyped using allele-specific primers (Hansson & Kawabe, 2005). In this method, two different allele-specific forward primers inducing a length difference at the 5′ end of the product are designed for a given SNP; the different alleles segregating in a family can then be scored with standard fragment-length analysis by capillary electrophoresis.

The PCR mix contained 2 pmol of each forward primer, 4 pmol reverse primer, 1 × NH_4-buffer (Bioline), 15 nmol $MgCl_2$, 2 nmol dNTP, 0·5 U BIOTAQ DNA polymerase (Bioline) and 1–5 ng template in 10 μl reaction volumes. PCR conditions were as follows: 94 °C for 2 min, then 35 cycles at 94 °C for 30 s/T_A for 30 s/72 °C for 30 s, followed by 72 °C for 10 min;

https://doi.org/10.1017/S0016672306008020 Published online by Cambridge University Press

where T_A is the locus-specific annealing temperature (Supplementary Table 1). The fluorescently labelled PCR products were separated and alleles detected in an ABI 3730 capillary sequencer (Applied Biosystems). We simultaneously multiplexed two to four loci in the sequencer. A few individuals with known genotypes (determined from sequence data) were run with the genotyping protocol to confirm that variants were correctly detected. In total, we designed primers for 19 loci, and of these co-dominant genotyping was possible for 15 loci, while one locus behaved as dominant (At1g32860) but could still be mapped. Three loci gave a single strong peak, or the same peaks, irrespective of the underlying SNP genotype and thus could not be scored with this method; one of these loci was genotyped by PCR-RFLP, but the others had no restriction enzyme recognition sites (Supplementary Table 1).

Two loci, At1g42470 and a related sequence we named 'At1g43980-deletion' (see below), differed by large indels (at least 50 bp) and their alleles were scored after separation on agarose gels. Three loci with short indels (Supplementary Table 1) were each scored using a single forward primer and an end-labelled reverse primer. The PCR mix contained 4 pmol of each primer, but the PCR and screening protocols were otherwise as described for the allele-specific PCR.

(v) *Analyses*

Mapping was done using JoinMap software 3.0 (van Ooijen & Voorrips, 2001). The method is based on the least squares method of Stam (1993). All data were included in the analysis, including eight markers with significant segregation distortion. The F2 genotypes for all markers were analysed, including the previous data for markers from all chromosomes (Kuittinen *et al*., 2004), in order to define the linkage groups. With a LOD score threshold of 5·0, the new markers were, as expected, associated with linkage groups one and two (LG1 and LG2). New maps for these linkage groups were constructed using the Kosambi mapping function for pairwise recombination frequencies and default parameter values for the mapping algorithm (linkage data for pairs of loci were used with the thresholds LOD 1·0 and recombination frequency less than 0·40, and the chi-square jump threshold for adding a new marker was 5·0). Recombination rates were not estimated separately for male and female meioses, because only 99 progeny were available in the family. This should be done in the future, using larger mapping families and fully informative markers.

To compare the maps of *A. thaliana* chromosome 1 with those of the homologous *A. lyrata* LG1 and LG2, we used the same loci as far as possible. A few of the loci mapped by us are also genetically mapped in *A. thaliana*, but, failing that, we often used data on mapped loci in *A. thaliana* that are physically closest to homologues of our genetically mapped loci. *A. thaliana* RI (recombinant inbred) mapping data and physical positions of markers were obtained from The Arabidopsis Information Resource (TAIR) database (http://www.arabidopsis.org, 1 April 2005).

3. Results

(i) *Gene duplications*

All the genes mapped in our study were studied using PCR primers derived from *A. thaliana* sequences (see Section 2). Thus we will be unable to map any loci that have been lost in *A. thaliana*; as discussed below, such gene losses may occur in the process of the chromosome fusions in which three centromeres were lost, and could account for absence of many genes from this plant's genome that are present in other dicotyledon species: two in the rosid division, like *A. thaliana*, and one asterid species (Allen, 2002). Although that analysis suggests loss of many genes in *A. thaliana*, this could be due to causes other than rearrangements, including the presence of TE sequences in plant genome databases (Bennetzen *et al*., 2004). A reliable estimate of the number lost in the evolution of its genome requires testing and mapping genes derived from close relatives, which is not yet feasible. We can, however, start to test this by asking whether any loci are present as single copies in *A. thaliana* but in multiple copies in *A. lyrata*, and, if so, whether these loci are close to the centromeres involved in fusions, or to other rearrangements. We indeed found evidence of two duplications in *A. lyrata* with respect to the state in *A. thaliana*, but neither appears to be due to loss of genes during chromosome fusions. We mapped the loci involved (At1g43980 and At1g01370, or HTR-12, described elsewhere; see A. Kawabe, S. Nasuda and D. Charlesworth, unpublished results). The HTR-12 genes are not located near regions involved in fusions or rearrangements. We next briefly describe the results from At1g43980.

PCR for the At1g43980 gene yielded three bands: one present in all plants tested (one parent and seven F2 progeny) and two other bands showing presence/absence variation. These three bands were gel-purified and cloned from plants with the three different band patterns. Four different sequences were present, indicating duplication. One band includes two sequences, both the same length as in *A. thaliana*. The two other bands differ in length and sequence from this 'thaliana type', sharing deletions of 1 bp and 23 bp (compared with either the *A. thaliana* sequence or the first two sequences), but are similar in sequence to each other. We refer to them as 'At1g43980-deletion' sequences. Intriguingly, At1g43980 is one

https://doi.org/10.1017/S0016672306008020 Published online by Cambridge University Press

of the loci with locations close to the *A. thaliana* centromere. However, it does not seem to represent a deletion in the fusion that formed the chromosome 1 centromere, because, in an outgroup species, *Arabis glabra*, only one copy is amplified, and individuals of this inbreeding species have only a single sequence (A. Kawabe, unpublished results). Given that this gene appears to be single-copy in both *A. glabra* and *A. thaliana*, it seems most likely that *A. lyrata* has experienced a duplication.

