



Comparative genome analyses of *Arabidopsis* spp.: Inferring chromosomal rearrangement events in the evolutionary history of *A. thaliana*

Krithika Yogeeswaran, Amy Frary, Thomas L. York, et al.

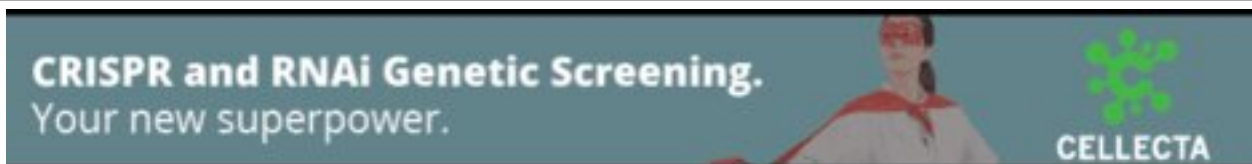
Genome Res. 2005 15: 505-515

Access the most recent version at doi:[10.1101/gr.3436305](https://doi.org/10.1101/gr.3436305)

References This article cites 49 articles, 22 of which can be accessed free at:
<http://genome.cshlp.org/content/15/4/505.full.html#ref-list-1>

License

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).



To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

Cold Spring Harbor Laboratory Press

Comparative genome analyses of *Arabidopsis* spp.: Inferring chromosomal rearrangement events in the evolutionary history of *A. thaliana*

Krithika Yogeeswaran,¹ Amy Frary,⁴ Thomas L. York,² Alison Amenta,^{4,5} Andrew H. Lesser^{1,6} June B. Nasrallah,¹ Steven D. Tanksley,³ and Mikhail E. Nasrallah^{1,7}

¹Department of Plant Biology, ²Department of Biological Statistics and Computational Biology, and ³Department of Plant Breeding, Cornell University, Ithaca, New York 14853, USA; ⁴Department of Biological Sciences, Mount Holyoke College, South Hadley, Massachusetts 01075, USA

Comparative genome analysis is a powerful tool that can facilitate the reconstruction of the evolutionary history of the genomes of modern-day species. The model plant *Arabidopsis thaliana* with its $n = 5$ genome is thought to be derived from an ancestral $n = 8$ genome. Pairwise comparative genome analyses of *A. thaliana* with polyploid and diploid *Brassicaceae* species have suggested that rapid genome evolution, manifested by chromosomal rearrangements and duplications, characterizes the polyploid, but not the diploid, lineages of this family. In this study, we constructed a low-density genetic linkage map of *Arabidopsis lyrata* ssp. *lyrata* (*A. l. lyrata*; $n = 8$, diploid), the closest known relative of *A. thaliana* (MRCA ~5 Mya), using *A. thaliana*-specific markers that resolve into the expected eight linkage groups. We then performed comparative Bayesian analyses using raw mapping data from this study and from a *Capsella* study to infer the number and nature of rearrangements that distinguish the $n = 8$ genomes of *A. l. lyrata* and *Capsella* from the $n = 5$ genome of *A. thaliana*. We conclude that there is strong statistical support in favor of the parsimony scenarios of 10 major chromosomal rearrangements separating these $n = 8$ genomes from *A. thaliana*. These chromosomal rearrangement events contribute to a rate of chromosomal evolution higher than previously reported in this lineage. We infer that at least seven of these events, common to both sets of data, are responsible for the change in karyotype and underlie genome reduction in *A. thaliana*.

[Supplemental material is available online at www.genome.org. The following individuals kindly provided reagents, samples, or unpublished information as indicated in the paper: C. Langley.]

Comparative genome analysis can be a powerful tool to address questions pertaining to the evolution of the structure of genomes (Paterson and Bennetzen 2001). In plant families that include model taxa, such as the *Brassicaceae*, the *Solanaceae*, and the *Poaceae*, comparative mapping studies allow evolutionary inferences related to the nature and number of rearrangements that distinguish the genomes of the model species from those of their less well-studied relatives (Bonierbale et al. 1988; Ahn and Tanksley 1993; Prince et al. 1993; Chittendan et al. 1994; Kurata et al. 1994; Periera et al. 1994; Lagercrantz and Lydiat 1996; Gale and Devos 1998; Lagercrantz 1998; Doganlar et al. 2002). Not only do such inferences facilitate the transfer of knowledge from model taxa to their relatives, but they also provide insight into the process of genome restructuring that can lead to reproductive isolation and ultimately speciation (Rieseberg 2001; Hall et al. 2002). The larger the number of taxa analyzed within a family, the more comprehensive the information is for inferring the nature of ancestral genomes and the evolutionary history that led to the speciation of the members of that family.

Present addresses: ⁵Dept. of Neuroscience, Brown Univ., Providence, RI 02912, USA; ⁶Dept. of Natural Resources, Cornell Univ., Ithaca, NY 14853, USA.

⁷Corresponding author.

E-mail men4@cornell.edu; fax (607) 255-5407.

Article and publication are at <http://www.genome.org/cgi/doi/10.1101/gr.3436305>.

The *Brassicaceae* (or *Cruciferae*) is a dicot family divided into 13 tribes with a total of 360 genera (Al-Shehbaz 1973) including the agronomically important genus *Brassica* and the model species *Arabidopsis thaliana*. The near-complete annotation of the genomic sequence of *A. thaliana*, including detailed information of physical map position, copy number of genes and intergenic sequences, and location of duplicated chromosomal segments (The *Arabidopsis* Genome Initiative 2000; Blanc et al. 2000; Vision et al. 2000; Bowers et al. 2003; Ermolaeva et al. 2003; Raes et al. 2003) provides a solid foundation for comparative mapping studies within this family. *A. thaliana*, prized for its small genome, is actually an anomaly within the tribe *Sisymbrieae* of the *Brassicaceae*. A base chromosome number of 8 is inferred to be the ancestral state in the tribe (Koch et al. 1999), and reduction in chromosome number to $n = 5-7$ is assumed to be a derived state (Koch et al. 1999, 2000). *A. thaliana*, with its five chromosomes, is the single most extreme example of this genome reduction in the tribe.

The nature and rate of occurrence of the chromosomal events that generated the reduced *A. thaliana* genome from the ancestral $n = 8$ genome are not well understood. As many as 90 rearrangements are estimated to differentiate *Brassica nigra* [$n = 8$] from *A. thaliana* (Lagercrantz 1998). Two recent mapping studies in which the *A. thaliana* genome was compared with the *Capsella* genome ($n = 8$; Boivin et al. 2004) and with the *Arabi-*

dopsis lyrata ssp. *petraea* genome (*A. l. petraea*; $n = 8$; Kuittinen et al. 2004), suggest that genome reduction was the result of chromosomal fusion rather than chromosomal elimination. Whatever the nature of these rearrangements, they must have occurred ~5 Mya, the estimated time of divergence (Koch et al. 2000, 2001; Kuittinen and Aguade 2000) between *A. thaliana* and its closest known relative *A. lyrata* ($n = 8$, genome ~116%–123% larger than that of *A. thaliana*; Dart et al. 2004; Schmuths et al. 2004).

Further insight into the evolutionary history of the *A. thaliana* genome is likely to be gained from a three-way comparison of the genomes of *A. thaliana* and its close relatives, *A. lyrata* and *Capsella* spp. In the study reported here, we used the wild, perennial, outcrossing *Arabidopsis lyrata* subspecies *lyrata* (hereafter *A. l. lyrata*; $n = 8$, diploid; Dart et al. 2004) which occurs in North America (in contrast to the *A. l. petraea* subspecies which occurs in Europe; Ramos-Onsins et al. 2004), to construct a low-density genetic map with 104 restriction fragment length polymorphism (RFLP) markers distributed on eight linkage groups, and we compared this map to the *A. thaliana* genome. We then performed comparative statistical analyses of rearrangements observed in *A. thaliana* versus *A. l. lyrata* and *A. thaliana* versus *Capsella* (based on data from Boivin et al. 2004), using an extension of the Bayesian method reported by Durrett et al. (2004). This method assigns probability estimates to the parsimony solutions as well as alternative scenarios, and is therefore more powerful than parsimony alone to infer the most probable events that underlie genome evolution. Based on these studies, the minimum number of major rearrangements that distinguish *A. thaliana* from the $n = 8$ genomes and their rate of occurrence are inferred.