The shorter of these two sequences has deletions of 18 bp, 1 bp and 31 bp, which cause frame-shifts inducing stop codons, suggesting that the deletion locus may have become a pseudogene after the duplication. Fourteen fixed substitutions were found between the 'thaliana type' and 'At1g43980-deletion' sequences, all of them 5′ to the 23 bp deletion. 3′ of this deletion, two variants are shared between the two types of sequences, suggesting that gene conversion has occurred between the duplicates. The At1g43980-deletion locus was typed using the length variants just described, and the other locus was typed by PCR-RFLP using a new locus-specific primer (Supplementary Table 1). Although we observed no amplification of very long PCR products containing the complete duplicated pair of sequences, the duplicated loci showed complete linkage in the F2 plants, and may represent a tandem duplication.

(ii) *Marker development and mapping results*

The study described here is the first test of the ASP marker development approach of Hansson & Kawabe (2005). Our results here show that it readily generates useful markers for mapping. Primers were tested for 42 loci (Supplementary Table 1). Including the two duplicated loci just mentioned, 44 loci were tested for polymorphism, and suitable variants were found in 36 of them (82%; Supplementary Table 1). As described above, the duplicated At1g43980 loci showed complete linkage in the F2 plants, and may represent a tandem duplication. These loci are therefore treated as a single locus in what follows, and depicted as one locus in Fig. 1, thus giving 35 new loci on *A. lyrata* LG1 and LG2.

A total of 55 genes from chromosome 1 of *A. thaliana* are now mapped in *A. lyrata* (Fig. 1): 25 new ones on LG1 and 10 on LG2 (in addition to the 15 and 5 loci already mapped on these two chromosomes by Kuittinen *et al.*, 2004). Figs. 2 and 3 show the results, together with data from chromosome 1 of *A. thaliana*. The linkage map for *A. lyrata* now spans 83 cM for chromosome 1 (on average, one locus per 2·1 cM) and 73 cM for chromosome 2 (on average, one locus per 4·9 cM), representing increases of 13% and 25%, respectively, compared with the previous map (Kuittinen *et al.*, 2004); this total of 156 cM compares with 135 cM for *A. thaliana* chromosome 1 (TAIR database, http://www.arabidopsis.org/).

Eight centromere-located loci (grey box in Fig. 1) were mapped to *A. lyrata* LG1, and they proved to be at the location corresponding to the centromere in the *A. thaliana* map. All these loci were completely linked (Figs. 1, 2), except for At1g43980, but this was also very closely linked to the other seven.

Our mapping effort shows that inversion 3, found in the *C. rubella* map of the chromosome corresponding to chromosome LG2 (Koch & Kiefer, 2005), is also present in *A. lyrata*, so that it can be inferred that the rearrangement occurred in the *A. thaliana* lineage. Apart from this, most markers are in the same order as in the *A. thaliana* map. The 12 exceptions are single out-of-place markers, and do not appear to indicate inversions (Fig. 1).

Figs. 2 and 3 show that the genetic map distances for LG1 of *A. lyrata* are remarkably similar to the *A. thaliana* ones, despite the probable larger DNA amount in *A. lyrata* (assuming that these chromosomes differ from the *A. thaliana* DNA content similarly to the genome-wide average: see Johnston *et al.*, 2005; Sanyal & Jackson, 2005). LG2 distances are, however, consistently larger than those in the *A. thaliana* map, in agreement with previous findings (Kuittinen *et al.*, 2004; Yogeeswaran *et al.*, 2005), though the slope of the linear regression for LG2 (Fig. 3*b*) is not significantly different from unity, by a *t*-test, nor are the differences in the map intervals significantly different by a non-parametric sign test.

4. Discussion

(i) *Chromosome evolution of* A. thaliana *chromosome: synteny, inversion and centromeres*

The new map data show that the break-point of the LG1 end of the chromosome fusion that created the *A. thaliana* chromosome 1 lies between the most distal gene mapped in *C. rubella* linkage group A, At1g56000 and At1g57820, which is proximal to other linkage group B markers (Boivin *et al.*, 2004). In *A. thaliana*, inversion 3 of Koch & Kiefer (2005) (which our denser map now clearly shows occurred in the *A. thaliana* lineage) has brought these genes physically close to one another. The other end of inversion 3 must lie between At1g64610 and At1g65450, which are within a region of less than 500 kb in the *A. thaliana* genome. In Fig. 1 we have drawn a single inversion, but the map resolution cannot exclude a more complex event in which several genes were rearranged.

It is very difficult to identify the locations of centromeres of a species' chromosomes (e.g. Khrustaleva *et al.*, 2005). Having a genome sequence of a related species might be helpful, if the chromosomes have

https://doi.org/10.1017/S0016672306008020 Published online by Cambridge University Press