Results

Genetic mapping of *A. l. lyrata* using *A. thaliana*-specific markers

A total of 432 EST markers were screened against our mapping population of 67 F₂ plants (see Methods), of which only 144 showed codominant polymorphisms. For the sake of comparison with other maps and to increase the utility of our map for the research community, all EST markers were named based on the gene locus ID number provided by the *Arabidopsis* Genome Initiative (2000) for the *A. thaliana* gene they represent (e.g., EST ATTS1092, which corresponds to locus At1g02150, is abbreviated as marker 1_02150). All mapping data were analyzed with Mapmaker version 3.0 (Lander et al. 1987) using the Kosambi cM function for genetic linkage analysis. When the LOD score was set at 3.0, 104 markers resolved into the expected eight linkage groups, hereafter designated Aly LG 1–8 (Fig. 1; see Supplemental Table S1 for details of mapped markers). The genetic map covers a total length of 406.6 cM with complete linkage groups ranging from 47.3 to 77.3 cM in length. The average spacing between markers is ~4 cM, with a range of 0.0–17.4 cM.

Several >20-cM gaps were found during preliminary mapping. All gaps were systematically filled, with the exception of a few regions that were consistently found to be monomorphic. The most pronounced monomorphic stretch of this nature extended over parts of Aly LGs 1 and 2, which corresponds to ~16 Mbp of the pericentric region of *A. thaliana* Chromosome I (At Chr I). Within this gap, 98 markers (76 EST-based and 22 RFLP markers previously used to map the corresponding region in the

Capsella map by Boivin et al. 2004) were found to be monomorphic in our *A. l. lyrata* mapping population.

Due to the homozygosity of all tested coding sequences in this gap, noncoding sequences were assessed in an effort to identify polymorphic markers. Previous work had suggested that many PCR primers designed to amplify simple sequence-length repeats (SSRs) from *A. thaliana* could amplify similar regions in *A. l. lyrata* as well (van Treuren et al. 1997; Nasrallah et al. 2000). Based on published reports (Lukowitz et al. 2000; Claus et al. 2002; http://genome.bio.upenn.edu/SSLP_info/SSLP.html), we selected 13 SSRs corresponding to the gap on At Chr I, some of which had been reported to amplify *A. lyrata*. To bridge the gap further, we also located five novel SSRs within this region (see Methods). However, none of the 18 SSRs screened were polymorphic in our mapping population, suggesting that both noncoding and coding sequences in this region are monomorphic. Additionally, a 25S rDNA-specific probe used to determine the position of the nucleolar organizing regions (NORs) in *A. l. lyrata* was also monomorphic, precluding the mapping of the NORs in our study. It should be noted that this unexpected lack of polymorphism is concordant with a recent multilocus analysis of variation and speciation among *Arabidopsis* spp., which indicated that *A. l. lyrata* has surprisingly low levels of variability for an outcrossing species, suggesting that a bottleneck occurred during its colonization of the New World from Eurasia (Ramos-Onsins et al. 2004).

The expected genotypic segregation ratio for codominant markers in an F₂ population is 1:2:1 with parental alleles occurring in approximately equal numbers. χ^2 -tests performed for 97 mapped markers with unambiguous scoring data demonstrated that a total of 20 marker loci (~21%) showed significant deviation from expected segregation ratios at a locus-by-locus significance level of $P = 0.05$. Deviations were observed in genotypic ratio for one locus, in overall allelic ratios for seven loci, and in both genotypic and overall allelic ratios for 12 loci (highlighted in Fig. 1). Ten loci on Aly LG 6 and six loci on Aly LG 3 showed skewing in genotypic or allelic ratio in favor of alleles from one parent (P1). Marker 5_34940 on LG 6 also showed skewing favoring P1 that is almost significant at a locus-by-locus significance level of $P = 0.05$ (0.052). On Aly LG 7, three loci showed deviation from expected ratios favoring the other parent (P2), and one locus demonstrated a significant overrepresentation of heterozygous individuals (12:33:5) and underrepresentation of the P2 homozygous class. Other mapping populations of *A. thaliana* relatives have also been reported to exhibit a large number of loci showing transmission distortion despite the range of different strategies used to generate them (Boivin et al. 2004; Kuittinen et al. 2004 and references therein).

Comparison of the genetic linkage map of *A. l. lyrata* with the physical map of *A. thaliana*

The comparative map illustrated in Figure 2 shows the gross chromosomal organization of *A. l. lyrata* relative to *A. thaliana* and the high degree of synteny between these two genomes. Barring small-scale rearrangements and duplications defined by single markers, the mapped markers assign ~79.2 Mbp of *A. thaliana* sequence (corresponding to 325 cM of the Aly map) to 16 colinear blocks of two or more markers (average length of a colinear block in *A. l. lyrata* = 20.3 cM, highlighted in Fig. 2).

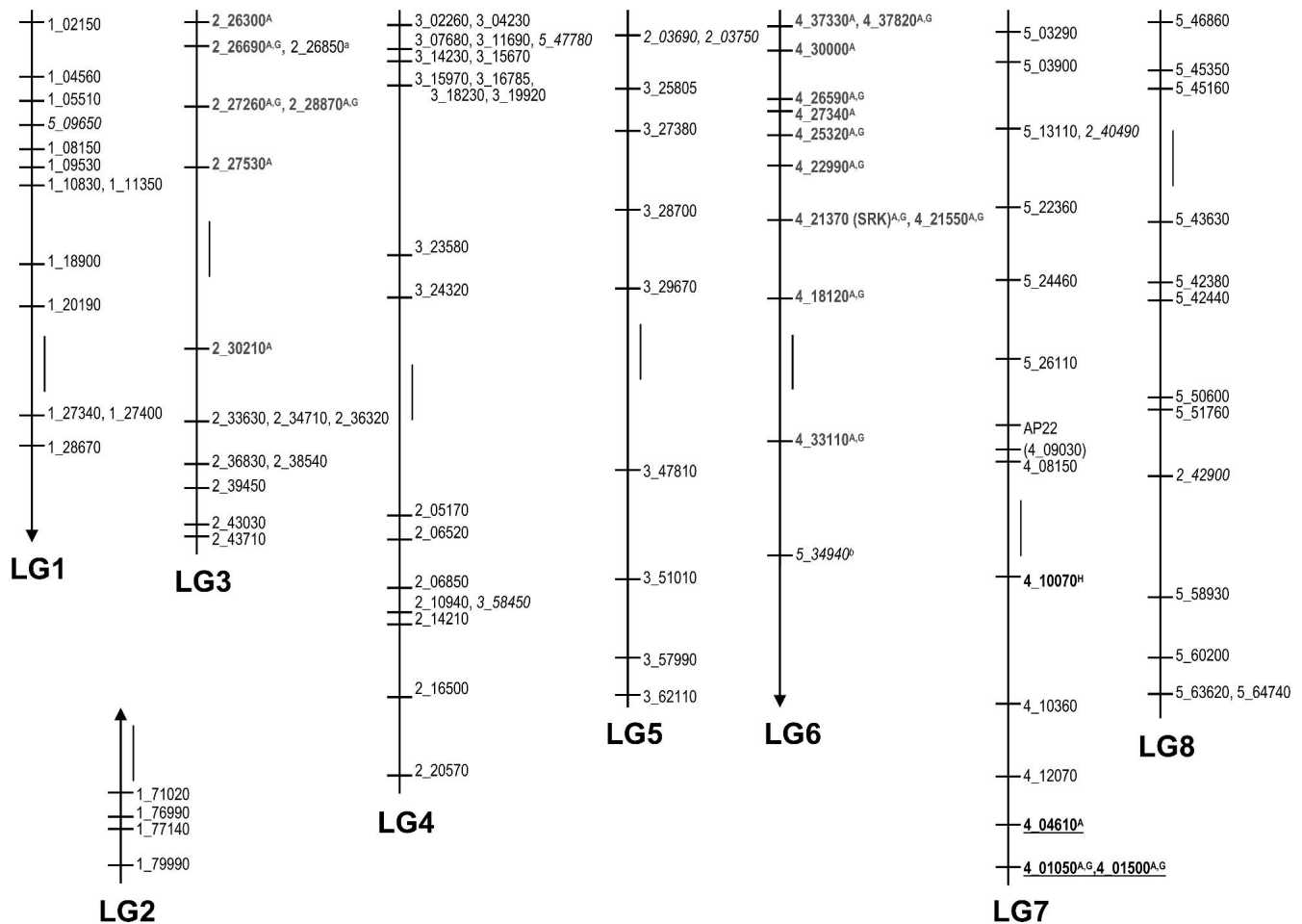


Figure 1. *A. lyrata* linkage map and segregation ratio distortion. The eight linkage groups of *A. lyrata* as resolved by the mapping of 104 markers are numbered LG1–8 at LOD 3 (markers in parentheses have LOD scores just below 3.0). Figure drawn to scale with vertical bars representing 5-cM mapping distance. Marker names signify the gene locus they represent (1_02150 = At1g02150). Italicized markers demonstrate disruption in synteny due to translocation or duplication followed by deletion. Markers in bold demonstrate transmission ratio distortion. Markers in gray show skewed segregation in favor of P1 alleles, and those that are underlined in favor of P2 alleles. ^A, distortion of allelic ratios; ^C, distortion of genic ratios; ^H, distortion ratio overrepresenting heterozygotes and underrepresenting P2 alleles. ^a, scoring data does not allow clear assessment of segregation ratios; ^b, allelic ratio nearly significantly skewed in favor of P1 alleles (0.052).

The simplest scenario for genome reduction from $n = 8$ to $n = 5$ in *A. thaliana* involves at least three chromosomal fusions and two reciprocal translocations (Fig. 2). The three fusions include a fusion (F1) between Aly LGs 1 and 2 to generate At Chr I, a second fusion (F2) between Aly LGs 3 and 4 to generate At Chr II, and a third fusion (F3) between Aly LGs 6 and 8 to generate part of At Chr V. In the first reciprocal translocation (T1), a likely event was fission at breakpoints towards the middle of Aly LG 4 and the top of Aly LG 5 followed by reciprocal exchange, which generated At Chr III and part of At Chr II. The second translocation (T2) probably involved breakage and reciprocal exchange between Aly LGs 6 and 7 to give rise to At Chr IV and part of At Chr V (Fig. 2). Additionally, within the limits of our map's resolution, at least two major inversions (I1: Aly LG 7, 34.6 cM and I2: Aly LG 8, 25.7 cM) and one minor inversion (i1: Aly LG 3, 5.4 cM) are predicted to have occurred, even at a higher LOD threshold of 5.0. Finally, the Aly LG 4 was found to contain a large gap (17.5 cM) flanking the breakpoint that gave rise to the upper ends of At Chr II and III, despite extensive end-to-end marker coverage of these two regions.

Forty-nine (46.7%) of the mapped markers were found to be either truly single-copy or to correspond to tandemly duplicated genes in *A. thaliana*, and their location on the *A. lyrata* map were clearly and directly ascertained. An additional 54 markers corresponded to members of gene families (see Methods) in the *A. thaliana* genome for which only single polymorphisms were detected in *A. lyrata*. Forty-nine of these markers mapped to a region of an Aly linkage group that was syntenic with an *A. thaliana* region containing one paralog. However, markers specific for paralogs of other gene families in *A. thaliana* showed complex hybridization patterns, likely reflecting polymorphisms at multiple paralogous loci. All but two of these (5_09650 and 3_58450) were considered multicopy and therefore were not used in map construction as they could not be unambiguously scored.

Statistical analysis of comparative genome mapping data and estimation of rates of chromosomal evolution

Bayesian analyses of our *A. thaliana*–*A. lyrata* comparative mapping data were performed under three permutations, namely

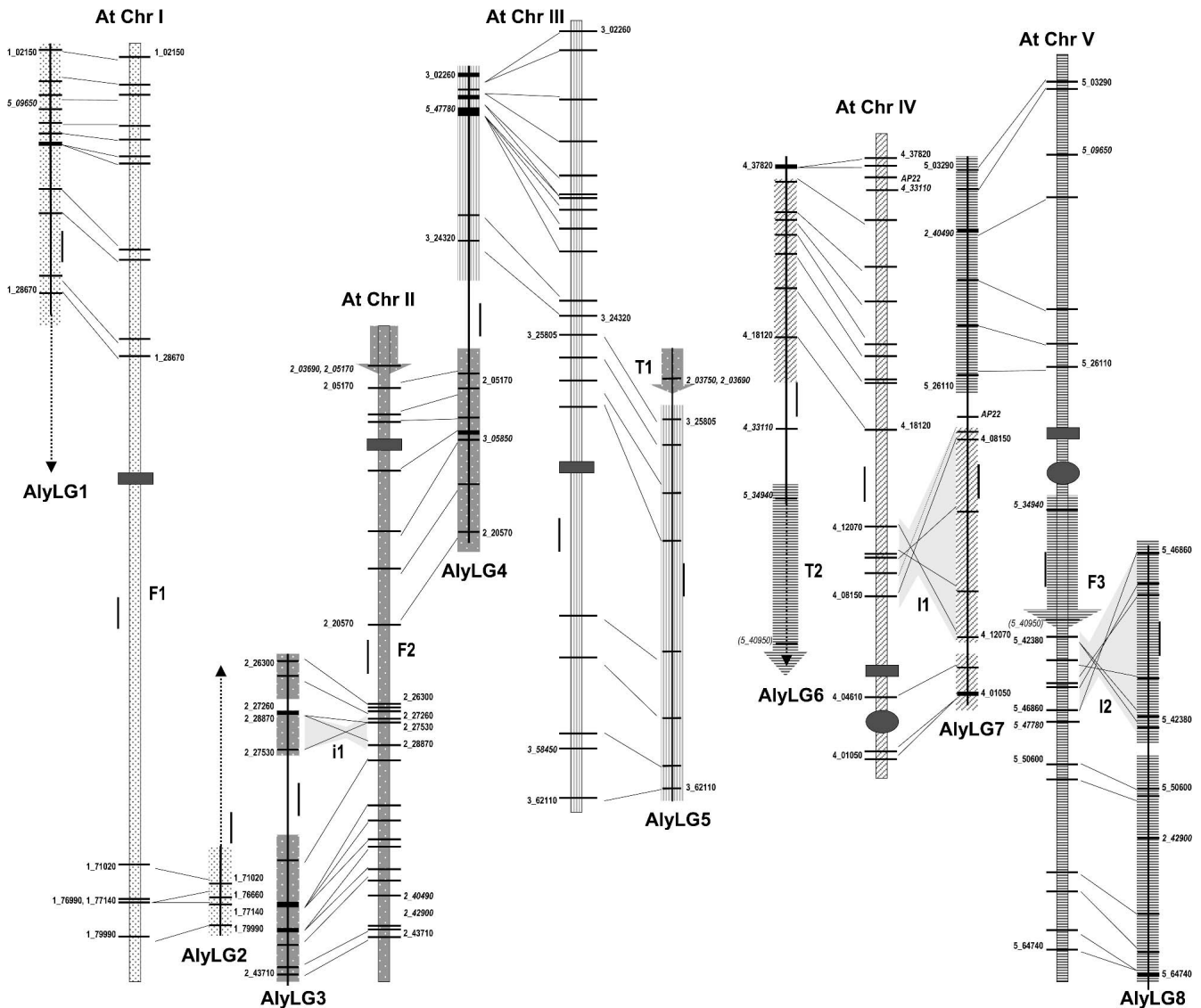


Figure 2. Colinearity of *A. lyrata* linkage map with the *A. thaliana* genome. *A. thaliana* chromosomes (At Chr I–V) are represented as patterned bars (drawn to scale, 1 unit = 1 Mbp; gray rectangles, centromeres; gray circles, heterochromatic knobs). *A. lyrata* linkage groups (Aly LG 1–8) are shown in black (drawn to scale, 1 unit = 5 cM). Sixteen colinear blocks are highlighted with the same pattern as the At chromosome to which they correspond. Markers defining the ends of each colinear block are shown on the map in black lettering. Markers mapping with LOD score less than 3.0 are featured in parentheses. Italicized markers map to translocated or nonsyntenic regions in *A. lyrata*. Translocations T1 and T2 are highlighted by arrows whose patterns correspond to the At chromosome where their colinear region lies. Major inversions I1 and I2 and minor inversion i1 are highlighted in light gray. Three chromosomal fusions are denoted as F1–F3.

marker order (1) without distances, (2) with distances in Mbp on the *A. thaliana* genome only, and (3) with distances in cM on the *A. lyrata* genome only (see Methods). A similar set of analyses were also performed using the raw mapping data from the *A. thaliana*–*Capsella* comparative mapping study by Boivin et al. (2004). The 95% credible intervals (CIs) for observable rearrangement events, namely inversions (Li), translocations (Lt), fusions/fissions (Lf), and total rearrangements (L), from all six analyses are summarized in Table 1 (see Methods). In all six analyses there was high probability in favor of only three fusion/fission events having occurred (Lf = [3]) with <~1% probability of observing more than three fusions/fissions.