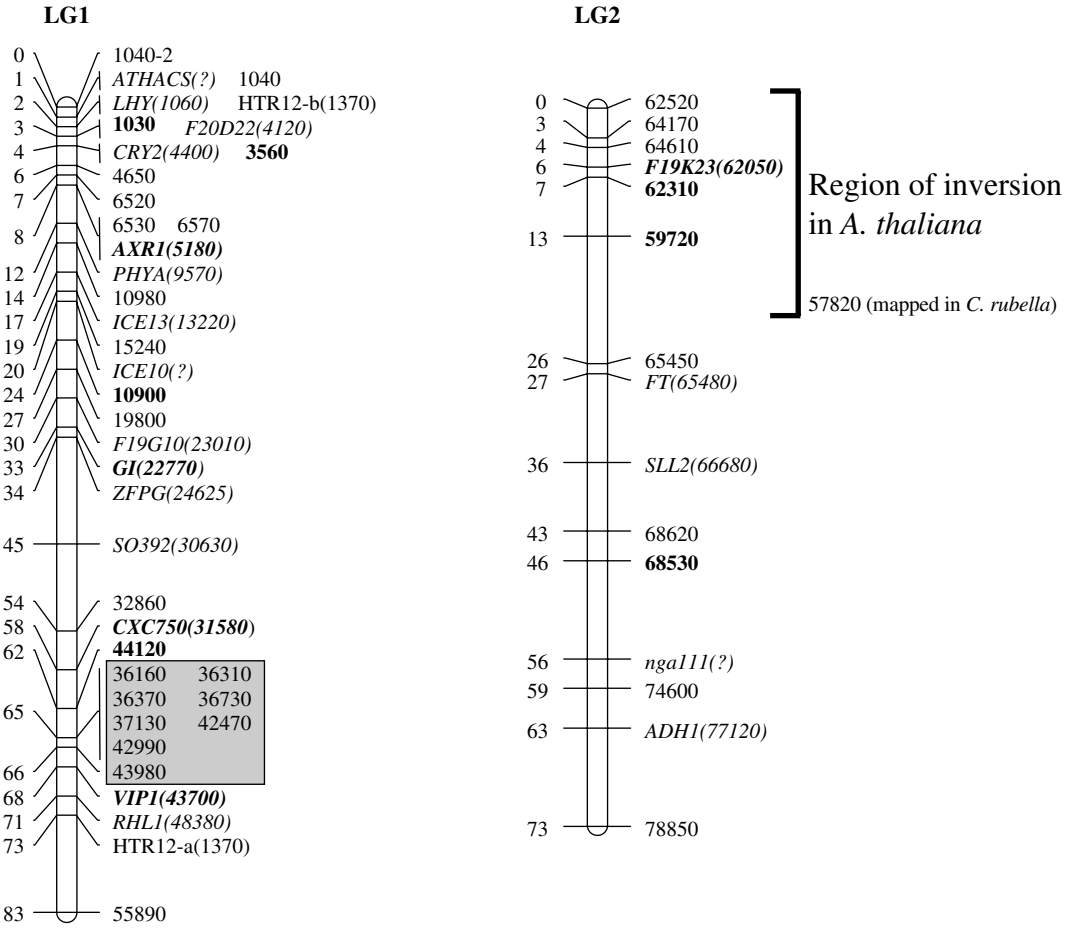

Fig. 1. Linkage map for *A. lyrata* chromosomes 1 and 2. Loci are identified by their MIPS code numbers (e.g. 10 980 = At1g10980), except for a few loci, mostly those mapped by Kuitinnen *et al.* (2004), where locus names were given for loci in the previously published maps; for these, we show the name and give the MIPS code in parentheses. For three loci the previous names are for putative genes, are indicated as '(?)', and are without MIPS numbers. Loci typed by Kuittinen *et al.* (2004) are in italics, and rearranged genes are indicated in bold type. Loci in the centromere region of *A. thaliana* chromosome 1 are boxed.

not been rearranged in the regions surrounding the centromeres. In *A. thaliana*, a region of suppressed recombination surrounds the core centromere regions (Copenhaver *et al.*, 1999), so that it may be possible to identify a centromere by testing whether loci in the region of suppressed recombination in this species are also completely linked in *A. lyrata*. We tested whether this approach can identify the centromere region of at least one of the *A. lyrata* linkage groups 1 and 2. The results also help test whether the fusion between the *A. lyrata* chromosomes 1 and 2, which created the existing *A. thaliana* chromosome 1, was a simple fusion event between intact chromosome arms or involved rearrangements. If the event was a simple one, we expect the markers to be in the same order in both species, though some markers may have been lost (Fig. 4*A*).

The eight centromere-located loci (grey box in Fig. 1) proved to be at the location corresponding to the centromere in the *A. thaliana* map, suggesting that the centromere of the *A. thaliana* chromosome 1 has retained the same location during the rearrangement that created this chromosome. This centromere does not appear to have moved, either by pericentric inversions or by neo-centromere formation, as has occurred in mammals and other organisms (Ventura *et al.*, 2001), though we cannot exclude small rearrangements, and the gene order in this region includes three out-of-place markers compared with *A. thaliana* (Fig. 1). The resolution of genetic mapping is, however, too coarse to definitively determine their order, which will probably require detailed cytogenetic studies involving *in situ* hybridization.

Our results also identify the lost centromere as that of the *A. lyrata* chromosome 2. There are two possible models for loss of the LG2 centromere region, which must presumably have occurred in or soon after the fusion event. First, a simple reciprocal translocation could have occurred between two chromosomes with short arms (Fig. 4, model A). This would cause loss of neighbouring loci. The fact that both the fusion and inversion 3 (see above) occurred in the *A. thaliana*

https://doi.org/10.1017/S0016672306008020 Published online by Cambridge University Press

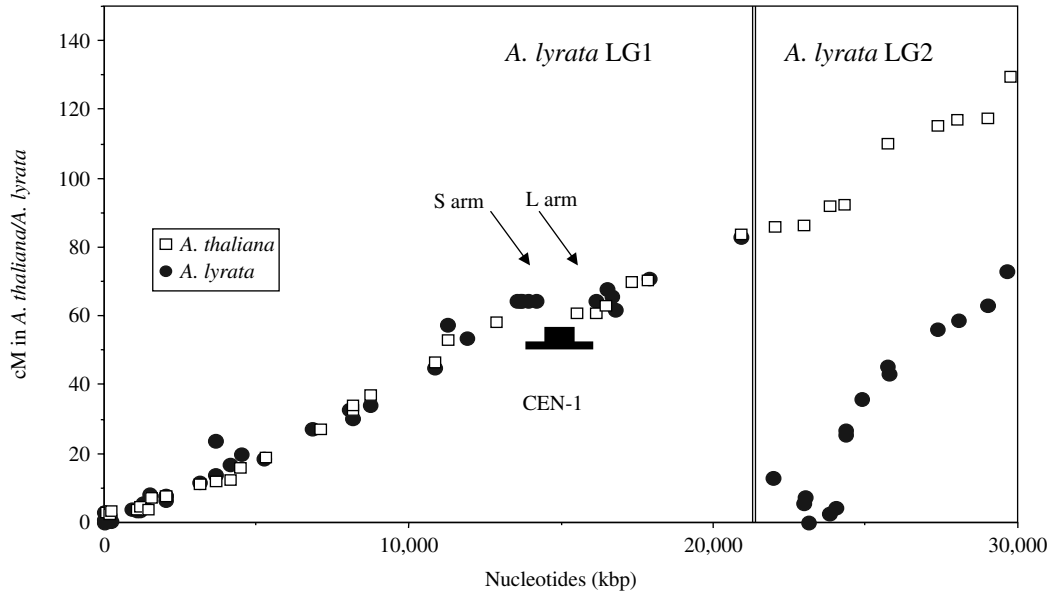

Fig. 2. Comparison between the genetic and physical maps of chromosome 1 of *A. thaliana* and the genetic maps *A. lyrata* of LG1 and LG2. The *x*-axis gives the physical positions of the *A. thaliana* markers. The total map length in *A. thaliana* is about 30 Mb and 135 cM (based on information in the TAIR database, http://www.arabidopsis.org/).

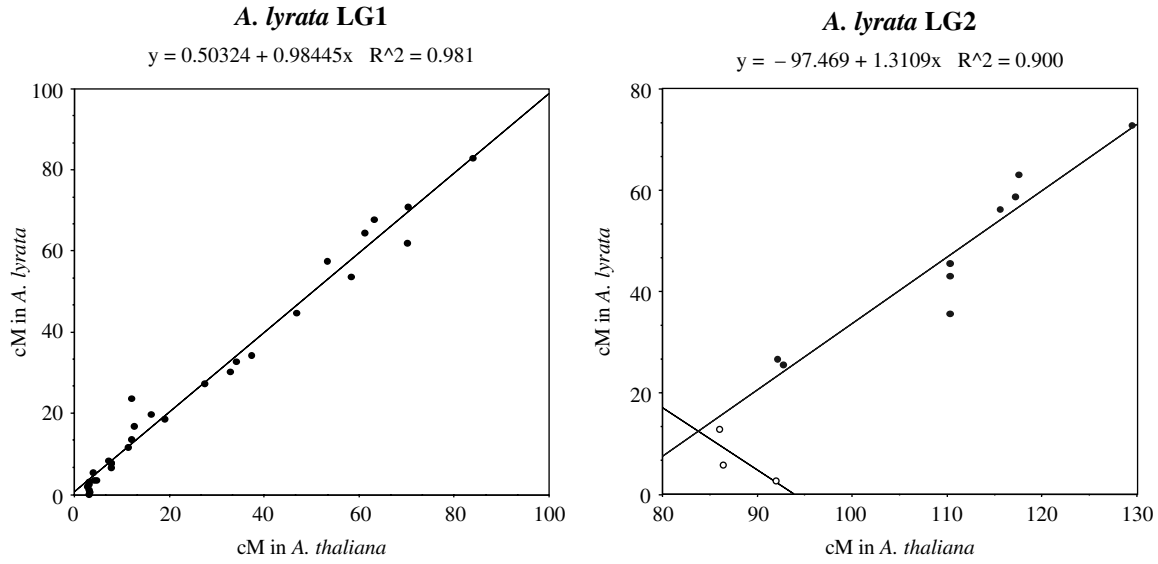

Fig. 3. Genetic map locations of chromosome 1 of *A. thaliana* and *A. lyrata* LG1 and LG2 markers, showing the similarity in map distances between the same markers for LG1, and differences for LG2 markers. The LG2 markers that are in inverted orientations in the two species are shown as open circles. We cannot compare all intervals between the two species, because no markers close to some of the genes we mapped have been mapped in *A. thaliana*. We excluded three loci in the region of LG2 that was involved in inversion 3 (At1g62310, At1g62520 and At1g64610), and seven other markers (At1g03560, At1g36160, At1g36310, At1g36370, At1g36730, At1g37130 and At1g43980) that are more than 500 kb distant from the closest mapped *A. thaliana* gene.

lineage suggests the alternative possibility that an inversion in the ancestral acrocentric chromosome could first have brought the LG2 centromere close to the tip of the chromosome, allowing fusion with chromosome 1 without loss of many genes (Fig. 4, model B).

The centromere of *A. lyrata* LG2 presumably lies beyond the inversion, closer to At1g65450 than to At1g59720. It may be possible to identify candidate centromere-linked markers for *A. lyrata* LG2 by adding more markers in this region, but we have not found suitable *A. thaliana* loci to develop markers, so molecular cytogenetic studies may be more promising to test whether loci in this region are missing from *A. thaliana* (which would be detectable as a gap in tiled FISH experiments with *A. thaliana* probes).

The density of LG2 markers remains low, even with our additional loci, and one large gap is indeed in the

https://doi.org/10.1017/S0016672306008020 Published online by Cambridge University Press

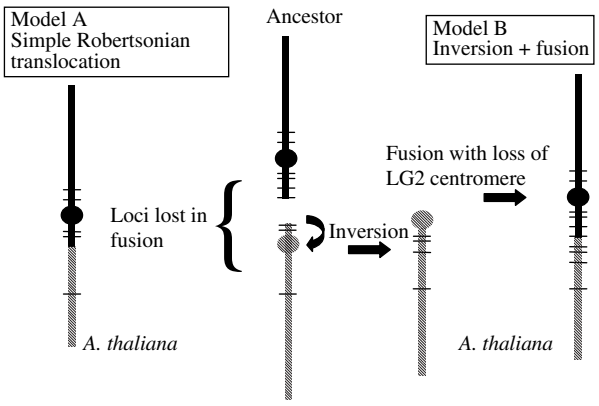

Fig. 4. Schemes for the creation of chromosome 1 of *A. thaliana* by the fusion of two ancestral acrocentric chromosomes corresponding to the current LG1 (black) and LG2 (hatched) of *A. lyrata*. Centromeres are shown as circles and markers are shown as short horizontal lines. Model A involves loss of genes and markers distal to the centromere of at least one chromosome (markers surrounding the centromere of the LG2 chromosome would be lost, plus potentially some genes from the tip of LG1). Model B involves an inversion, bringing the LG2 centromere very close to the end of the chromosome, followed by the fusion, during which the centromere was lost, together with smaller genome regions than in *A. thaliana*.