Plots of the estimated posterior probability distributions for the total number of observable rearrangements (L) are shown in

Figure 3 for (A) *A. thaliana* versus *A. lyrata* and (B) *A. thaliana* versus *Capsella* under the simplest case of marker order without distance (see Supplemental Fig. S1 for plots from the other four analyses). There was strong consensus of the 95% CIs for observable rearrangement events, albeit with small deviations, from these two comparisons considering marker order only. In going from either *A. lyrata* or *Capsella* to *A. thaliana* there are at least 10 major rearrangements. The joint posterior probability, estimated for various combinations of inversions and translocations for mapping data with marker order but without distances, is summarized in Table 2. There is high joint posterior probability in favor of five inversions and two translocations for observable rearrangements, specifically 71.8% and 78.8% for the At–Aly and At–Cap comparisons, respectively. Therefore, in both cases, the

Table 1. 95% Credible intervals for observable rearrangement events

Genomes/95% CI for rearrangements	Marker order (no distances)	Marker order with At distances in Mbp	Marker order with Aly or Cap distances in cM
<i>A. thaliana</i> vs. <i>A. lyrata</i>			
L = Total rearrangements	[10, 11]	[10, 13]	[10, 12]
Li = Inversions	[4, 7]	[4, 7]	[4, 7]
Lt = Translocations	[2, 3]	[2, 4]	[2, 4]
Lf = Fusions/fissions	[3]	[3]	[3]
<i>A. thaliana</i> vs. <i>Capsella</i> spp.			
L = Total rearrangements	[10, 11]	[10, 12]	[10, 12]
Li = Inversions	[4, 6]	[4, 7]	[4, 7]
Lt = Translocations	[2, 3]	[2, 3]	[2, 3]
Lf = Fusions/fissions	[3]	[3]	[3]

Note: 95% of the posterior probability mass falls within the intervals defined above (i.e., L = [10, 11] denotes a 95% posterior probability of 10–11 total rearrangements having occurred).

most likely scenario is what we may refer to as the parsimony scenario, with the minimal numbers of total rearrangements (10), fusions/fissions (three), translocations (two), and therefore $10 - 3 - 2 = 5$ inversions. The Bayesian analysis also suggests that the next likely scenario, involving four inversions and three translocations, has only ~11.5% probability of occurring in both cases (see Table 2).

Incorporating distances between markers from one or the other genome in these data sets results in a small but significant shift in posterior probability mass in favor of alternative scenarios (see Supplemental Table S2). Some of these scenarios involve greater than parsimony estimates of inversions and/or translocations to give > 10 observable rearrangements. Alternative scenarios observed to gain some support by incorporating distances, include one involving two translocations and seven inversions (up to 8.8% in At-Aly and 5.2% in At-Cap) and one involving three translocations and five inversions (up to 10.7% in At-Aly). However, in all cases the most likely scenario predicted by the Bayesian analyses with distances is still the parsimony scenario.

Figure 4 illustrates the rearrangement steps in the parsimony solutions. Seven rearrangement steps (A–G) were found to be shared by the *A. thaliana*–*A. l. lyrata* and *A. thaliana*–*Capsella* spp. genome data, suggesting that a minimum of seven steps underlie genome reduction in the *A. thaliana* lineage ($n = 5$). These include two reciprocal translocations (B = T1 and D = T2), three fusion events (A = F1, C = F2, and E = F3), and at least two inversions F (= I2) and G (= I3).

H–J represent inversions deduced only from the *A. thaliana*–*A. l. lyrata* comparative map, and V–Z represent rearrangements deduced only from the *A. thaliana*–*Capsella* spp. comparative map. Therefore, both sets of data show the most probable number of inversions to be five, two in common and three unique to each set of data. Whether the latter rearrangements are unique to the *Arabidopsis* or *Capsella* lineages cannot be ascertained, since detailed marker information is lacking for the corresponding regions on one of the maps. I and J represent a case where an inversion within an inversion is required to explain the marker arrangement observed with high LOD score (5.0) on the At-Aly map.

Applying the largest 95% CIs for rearrangement events (L) predicted by three permutations of the Bayesian analyses, the rates of chromosomal evolution for At-Aly and At-Cap were estimated to be ~0.9–1.7 rearrangements/Myr/genome and ~0.5–0.6 rearrangements/Myr/genome (Table 3), respectively, with the assumption of

divergence times from *A. thaliana* of 3.8–5.8 Myr for *A. lyrata* (Kuittinen and Aguade 2000) and ~10 Myr for *Capsella* (Koch et al. 2000, 2001). These estimated rate ranges include the rates predicted by the parsimony estimate of the number of rearrangements, namely 0.9–1.3 and 0.5 rearrangements/Myr/genome for the *A. thaliana*–*A. lyrata* and *A. thaliana*–*Capsella* comparisons, respectively.

Discussion

Intraspecific genome comparisons of the *A. l. lyrata* and *A. l. petraea* subspecies

The *A. l. lyrata* genetic linkage map we generated is ~406 cM in length and contains 104 markers, while the *A. l. petraea* genetic

linkage map generated by Kuittinen et al. (2004) is ~515 cM in length and contains 72 markers. The average LG length determined in the two mapping studies (64.4 cM for *A. l. petraea*, Kuittinen et al. 2004; 60.8 cM for *A. l. lyrata* LGs 3–8, from this study) are quite similar, and we estimate the *A. l. lyrata* map, with full resolution of LGs 1 and 2, to be ~487 cM long. This difference in the overall map lengths inferred for the two subspecies falls well within the range of the 20% variation in map lengths often produced by analysis of different populations of the same species (Burr et al. 1988; Beavis and Grant 1991), and thus likely reflects variation in some of the genetic or environmental factors that are known to affect recombination frequencies (de Vienne 2003).

Three-way comparisons of the *A. l. lyrata*, *A. l. petraea*, and *Capsella* maps

While the markers used in our mapping study were, for the most part, different from those used to map *Capsella* (Boivin et al. 2004) and *A. l. petraea* (Kuittinen et al. 2004), the physical positions of the three sets of markers, with reference to the *A. thaliana* genome, jointly reveal overall gross conservation of marker order and chromosomal structure between these $n = 8$ genomes, with a few minor differences (see Fig. 2; refer to the Methods section for correspondence of LGs between maps). Thus, despite their resolution, these three maps are likely to be collectively useful for predicting the location of specific loci in $n = 8$ crucifers. A case in point relates to the NORs, which we were unable to map in our study. The NORs, which are found at subtelomeric positions on

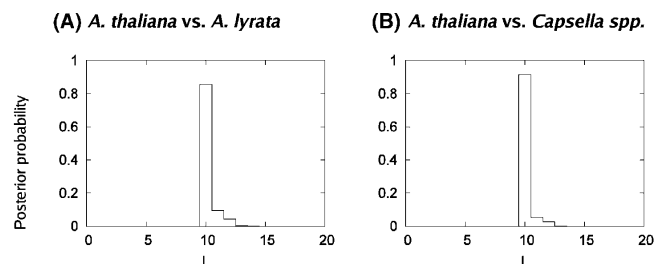


Figure 3. Posterior probability plots of *A. thaliana*–*A. l. lyrata* and *A. thaliana*–*Capsella* mapping data considering marker order only. Posterior probability plots generated from Bayesian analyses of mapping data accounting for marker orders without distances. (A) *A. thaliana*–*A. l. lyrata* comparison. (B) *A. thaliana*–*Capsella* comparison. x-axis: L = estimate of the total number of observable rearrangements; y-axis: Posterior probability estimate for the number of observable rearrangements.

Table 2. Joint posterior probabilities for inversions and translocations (marker order only)

(A) <i>A. thaliana</i> – <i>A. l. lyrata</i>						
Li/Lt	1	2	3	4	5	6
1	—	—	—	—	—	—
2	—	—	—	—	0.3	—
3	—	—	—	1.8	0.1	—
4	—	—	11.6	0.8	0.1	—
5	—	71.8	3.6	0.5	—	—
6	—	5.0	0.6	—	—	—
7	—	3.2	0.1	—	—	—
8	—	0.2	—	—	—	—
9	—	0.1	—	—	—	—
10	—	—	—	—	—	—
(B) <i>A. thaliana</i> – <i>Capsella</i> spp.						
Li/Lt	1	2	3	4	5	6
1	—	—	—	—	—	—
2	—	—	—	—	0.1	—
3	—	—	—	1.2	0.1	—
4	—	—	11.5	0.5	0.1	—
5	—	78.8	2.1	0.3	—	—
6	—	2.7	0.3	—	—	—
7	—	2.1	0.1	—	—	—
8	—	0.1	—	—	—	—
9	—	—	—	—	—	—

Percentage joint posterior probabilities (unit: %) for various combinations of observable inversions and translocations generated from Bayesian analyses of (A) *A. thaliana*–*A. lyrata* and (B) *A. thaliana*–*Capsella* spp. mapping data, considering marker orders only. Combinations with joint posterior probability of <0.1% are denoted with “—”. The percentage posterior probability for the combinations of inversions (Li) and translocations (Lt) within their respective 95% CIs are shown in bold, and the most likely scenario for these rearrangements is italicized in each data set.

the short arms of At Chr II and IV of *A. thaliana* (Copenhaver and Pikaard 1996), were localized on the *Capsella* map to the center of linkage group B (=Aly LG2) and one end of linkage group G (=Aly LG6), and it is likely that they are located in the corresponding regions of the *A. lyrata* genome as well.

Importantly, three-way comparisons of these maps allow substantiation of putative chromosomal rearrangements defined by one or two markers in one map based on the occurrence of similar rearrangements in one or more of the other maps. They can also compensate for the deficiencies of any one of these low-density genetic maps and identify chromosomal rearrangements that escaped detection on a particular map, and they can clarify the time of occurrence of these rearrangements, at least in some cases. Several pertinent inferences of these three-way comparisons are discussed in the following sections.

Major inversions

Two large inversions relative to the *A. thaliana* map appear consistently on the *Capsella* and at least one of the *A. lyrata* maps. One inversion (I1) on Aly LG 7 was detected on our *A. l. lyrata* map as well as in the *Capsella* and *A. l. petraea* maps, though the extent of this inversion is best defined on our map (~34.6 cM = ~2.08 Mbp on *A. thaliana*). A second large inversion (I2) on Aly LG 8 was observed on the *Capsella* map and our map, but escaped detection on the *A. l. petraea* map, likely due to the mapping of only one marker within the inverted region. Conservation of marker order in these regions of *A. lyrata* and *Capsella* suggests that they represent the ancestral state, and that inversions occurred in the *A. thaliana* lineage after its divergence from *A. lyrata* to give rise to the marker order observed in the *A. thaliana* genome.

In contrast, comparison of the *Capsella* and *A. lyrata* maps offered no insight on whether the large inversion observed at the upper end of LG B in the *Capsella* map (=Aly LG2) was specific to the *Capsella* lineage, and therefore postdated the divergence of *Capsella* and *Arabidopsis*. This inversion was observed neither on our *A. l. lyrata* map, where it falls within the region of missing information, nor in the *A. l. petraea* map on which this region was defined by only one mapped marker.

Reciprocal translocations

All three maps identified two major translocations in *Capsella* and *A. lyrata* relative to *A. thaliana*. One translocation, T1 (Fig. 2), involved the transfer of a small chromosomal segment from the top of Aly LG 5 to the top of At Chr II. A second major translocation, T2, from the bottom of Aly LG 6 to the center of At Chr V, is resolved in better detail on the *Capsella* and *A. l. petraea* maps than on the *A. l. lyrata* map, largely because most markers screened in this region were monomorphic in our mapping population, with only one marker, 5_34940, defining this region at LOD 3 and 5. Marker 5_40950 was also found to map to the end of Aly LG 6 as part of this translocation, but its position remains tentative (indicated by parentheses on the comparative map in Fig. 2) because the distance between 5_34940 and 5_40950 was very long (19.5 cM) and the LOD score below 3.0. On closer inspection of the data, we found that unlike the other 11 markers on LG 6, this marker did not show skewed segregation in favor of P1 alleles, possibly exaggerating its apparent genetic distance from other linked markers. Based on our mapping data, the translocated fragment includes, at the very least, the 3.2-Mbp region of At Chr V between genes 5_34940 and 5_40950 (all markers on the corresponding regions of the other maps fall within these boundaries).

Small-scale rearrangements

While several small-scale genome rearrangements are likely to have been missed in all three mapping studies, a few small-scale rearrangements were observed in only one or the other of the three maps, as expected for low-density maps generated with different markers. For example, a small inversion on the bottom end of *Capsella* LG H (=Aly LG 8) involving 0.11 Mbp containing the orthologs of At5g66960–At5g67260 was not detected on our *A. l. lyrata* map, since the most terminal marker we mapped, 5_64740, lies just before this region. A small inversion (i1) on Aly LG 3 (from 2_27530 to 2_28870) is not observed on *Capsella* LG D, where neither these nor other intervening markers are mapped. In such cases we cannot ascertain whether the rearrangements are specific to *Capsella* or *A. lyrata*, and therefore cannot infer the approximate time of occurrence for these arrangements.

Another interesting observation pertains to the major inversion I1 on Aly LG 7. Our data suggest that within this primary inversion (Aly: 4_08150, 4_09030, 4_10070, 4_10360, 4_12070), there might have been a small-scale secondary inversion (Aly: 4_09030, 4_08150, 4_10070, 4_10360, 4_12070) that restored the marker order (with reference to *A. thaliana*) to part of the region. However, this possibility must be confirmed by more detailed mapping in this region, since marker 4_09030 was mapped with an LOD score just below 3, and both the *Capsella* and *A. l. petraea* maps are poorly resolved in this region.

The *A. l. petraea* mapping data (Kuittinen et al. 2004) suggest the presence of three minor inversions relative to *A. thaliana* on

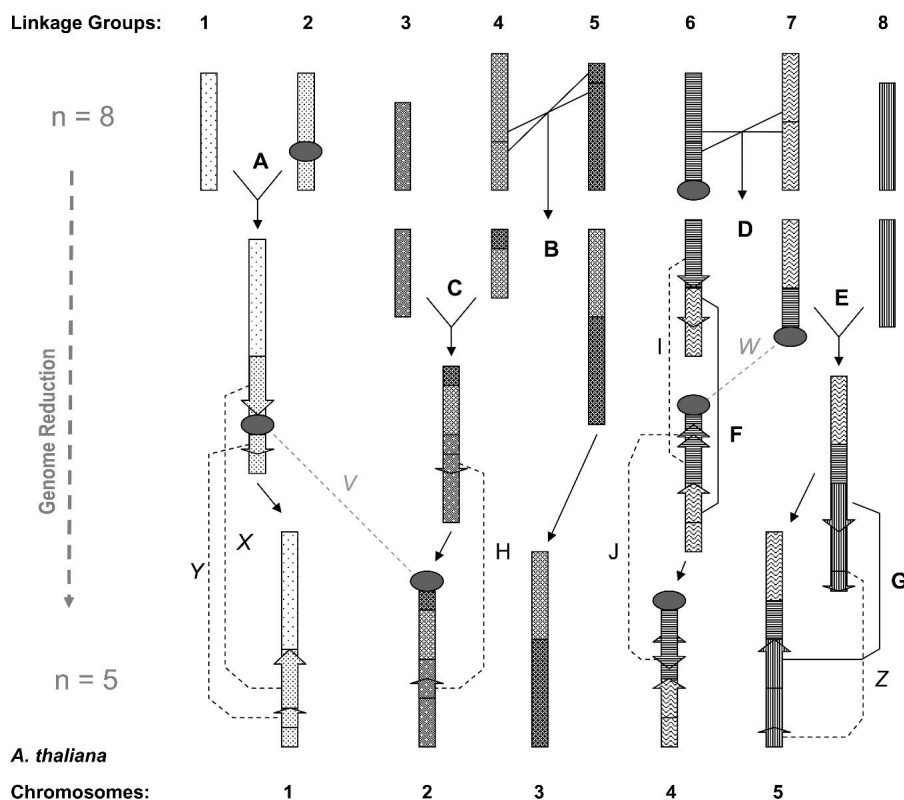


Figure 4. Genome reduction in *A. thaliana*: Parsimony solution of observable chromosomal rearrangement steps. Linkage groups 1–8 of *A. l. lyrata*, represented as patterned blocks, are shown across the top of the figure; *A. thaliana* chromosomes I–V at the bottom. A–G: Seven steps that are found to be in common between *A. thaliana*–*A. l. lyrata* and *A. thaliana*–*Capsella* spp. genome data; namely B and D, reciprocal translocations; A,C,E, fusion events; F and G, inversions. These seven steps could have occurred in any sequence during genome reduction in *A. thaliana*'s evolutionary history. (H–J) Inversions seen only in the *A. thaliana*–*A. l. lyrata* comparative map. (V–Z) Rearrangements seen only in the *A. thaliana*–*Capsella* spp. comparative map, namely X,Y, Z, inversions and V and W, translocations of NORs (V and W not included in statistical analysis).

A. l. petraea LG 1. Two of these regions are unresolved on the two other maps. The third region is unresolved in *Capsella*, but is defined by markers 1_05510, 1_08150, and 1_09530 on *A. l. lyrata* LG 1 where no inversion is observed. However, this discrepancy between these *A. lyrata* species-specific maps is unlikely to reflect a rearrangement that distinguishes these subspecies, since the marker order in these three regions of *A. l. petraea* is not certain (Kuittinen et al. 2004).