region between At1g65450 and At1g59720 (Fig. 1). Our efforts failed to eliminate this gap in this region, although it is reduced from 17 cM in the previous *A. lyrata* genetic map (Kuittinen *et al.*, 2004) to 13 cM in our map (Fig. 1); there are also some minor gaps. These gaps are not due to failure to find polymorphisms in genes from the corresponding region of the *A. thaliana* map, since, as explained above, we found suitable variants in 82% of all loci tested, and no region stands out as consistently failing to yield variants (see Supplementary Table 1). The large remaining gap in LG2 may therefore truly represent a region where sequences are missing in *A. thaliana*, perhaps indicating loss of some genes. In the corresponding *C. rubella* chromosome (B), a marker based on the *A. thaliana* chromosome 4 gene At4g08590 maps in this region (Boivin *et al.*, 2004). However, this is probably not due to loss of this gene from chromosome 1 of *A. thaliana*, because we found an *A. thaliana* chromosome 1 gene with a similar sequence (we did not map this gene, as in *A. thaliana* it is located close to At1g59720, near the end of the inversion, see Fig. 1).

Another possible reason for the gap is presence of an rDNA locus (nucleolar organizer region, NOR) in this region in *A. lyrata*. No such sequences are present in *A. thaliana* chromosome 1, so, if present in *A. lyrata*, they must have been lost in *A. thaliana*. An rDNA locus maps in this region in *C. rubella* (Boivin *et al.*, 2004). Because all cytologically detectable *C. rubella* NORs are at chromosome ends (Ali *et al.*, 2005), the mapping of rDNA in the middle of chromosome B suggests that in *C. rubella* this region may contain a small rDNA gene, not a large repetitive array. There is no cytological evidence for rDNA on *A. lyrata* LG2, and the number of cytologically detectable NOR loci differs between these two species (M. Lysak, personal communication), so there is no reason to expect one on this chromosome, but at present we cannot exclude presence of a small rDNA array on *A. lyrata* LG2.

(ii) *Differences in map distances and genome sizes in* A. lyrata *and* A. thaliana

If the physical sizes of the intervals do not differ between the two species compared here, we would expect map distances either to be similar (if the *A. thaliana* lineage lost self-incompatibility very recently, and has not had enough time for recombination rates to change), or else to be larger in *A. thaliana* (since the generally greater chiasma numbers per chromosome arm in inbreeding plants would presumably predict higher crossing-over rates per base pair). If the physical sizes of the intervals are smaller in *A. thaliana*, as suggested by the two species' genome size differences, we might expect map distances to be smaller in *A. thaliana*. These predictions are not, of course, precise, because crossover frequencies are clearly not strictly proportional to physical distances, since recombination hot- and cold-spots are known in humans (e.g. Jeffreys & May, 2004) and in other organisms, including *A. thaliana* (Drouaud *et al.*, 2005), and chiasma frequencies differ between *A. thaliana* strains, which probably do not differ greatly in the physical sizes of the chromosomes (Sanchez-Moran *et al.*, 2002).

The evolutionary time available is restricted to the time since the split between the ancestor of *A. thaliana* from related species, such as *A. lyrata*, since outgroups show that most of the differences evolved as changes in the *A. thaliana* lineage (see above). Any changes due to the transition to self-fertilization mentioned above must have occurred during the time when the *A. thaliana* lineage lost self-incompatibility. This time is not known, but evidence for selection for a non-functional pollen incompatibility gene suggests that it may be very recent (Shimizu *et al.*, 2004). This might make these *Arabidopsis* species unsuitable for testing these and other predictions about evolutionary changes in genomes driven by the evolution of self-fertilization. However, it is clear that a surprising number of chromosome fusions have occurred, and some increase in the recombination rate per base pair.

Even if sites of recombination and recombination rates per base pair have not changed, differences in the physical sizes of many map intervals probably have changed, given the large difference in genome

size between these two species. However, it is not known whether the smaller genome size of *A. thaliana* is due to pervasive, and presumably slow, small changes throughout the genome, affecting all intervals between genes similarly, or to rapid changes, perhaps due to a few events involving loss of large chunks of DNA (e.g. loss during chromosome rearrangement events, particularly fusions). There is some evidence for pervasive small changes in the *A. thaliana* genome (Sadowski & Quiros, 1998; Devos *et al.*, 2002); intron sizes tend to be smaller than in *A. lyrata* orthologues, based on comparing a sample of 19 genes, mostly from *A. thaliana* chromosome 1 (Wright *et al.*, 2002) and 41 further genes, mostly from chromosome 4 (A. Kawabe, unpublished data). Such changes could be due to selection favouring reduced DNA content of inter-genic regions and/or introns, perhaps to allow speedy cell division and the rapid growth required for a weedy lifestyle, or to loss of transposable elements in these regions, i.e. not to selection acting on the plant genome but instead to a changed balance between different selective forces acting on the elements themselves.

If the physical sizes of intervals have indeed become shorter in *A. thaliana*, the similarity of the map distances in LG1 and the corresponding *A. thaliana* chromosome arm is consistent with higher recombination in *A. thaliana*, as predicted. For the mapped region of the long arm of LG1, our estimate is 22·70 cM in *A. thaliana* versus 18·30 in *A. lyrata* (4·76 versus 3·84 cM/Mb, respectively, based on *A. thaliana* physical distances); the short arm of this chromosome has similar genetic map lengths in the two species. Our mapping could have detected a difference in map lengths much smaller than the estimated 50% difference in genome size. The length difference in introns is, however, slight (for *A. thaliana* chromosome 1, the mean reduction in intron size relative to the *A. lyrata* value is 5·0%, and for the somewhat smaller introns surveyed from chromosome 4 it is 3·1%; both data sets include one intron which is much larger in *A. thaliana*, so the difference in a 'typical' gene or intron would be slightly larger than these values). Inter-genic regions differ in size in the same direction (smaller in *A. thaliana*, in the FLC region of chromosome 4, see Sanyal & Jackson, 2005), but do not seem to differ much more than introns. It is not yet known whether this difference is general throughout the genomes of these species (Sadowski & Quiros, 1998; Devos *et al.*, 2002).

Another possible reason for the similarity in map distances is that recombination might occur largely within coding sequences. If so, any size reduction in *A. thaliana* non-coding regions would leave the map distances unaffected. Such a restriction of crossovers to the genes has previously been inferred from the similarity of genetic map lengths between other species with differing DNA amounts (Thuriaux, 1977), and indeed in maize most recombination seems to occur within genes (Schnable *et al.*, 1998; Xu *et al.*, 1995; Dooner & Martinez-Férez, 1997). However, this would not cause different chiasma numbers per chromosome arm between inbreeding and outcrossing species. Chiasma numbers have not yet been estimated in *A. lyrata*, and cannot therefore be compared between the two *Arabidopsis* species, but if they do differ, as in other related plants with different breeding systems, the similarity of the map distances will require some other explanation, and would suggest that sites of crossover events have changed.

For linkage group 2, distances between markers are consistently larger than those for corresponding *A. thaliana* regions. However, the difference is non-significant, based on our current data, and more markers and larger numbers of informative offspring are still needed to test whether this is real. Thus, only the LG1 results suggest more frequent crossing-over in the inbreeding *A. thaliana*. Another potential explanation for the larger LG2 map distances, compared with those in *A. thaliana*, might be the presence of nucleolar organizers. In *A. thaliana*, the two chromosomes with NORs (chromosomes 2 and 4) have lower recombination rates per base pair than the others (Copenhaver *et al.*, 1998), although a very high recombination rate in the NOR-carrying arm of chromosome 4 was found, distal to the heterochromatic knob on this chromosome (Drouaud *et al.*, 2005). An NOR is unlikely to be causing the difference observed here, because it is unlikely that either of the *A. lyrata* chromosomes mapped here carries a large NOR, though (as discussed above) we cannot exclude the presence of a small rDNA locus in *A. lyrata* LG2. Another possibility that should be tested in the future is that the DNA content of the LG2 region mapped here may be much larger in *A. lyrata* than its counterpart *A. thaliana*, i.e. that DNA content has changed across most inter-marker intervals.

(iii) *Conclusions*

Although our results suggest a recombination rate difference in LG1 in the predicted direction, a study of a single genus is clearly not definitive, and future studies in other closely related inbreeder/outbreeder pairs will be needed in order to test whether this is a general pattern. Given the possibility of using species whose genomes have been sequenced to develop markers in related species, including wild, non-model species, such comparisons are at last feasible, provided that genes tend usually to be in the same order, on largely non-rearranged chromosome arms. Evidence is accumulating, not just for synteny of genes, but for similar gene orders across large tracts of the genomes of related plants, despite many minor differences in

lengths of small regions (Ma *et al*., 2004; Bennetzen *et al*., 2005; Sanyal & Jackson, 2005).

The finding that the genes are largely in the same order, on largely non-rearranged chromosome arms, is encouraging in that it may be possible to use the *A. thaliana* physical map to find *A. lyrata* genes close to regions or genes of interest, for instance to obtain data on reference loci in the same genome region as a locus of interest – an approach that is becoming increasingly practicable in plants (e.g. Hackauf & Wehling, 2005). It is also encouraging for comparing the evolution of genes in regions with high and low recombination frequencies. It is of great interest to test whether sequence diversity is affected by the local recombinational environment, but this has not been possible in plants, because recombination rates per base pair are not known for most genome regions of outcrossing species (Baudry *et al*., 2001); inbreeding species are not predicted to show strong effects on diversity, because effective recombination rates are low due to inbreeding (Charlesworth *et al*., 1997). If crossover rates in homologous intervals are similar in the two *Arabidopsis* species, and if the overall difference in DNA content between these two species is mainly caused by small differences fairly uniformly distributed throughout the genome, it may be possible to use the whole genome size difference to correct for the difference in physical distances.

We thank S. I. Wright (York University, Toronto) for information about loci on *A. thaliana* chromosome 1 and for the primers, and O. Savolainen (University of Oulu, Finland) for DNA samples from the *A. lyrata* mapping family. Support is gratefully acknowledged of a postdoctoral grant from the Swedish Research Council and STINT (to B.H.), and a small research grant from the National Environmental Research Council of the UK (to D.C.).

References

Ali, H. B., Lysak, M. A. & Schubert, I. (2005). Chromosome localization of rDNA in the Brassicaceae. *Genome* **48**, 341–346.

Allen, K. D. (2002). Assaying gene content in *Arabidopsis*. *Proceedings of the National Academy of Sciences of the USA* **99**, 9568–9572.

Alonso-Blanco, C., Peeters, A. J. M., Koornneef, M., Lister, C., Dean, C., Bosch, N. v. d., Pot, J. & Kuipe, M. T. R. (1998). 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. *The Plant Journal* **14**, 259–271.

Arroyo, M. T. K. (1975). Electrophoretic studies of genetic variation in natural populations of allogamous *Limnanthes alba* and autogamous *Limnanthes floccosa* (Limnanthaceae). *Heredity* **35**, 153–164.

Arumuganathan, K. & Earle, E. D. (1991). Nuclear DNA content of some important plant species. *Plant Molecular Biology Reporter* **9**, 208–218.