Gene families and duplications

The mapping of markers that correspond to members of gene families can often provide information on the evolution of these gene families. For example, in the *Capsella* mapping study

(Boivin et al. 2004) multiple (two or more) scorable polymorphisms were detected for such markers, allowing the simultaneous mapping of several paralogs to various locations in the *Capsella* genome. In some cases, deletion following duplication could be inferred due to absence of a corresponding copy in syntenic regions of either *Capsella* or *A. thaliana* (Boivin et al. 2004). In contrast, in our mapping study, markers specific for individual members of a gene family were often observed to identify only one polymorphism that corresponded to the expected location of one of the paralogs of that particular gene. Only two of the markers screened, 5_09650 and 3_58450, could clearly be scored for two polymorphisms, but only a single paralog was successfully mapped in each case. At5g09650 is a single-copy gene in *A. thaliana*, and an *A. lyrata* homolog was mapped to Aly LG 1 in our study, suggesting that it may correspond to a copy lost in *A. thaliana* but still present in *A. lyrata*. Marker 3_58450 corresponds to two *A. thaliana* genes, At3g58450 and At3g11930, but the position of neither of these genes in *A. thaliana* explains its placement in *A. lyrata* LG 4, suggesting a case of duplication followed by deletion of different paralogs in the two genomes.

Chromosomal evolution in the *A. thaliana* lineage

Three-way comparisons of the *A. thaliana* genome with *A. lyrata* and *Capsella* spp. can begin to address the

question of how reduction in chromosome number occurred in the *A. thaliana* lineage. Figure 5 illustrates a simplified phylogeny relating the three taxa *A. thaliana*, *A. lyrata*, and *Capsella* species. The most parsimonious explanation for the synteny observed between modern day *A. lyrata* and *Capsella* spp. is that genome arrangement has, to a large extent, been conserved since the last common ancestor of these taxa (Fig. 5A). Most importantly, the extent of colinearity demonstrated by the maps of these two taxa strongly suggests that major chromosomal rearrangement events, namely chromosomal fusions, reciprocal translocations, and some inversions, must have occurred in the evolutionary history of *A. thaliana* since its divergence from *A. lyrata* ~5 Mya (Fig. 5B). The presence of a few large gaps on the *A. l. lyrata* map despite extensive marker coverage of the contributing regions,

Table 3. Estimated rates of chromosomal evolution

Genomes under comparison	Estimated # of rearrangements (largest 95%CI)	Estimated divergence time (Myr)	Rate of chromosomal evolution (rearrangements/Myr/genome)
<i>A. thaliana</i> vs. <i>A. lyrata</i>	[10, 13]	3.8–5.8 ^a	–0.9–1.7
<i>A. thaliana</i> vs. <i>Capsella</i> spp.	[10, 12]	~10 ^b	–0.5–0.6

For calculation see Methods.

Within the Bayesian 95% CI estimate of the # of rearrangements, parsimony estimates are denoted in bold numbering.

^aAfter Kuittinen and Aguade (2000).

^bAfter Koch et al. (2000, 2001).

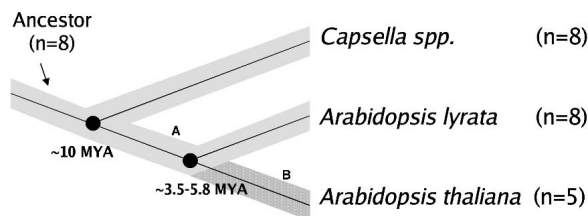


Figure 5. Phylogenetic relationship between *Arabidopsis* spp. and *Capsella* spp. The phylogenetic relationship between *A. thaliana*, *A. lyrata*, and *Capsella* spp. is illustrated in this simple cartoon (modified from Koch et al. 1999). Extensive colinearity of mapping data suggests (A) conservation of genome arrangement since the ancestor of the tribe Sisymbrieae ($n = 8$) until these modern-day taxa with haploid complement of $n = 8$, and (B) the occurrence of most major genome rearrangement events in the last 5 Myr of *A. thaliana*'s evolutionary history.

such as the gaps on Aly LG 3 and Aly LG 4 (Fig. 2), supports the notion that chromosomal loss may have been a player during genome reduction in the *A. thaliana* lineage, and suggests potential sites for chromosomal loss in this lineage. Such chromosomal loss would have presumably been tolerated due to genome-wide and partial duplications that predated the origin of the family Brassicaceae (The *Arabidopsis* Genome Initiative 2000; Blanc et al. 2000; Vision et al. 2000; Bowers et al. 2003; Ermolaeva et al. 2003; Raes et al. 2003).

The Bayesian analyses of the *A. thaliana*-*A. lyrata* (At-Aly) and *A. thaliana*-*Capsella* (At-Cap) mapping data allow us to infer the most likely numbers of rearrangements that distinguish these genomes and to deduce rates of chromosomal evolution. High posterior probabilities in favor of the most parsimonious scenarios of 10 observable rearrangement events, namely five inversions, two translocations, and three fusion/fissions separating these genomes (Fig. 3, Table 2) were estimated. Within the resolution of the mapping data, at least three alternative scenarios for observable rearrangement events were also predicted at low probability ($\leq 13\%$), but overall, there is strong evidence to suggest that the parsimony solution is a good approximation of what occurred during evolution. Thus, this statistical approach affords an opportunity to determine how likely the parsimony estimate is, to offer alternative probable scenarios, and therefore to arrive at a comprehensive view of the probable rearrangements that distinguish genomes. It should be noted, however, that this approach, though powerful, is constrained by the resolution of the mapping data, and additional comparative data will be required to infer which rearrangement events predated others during genome reduction in *A. thaliana* and when in the last ~5–10 Myr they occurred.

A detailed look at the paths with the highest posterior probability of occurring revealed the observable rearrangements that most likely occurred during evolution, as summarized in Figure 4. In these cases, the most likely sets of rearrangements are also parsimonious. A parsimonious view of the high degree of genome colinearity observed between *A. lyrata* and *Capsella* also implies that at least seven of the rearrangements (A–G) occurred in the *A. thaliana* lineage since the last common ancestor ($n = 8$) of *A. thaliana* and *A. lyrata* ~5 Mya. An additional three inversion events (H–J) are also predicted to have occurred sometime in the last 5 Myr, but these changes cannot be inferred to be specific to the *A. lyrata* or the *A. thaliana* lineage, because the affected regions are unresolved on the *Capsella* map. Similarly, events X–Z, which are observed only on the *Capsella* map, cannot be assumed

to be specific to the *Capsella* lineage, because they are unresolved on the *A. lyrata* map. It must be noted that two additional translocations involved in the repositioning of the NORs are likely to have occurred, based on the *Capsella* map (Boivin et al. 2004) and evidence from FISH-based studies in *A. thaliana* relatives (Lysak et al. 2003). However, these translocations were not accounted for in the Bayesian analyses, because raw position and distance data were not available (see Methods), but they are included in the scenario for karyotype evolution in Figure 4 (events V and W).

Clearly, a large number of rearrangements have occurred since the divergence of *A. thaliana* from its closest relative *A. lyrata*, based on our mapping data. According to the Bayesian estimates of the most likely number of rearrangement events, we infer a frequency of ~0.5–0.6 rearrangements/Myr/genome from the *Capsella*-*A. thaliana* comparison and a higher rate of ~0.9–1.7 rearrangements/Myr/genome from the *A. lyrata*-*A. thaliana* comparison (Table 3). These estimates suggest that the rate of chromosomal evolution almost doubled during the last ~5 Myr of *A. thaliana*'s evolutionary history. Although direct comparisons of these estimates to those derived from other comparative mapping studies cannot be made, the estimate of ~0.9–1.7 rearrangements/Myr/genome is comparable to the rate of chromosome evolution inferred from a comparison of the *B. nigra* and *A. thaliana* genomes. In a comparative mapping study performed prior to the release of the full *A. thaliana* genome sequence, Lagercrantz (1998) inferred that 90 chromosomal rearrangements differentiate the *A. thaliana* and *B. nigra* genomes, which translates to an estimate of 1.9–3.2 rearrangements/Myr/genome (Kuittinen et al. 2004) based on a divergence time between the two taxa of ~14–24 Myr (Yang et al. 1999; Koch et al. 2000, 2001). This large number of chromosomal rearrangements has suggested a more rapid rate of chromosomal evolution in the Brassicaceae than in other plant family genomes for which maps of comparable resolution are available (Lagercrantz 1998).

While any measure of the rate of chromosomal evolution is dependent on the density of the maps analyzed, the degree to which deviations from colinearity can be verified, and the accuracy of divergence time estimates, our results support the hypothesis of rapid chromosomal evolution in the Brassicaceae (Lagercrantz 1998). In contrast, Kuittinen et al. (2004) recognized only six major rearrangements in *A. lyrata* relative to *A. thaliana*, and they inferred a relatively slow rate of 0.6 rearrangements/Myr/genome. A Bayesian analysis of the *A. lyrata* mapping data of Kuittinen et al. (2004) resolves this discrepancy and predicts a rate of 0.8–1.3 rearrangements/Myr/genome (using the 3.8–5.8 Myr divergence estimate), which is comparable to *A. lyrata* despite the lack of resolution in some regions of major chromosomal rearrangement (data not shown).

The relatively rapid rates of chromosomal evolution inferred from our Bayesian analyses of two independent *A. lyrata*-*A. thaliana* comparative mapping studies support the hypothesis that rapid chromosomal evolution is a characteristic of lineages that differ by major changes in chromosome number and is not restricted to the polyploid lineages of *Brassica* spp. as previously suggested (Koch 2003; Kuittinen et al. 2004). This hypothesis is consistent with the high degree of colinearity observed by Lukens et al. (2003) between *B. oleracea* and the *Capsella* linkage group 4 (Acarkan et al. 2000), which counters the notion that extensive genome rearrangements occurred exclusively in the Brassica lineage. Indeed, Lukens et al. (2003) suggested that the last common ancestor of *Arabidopsis*, *Capsella*, and *Brassica* might have had a Brassica-like genome structure.

Genomic plasticity in the *Brassicaceae*, manifested by high levels of duplications and chromosomal rearrangements, has been suggested to have enhanced the ability of crucifer species to cope and adapt to change during alternating cycles of glacial and interglacial migrations through Eurasia in the last 10 Myr (McClintock 1984; Kianian and Quiros 1992; Kowalski et al. 1994). Glaciations may have facilitated the fixation of certain chromosomal rearrangements in *Brassica* spp. in small refugial populations while others died out, leading to modern day karyotypic diversity (Kowalski et al. 1994). *A. thaliana* is thought to have colonized Europe during the last glaciation (Pleistocene: 1.8 Myr to 11,000 years ago; Sharbel et al. 2000), and genome reduction may have resulted from such a major cataclysmic event. Fixation of perturbations in chromosomal structure may have been responsible for the speciation of *A. thaliana* as well as several *Arabidopsis* spp. ($n = 6-7$; Koch et al. 1999) that demonstrate karyotypic changes from the ancestral state of $n = 8$. Clearly, there is a need for further comparative mapping between members of the *Brassicaceae*, to identify larger colinear blocks, better predict the structure of the ancestral genomes, and help define the number and sequence of rearrangement events specific to the various lineages. Application of the statistical approaches described here, in conjunction with other recently developed methods such as chromosome painting of whole chromosomes with BAC-specific probes (Lysak et al. 2003), should allow a more rational and detailed reconstruction of karyotype evolution in the *Brassicaceae*.

Methods

Plant material and mapping population

The accession 97.19 of the self-incompatible *Arabidopsis lyrata* ssp. *lyrata* (*A. l. lyrata*) used in this study was collected from Northern Michigan (USA) and kindly provided by C. Langley (U.C. Davis). A single plant of this accession was considered an "F₁" plant (hereafter 97.19 in the text will refer to this F₁ plant) with sufficient heterozygosity to generate a mapping population, because of the outcrossing nature of *A. l. lyrata*. Thus 97.19 was selfed to generate an F₂ population of 67 individuals for mapping.

RFLP markers

EST markers spanning the BAC-tiling path of *A. thaliana* were used in BLASTN searches (Altschul et al. 1990) of the *A. thaliana* genome sequence to determine copy number. Clones of ESTs that appeared to be single copy (or tandem duplicates) were preferentially selected to minimize scoring ambiguities, and these clones were obtained from the *Arabidopsis* Biological Resource Center (ABRC) at Ohio State University. It must be noted that the earliest marker screen of ~150 markers (of which ~70 anchor markers were placed on the map) was done prior to the release of the full-genome sequence of *A. thaliana*. As a result, this marker set contained several markers belonging to gene families that had slipped through our screen for single-copy markers, and some of these markers were informative. Clones were grown overnight in Luria-Broth supplemented with appropriate antibiotics, and the cultures were used as a source of template for colony amplification using the polymerase chain reaction (PCR) under the following conditions: 99°C for 10 min to lyse cells, followed by 35 cycles, each of 94°C for 30 sec, 55°C for 30 sec, and 72°C for 1 min and 30 sec. Amplified PCR products were used as probes to detect restriction fragment length polymorphisms by DNA gel blot analysis as described below. In addition, several RFLP markers used to map the *Capsella* genome as described in Boivin et al. (2004) and a 28S rDNA-specific clone were also used.

DNA gel blot analysis

Genomic DNA was isolated from leaf and inflorescence tissue collected from plant 97.19 and individual plants in the mapping population using the CTAB protocol (Fulton et al. 1995). The DNA was digested with EcoRI, EcoRV, DraI, and ScaI, and the digests were subjected to electrophoresis on 0.8% (w/v) agarose gels followed by capillary transfer to Hybond N+ membranes under alkaline conditions (0.25N HCl depurination, 10 min, 0.4N NaOH for 15 min, and transfer in 0.4N NaOH). Probes were generated from PCR products using the Random Primed DNA labeling kit from Roche Applied Science and ³²P- α -dATP. The blots were hybridized at 65°C overnight, washed twice at low-stringency conditions (2% SSC / 0.1% SDS, 65°C, 20 min), and exposed to autoradiography film or phosphor screens.

Restriction fragment length polymorphisms were identified first by screening survey blots containing DNA digests from a few members of the F₂ population. Each of the probes that identified scorable polymorphisms was then used to genotype 50-67 F₂ plants. Several markers were discarded either because they were monomorphic and thus uninformative, or because they were multicopy and would likely lead to scoring errors.

SSR markers for Chromosome I

Previously reported polymorphic simple sequence-length repeats (SSRs) from *A. thaliana* Chromosome I were used to screen the *A. l. lyrata* mapping population. These SSRs included the dinucleotide repeats AtSRP54A, nF19G10, AtZFPG, T27K12, ciw1, nga280, AthGENEA, nF19K23, nF5114, AthATPase, nga692, and NYUP8H12 (Bell and Ecker 1994) and the trinucleotide repeat ATTS0392 (Clausen et al. 2002). For regions within the gap on chromosome I of *A. thaliana* that contained no previously described SSR markers, the SSRIT (simple sequence repeat identification tool: <http://www.gramene.org/db/searches/ssrtool>; Temnykh et al. 2001) was used to locate SSRs. PCR-based primers were designed to 100-base pair regions flanking identified SSRs in *A. thaliana* using the PrimerSelect program in the DNASTAR software package and synthesized by Integrated DNA Technologies. These were tested to determine their ability to amplify *A. l. lyrata*, and then used to screen the mapping population. PCR and gel electrophoresis conditions were as described by Lukowitz et al. (2000).

Genetic linkage analysis

In the absence of information regarding parental genotypes, each polymorphic marker was scored twice taking into account both parental possibilities, and both sets of data were pooled and analyzed using Mapmaker 3.0 for linkage group construction, with the LOD score (minimum logarithm of the odds ratio) set at 3.0. Positional information regarding the 104 mapped markers is provided in Supplemental Table S1.

Sequencing

Eight markers (two single copy, one tandem duplicate, and five gene family members) were found to map to regions in *A. l. lyrata* that were not colinear with the corresponding regions of *A. thaliana*, such that their locations could not be explained within the context of the inferred large-scale genome rearrangements. In these cases of interrupted synteny, the EST clones obtained from ABRC were sequenced to verify their identity. DNA sequencing was performed using the Applied Biosystems automated 3730xl DNA analyzer, big dye terminator chemistry, and AmpliTaq-FS DNA polymerase by the BioResource Center at Cornell University. Sequence identity was verified using a BLASTN query of the *A. thaliana* whole-genome sequence. In one case, sequence analysis revealed that the EST clone was mislabeled: It did not corre-

spond to the expected At3g18310 locus, but instead it matched the gene locus At5g42440 whose position in *A. thaliana* did correspond to a colinear block with *A. lyrata*. The correct identity of this marker is indicated on the map.

Segregation ratios

Genotypic data for plants in the mapping population were used to calculate segregation ratios for each of 97 markers with clear scoring data. χ^2 -tests were performed to determine which markers segregated according to the expected 1:2:1 ratio among genotypes and the expected 1:1 ratio between parental alleles, and which markers exhibited significant deviations from these ratios at a locus-by-locus *P*-value of 0.05.