Baudry, E., Kerdelhué, C., Innan, H. & Stephan, W. (2001). Species and recombination effects on DNA variability in the tomato genus. *Genetics* **158**, 1725–1735.

Bennett, M. D., Leitch, I. J., Price, H. J. & Johnston, J. S. (2003). Comparisons with *Caenorhabditis* (~100 Mb) and *Drosophila* (~175 Mb) using flow cytometry show genome size in *Arabidopsis* to be ~157 Mb and thus ~25% larger than the *Arabidopsis* genome initiative estimate of similar to 125 Mb. *Annals of Botany* **91**, 547–557.

Bennetzen, J. L., Coleman, C., Liu, R., Ma, J. & Ramakrishna, W. (2004). Consistent overestimation of gene number in complex plant genomes. *Current Opinion in Plant Biology* **7**, 732–736.

Bennetzen, J. L., Ma, J. X. & Devos, K. M. (2005). Mechanisms of recent genome size variation in flowering plants. *Annals of Botany* **95**, 127–132.

Boivin, K., Acarkan, A., Mbulu, R.-S., Clarenz, O. & Schmidt, R. (2004). The *Arabidopsis* genome sequence as a tool for genome analysis in Brassicaceae. A comparison of the *Arabidopsis* and *Capsella rubella* genomes. *Plant Physiology* **134**, 735–744.

Borts, R. H. & Haber, J. E. (1987). Meiotic recombination in yeast: alteration by multiple heterozygosities. *Science* **237**, 1459–1465.

Charlesworth, B. (1992). Evolutionary rates in partially self-fertilizing species. *American Naturalist* **140**, 126–148.

Charlesworth, B., Nordborg, M. & Charlesworth, D. (1997). The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided inbreeding and outcrossing populations. *Genetical Research* **70**, 155–174.

Charlesworth, D. & Charlesworth, B. (1995). Transposable elements in inbreeding and outbreeding populations. *Genetics* **140**, 415–417.

Charlesworth, D., Charlesworth, B. & Strobeck, C. (1977). Effects of selfing on selection for recombination. *Genetics* **68**, 213–226.

Charlesworth, D., Charlesworth, B. & Strobeck, C. (1979). Selection for recombination in self-fertilising species. *Genetics* **93**, 237–244.

Copenhaver, G. P., Browne, W. E. & Preuss, D. (1997). Measuring genome-wide recombination and centromere functions in *Arabidopsis* meiosis. *Molecular Biology of the Cell* **8**, 1557.

Copenhaver, G. P., Browne, W. E. & Preuss, D. (1998). Assaying genome-wide recombination and centromere functions with *Arabidopsis* tetrads. *Proceedings of the National Academy of Sciences of the USA* **95**, 247–252.

Copenhaver, G. P., Nickel, K., Kuromori, T., Benito, M. I., Kaul, S., Lin, X. Y., Bevan, M., Murphy, G., Harris, B., Parnell, L. D., McCombie, W. R., Martienssen, R. A., Marra, M. & Preuss, D. (1999). Genetic definition and sequence analysis of *Arabidopsis* centromeres. *Science* **286**, 2468–2474.

Devos, K. M., Brown, J. K. M. & Bennetzen, J. L. (2002). Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. *Genome Research* **12**, 1075–1079.

Dooner, H. K. & Martinez-Férez, I. M. (1997). Recombination occurs uniformly in the *bronze* gene, a recombination hotspot in the maize genome. *The Plant Cell* **9**, 1633–1646.

Drouaud, J., Camilleri, C., Bourguignon, P.-Y., Canaguier, A., Bérard, A., Vezon, D., Giancola, S., Brunei, D., Colot, V., Prum, B., Quesneville, H. & Mézard, C. (2005). Variation in crossing-over rates across chromosome 4 of *Arabidopsis thaliana* reveals the presence of meiotic recombination 'hot spots'. *Genome Research* **December 12**, doi: 10.1101/gr.4319006

Gao, L.-z. & Innan, H. (2004). Very low gene duplication rate in the yeast genome. *Science* **306**, 1367–1370.

Garber, E. D. (1956). The genus *Collinsia*. I. Chromosome numbers and chiasma frequencies of species in the two sections. *Botanical Gazette* **118**, 71–73.

Grant, V. (1958). The regulation of recombination in plants. *Cold Spring Harbor Symposia in Quantitative Biology* **23**, 337–363.

Hackauf, B. & Wehling, P. (2005). Approaching the self-incompatibility locus Z in rye (*Secale cereale* L.) via comparative genetics. *Theoretical and Applied Genetics* **110**, 832–845.

Hansson, B. & Kawabe, A. (2005). A simple method to score single nucleotide polymorphisms based on allele-specific PCR and primer-induced fragment-length variation. *Molecular Ecology Notes* **5**, 692–696.

Hedrick, P. W., Jain, S. K. & Holden, L. (1978). Multilocus systems in evolution. *Evolutionary Biology* **2**, 104–184.

Hosouchi, T., Kumekawa, N., Tsuruoka, H. & Kotani, H. (2002). Physical map-based sizes of the centromeric regions of *Arabidopsis thaliana* chromosomes 1, 2, and 3. *DNA Research* **9**, 117–121.

Jeffreys, A. J. & May, C. A. (2004). Intense and highly localized gene conversion activity in human meiotic crossover hot spots. *Nature Genetics* **36**, 151–156.

Johnston, J. S., Pepper, A. E., Hall, A. E., Chen, Z. J., Hodnett, G., Drabek, J., Lopez, R. & Price, H. J. (2005). Evolution of genome size in Brassicaceae. *Annals of Botany* **95**, 229–235.

Khrustaleva, L. I., Melo, P. E. d., Heusden, A. W. v. & Kik, C. (2005). The integration of recombination and physical maps in a large-genome monocot using haploid genome analysis in a trihybrid *Allium* population. *Genetics* **169**, 1673–1685.

Koch, M. & Kiefer, M. (2005). Genome evolution among Cruciferous plants: a lecture from the comparison of the genetic map of three diploid species – *Capsella rubella*, *Arabidopsis lyrata* subsp. *petraea*, and *A. thaliana*. *American Journal of Botany* **92**, 761–767.

Koch, M., Haubold, B. & Mitchell-Olds, T. (2000). Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis*, *Arabis* and related genera (Brassicaceae). *Molecular Biology and Evolution* **17**, 1483–1498.

Kuittinen, H., Haan, A. A. d., Vogl, C., Oikarinen, S., Leppälä, J., Koch, M., Mitchell-Olds, T., Langley, C. & Savolainen, O. (2004). Comparing the linkage maps of the close relatives *Arabidopsis lyrata* and *Arabidopsis thaliana*. *Genetics* **168**, 1575–1584.

Lande, R. (1984). The expected fixation rate of chromosomal inversions. *Evolution* **38**, 743–752.

Lister, C. & Dean, C. (1993). Recombinant inbred lines for mapping RFLP and phenotypic markers in *Arabidopsis thaliana*. *Plant Journal* **4**, 745–750.

Lysak, M. A., Pecinka, A. & Schubert, I. (2003). Recent progress in chromosome painting of *Arabidopsis* and related species. *Chromosome Research* **11**, 195–204.

Ma, J., Devos, K. M. & Bennetzen, J. L. (2004). Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. *Genome Research* **14**, 860–869.

Rees, H. (1961). Genotypic control of chromosome form and behaviour. *Botanical Review* **27**, 288–318.

Sadowski, J. & Quiros, Q. F. (1998). Organization of an *Arabidopsis thaliana* gene cluster on chromosome 4 including the *RPS2* gene, in the *Brassica nigra* genome. *Theoretical and Applied Genetics* **96**, 468–474.

Sanchez-Moran, E., Armstrong, S. J., Santos, J. L., Franklin, F. C. & Jones, G. H. (2002). Variation in chiasma frequency among eight accessions of *Arabidopsis thaliana*. *Genetics* **162**, 1415–1422.

Sanyal, A. & Jackson, S. A. (2005). Comparative genomics reveals expansion of the FLC region in the genus *Arabidopsis*. *Molecular Genetics and Genomics* doi: 10.1007/s00438-00500063-y.

Schnable, P. S., Hsia, A. P. & Nicolau, B. J. (1998). Genetic recombination in plants. *Current Opinion in Plant Biology* **1**, 123–129.

Shimizu, K. K., Cork, J. M., Caicedo, A. L., Mays, C. A., Moore, R. C., Olsen, K. M., Ruzsa, S., Coop, G., Bustamante, C. D., Awadalla, P. & Purugganan, M. D. (2004). Darwinian selection on a selfing locus. *Science* **306**, 2081–2084.

Stam, P. (1993). Construction of integrated genetic-linkage maps by means of a new computer package: JoinMap. *Plant Journal for Cell and Molecular Biology* **3**, 739–744.

Thuriaux, P. (1977). Is recombination confined to structural genes on the eukaryotic genome? *Nature* **268**, 460–462.

True, J. R., Mercer, J. M. & Laurie, C. C. (1996). Differences in crossover frequency and distribution among three sibling species of *Drosophila*. *Genetics* **142**, 507–523.

van Ooijen, J. W. & Voorrips, R. E. (2001). *JoinMap 3.0*, *Software for the Calculation of Genetic Linkage Maps*. Wageningen, the Netherlands: Plant Research International.

van Treuren, R., Kuittinen, H., Karkkainen, K., Baena-Gonzalez, E. & Savolainen, O. (1997). Evolution of microsatellites in *Arabis petraea* and *Arabis lyrata*, outcrossing relatives of *Arabidopsis thaliana*. *Molecular Biology and Evolution* **14**, 220–229.

Ventura, M., Archidiacono, N. & Rocchi, M. (2001). Centromere emergence in evolution. *Genome Research* **11**, 595–599.

Winckler, W., Myers, S. R., Richter, D. J., Onofrio, R. C., McDonald, G. J., Bontrop, R. E., McVean, G. A. T., Gabriel, S. B., Reich, D., Donnelly, P. & Altshuler, D. (2005). Comparison of fine-scale recombination rates in humans and chimpanzees. *Science* **308**, 107–111.

Wright, S. I. (2003). Effects of recombination rate and mating system on genome evolution and diversity in *Arabidopsis lyrata*. PhD thesis, University of Edinburgh.

Wright, S. I. & Schoen, D. J. (1999). Transposon dynamics and the breeding system. *Genetica* **107**, 139–148.

Wright, S. I., Le, Q. H., Schoen, D. J. & Bureau, T. E. (2001). Population dynamics of an Ac-like transposable element in self- and cross-fertilizing *Arabidopsis*. *Genetics* **158**, 1279–1288.

Wright, S. I., Lauga, B. & Charlesworth, D. (2002). Rates and patterns of molecular evolution in inbred and outbred Arabidopsis. *Molecular Biology and Evolution* **19**, 1407–1420.

Xu, X., Hsia, A. P., Zhang, L., Nikolau, B. J. & Schnable, P. S. (1995). Meiotic recombination break points resolve at high rates at the 5′ end of a maize coding sequence. *The Plant Cell* **7**, 2151–2161.

Yogeeswaran, K., Frary, A., York, T. L., Amenta, A., Lesser, A. H., Nasrallah, J. B., Tanksley, S. D. & Nasrallah, M. E. (2005). Comparative genome analyses of *Arabidopsis* spp.: inferring chromosomal rearrangement events in the evolutionary history of *A. thaliana*. *Genome Research* **15**, 505–515.

Zarchi, Y., Simchen, G., Hillel, J. & Schaap, T. (1972). Chiasmata and the breeding system in wild populations of diploid wheats. *Chromosoma* **38**, 77–94.