Comparisons with maps of other species

For comparisons with maps of other related species, the correspondence of various linkage groups is as follows:

	<i>A. l. lyrata</i>	<i>A. l. petraea</i>	<i>Capsella spp.</i>
	1	1	A
	2	2	B
	3	4	D
Linkage Groups	4	3	C
	5	5	E
	6	7	G
	7	6	F
	8	8	H

Mapping file preparation and statistical analysis

Mapping files were generated which list marker order (with locus ID names) per chromosome in both genomes (1) without distances, (2) with distance in Mbp on the *A. thaliana* genome only, and (3) with distance in cM on the *A. l. lyrata* genome only. To prevent spurious inflation of the Bayesian estimates of 95% credible intervals (CIs) for rearrangements, eight of the 104 mapped markers were disregarded because they exhibited either lower-than-threshold LOD scores or mapping positions more consistent with duplication followed by deletion than with genome-level rearrangements. Eighteen loci on the Aly map were defined by two or more cosegregating markers. In these cases, only one marker was retained to define the locus on the mapping files, as the local order of cosegregating markers could not be ascertained on the Aly map.

Mapping file data for the remaining 78 markers were subjected to Bayesian analyses using Markov chain Monte Carlo (MCMC) code, based on theory and methodology described by Durrett et al. (2004). Briefly, in this method, one makes a reasonable assumption about the prior probability of occurrence of each possible path (sequence of marker arrangements); this is multiplied by the probability of the data given the path (in this case either 1 or 0) to give the relative posterior probability. These probabilities are used in constructing a Markov chain which, when run for sufficiently long, provides samples from the posterior distribution. These samples are then analyzed to yield, for instance, the posterior distribution of the number of inversions, and the joint posterior distribution of inversions and translocations. The original published form of the code uses information about marker order only. The capability of using distance information from one of the genomes (but not both) has recently been added to this code (T.L. York, pers. comm., in prep.) and the program was used in this capacity to perform the analyses. The analysis provides 95% CIs that define the most likely range of possible rearrangements that best fit the data. These intervals are

denoted in square brackets and collectively correspond to 95% of the posterior probability (i.e., $L = [3, 4]$ denotes a 95% posterior probability of 3–4 events having occurred). The analyses provided individual 95% CIs for the numbers of observable rearrangements, namely, inversions (Li, defined by at least two markers), translocations (Lt), fusions/fissions (Lf), as well as the total number of rearrangements (L) that are required to convert one genome into another.

A similar comparative statistical analysis of *A. thaliana* and *Capsella* genome data was also performed with mapping files generated from the supplemental data of Boivin et al. (2004). A BLASTN search with six gene-specific markers used in that study against the *A. thaliana* genome revealed the paralogs whose positions in *A. thaliana* are syntenic with their mapped location in the *Capsella* genome [illustrated in Fig. 2 of Boivin et al. 2004 but not documented in their supplemental data]. These paralogous loci (specifically, E71b = At1g50010, E13 = At1g57820, E71a = At1g64740, mi320 = At1g66530, E26 = At3g53870, CA6 = At5g66450) were included in the raw mapping file. Of the 117 markers on the *Capsella* map, 95 markers defining distinct loci (i.e. allowing only one marker to define each locus where cosegregating markers mapped) were used to generate the mapping file for statistical analyses. Raw position and distance data for the rDNA marker used to locate the NORs was not provided in the supplemental data of Boivin et al. (2004), and hence could not be included in the mapping file. Only two markers were found to be mapped on both *Capsella* and *A. l. lyrata*; all others were unique to each map.

Estimation of rates of chromosomal evolution

The equation used by Kuitinen et al. (2004) to calculate the number of rearrangements/Myr/genome was used to estimate the rate of chromosomal evolution (Table 3), taking the smallest lower bound and the largest upper bound of the 95% CIs, as defined by three permutations of the Bayesian analysis, to be the number of rearrangement events for each data set (i.e., [10, 13] for At-Aly, see Table 1). In cases where the divergence time was defined by a range (i.e., 3.8–5.8 Myr for At-Aly), four values were obtained, and the smallest and largest of these values were used to define the estimate of the rate of chromosomal evolution (i.e., the values 1.3, 0.9, 1.7 and 1.1 for At-Aly give an estimate of 0.9–1.7).

Acknowledgments

We thank Charles Langley (Univ. of California, Davis, CA) for kindly providing *A. lyrata* seed, Susan McCouch and Teresa Fulton (Cornell Univ., Ithaca, NY) for helpful suggestions during the course of the mapping project, Renate Schmidt (Max-Planck-Inst., Potsdam, Germany) and Outi Savolainen (Univ. of Oulu, Finland) for kindly providing manuscripts prior to publication, Patrick J. Brown and Sue Sherman-Broyles (Cornell Univ.) for critical reading of the manuscript, and the *Arabidopsis* Biological Resource Center for providing EST stocks. This work was supported by a grant from the NSF to M.E.N. T.L.Y. was supported by Grant # 0201037 from a joint program of the NSF and National Inst. of General Medical Sciences.

References

- Acaran, A., Rossberg, M., Koch, M., and Schmidt, R. 2000. Comparative genome analysis reveals extensive conservation for *Arabidopsis thaliana* and *Capsella rubella*. *Plant J.* **23**: 55–62.

- Ahn, S. and Tanksley, S.D. 1993. Comparative linkage maps of rice and maize genomes. *Proc. Natl. Acad. Sci.* **90**: 7980–7984.
- Al-Shehbaz, I.A. 1973. The biosystematics of the genus *Thelypodium* (*Cruciferae*). *Contributions to the Gray Herbarium, Harvard University* **204**: 3–148.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and D. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**: 403–410.
- The *Arabidopsis* Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**: 796–815.
- Beavis, W.D. and Grant, D. 1991. A linkage map based on information from F2 populations of maize (*Zea mays*). *Theor. Appl. Genet.* **82**: 636–644.
- Bell, C.J. and Ecker, J.R. 1994. Assignment of 30 microsatellite loci to the linkage map of *Arabidopsis*. *Genomics* **19**: 137–144.
- Blanc, G., Barakat, A., Guyot, R., Cooke, R., and Delseny, M. 2000. Extensive duplication and reshuffling in the *Arabidopsis* genome. *Plant Cell* **12**: 1093–1101.
- Boivin, K., Acarkan, A., Mbulu, R.-S., Clarenz, O., and Schmidt, R. 2004. The *Arabidopsis* genome sequence as a tool for genome analysis in the *Brassicaceae*. A comparison of the *Arabidopsis* and *Capsella rubella* genomes. *Plant Physiol.* **135**: 735–744.
- Bonierbale, M.D., Plaisted, R.L., and Tanksley, S.D. 1988. RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato. *Genetics* **120**: 1095–1103.
- Bowers, J.E., Chapman, B.A., Rong, J., and Paterson, A.H. 2003. Unraveling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* **422**: 433–438.
- Burr, B., Burr, A., Thompson, K.H., Albertson, M.C., and Stuber, C.W. 1988. Gene mapping with recombinant inbreds in maize. *Genetics* **118**: 519–526.
- Chittendan, L.M., Schertz, K.F., Lin, Y.-R., Wing, R.A., and Paterson, A.H. 1994. A detailed RFLP map of *Sorghum bicolor* X *S. propinquum* suitable for high-density mapping suggests ancestral duplication of chromosomes or chromosomal segments. *Theor. Appl. Genet.* **87**: 925–933.
- Clauss, M.J., Cobban, H., and Mitchell-Olds, T. 2002. Cross-species microsatellite markers for elucidating population genetic structure in *Arabidopsis* and *Arabis* (*Brassicaceae*). *Mol. Ecol.* **11**: 591–601.
- Copenhaver, G.P. and Pikaard, C.S. 1996. RFLP and physical mapping with an rDNA-specific endonuclease reveals that nucleolus organizer regions of *Arabidopsis thaliana* adjoin the telomeres on chromosome 2 and 4. *Plant J.* **9**: 259–272.
- Dart, S., Kron, P., and Mable, B.K. 2004. Characterizing polyploidy in *Arabidopsis lyrata* using chromosome counts and flow cytometry. *Can. J. Bot.* **82**: 185–197.
- de Vienne, D. 2003. Construction of genetic linkage maps. In *Molecular markers in plant genetics and biotechnology* (ed. D. de Vienne), pp. 47–79. Science Publishers, Enfield, NH.
- Doganlar, S., Frary, A., Daunay, M.-C., Lester, R.N., and Tanksley, S.D. 2002. A comparative genetic linkage map of eggplant (*Solanum melongena*) and its implications for genome evolution in the *Solanaceae*. *Genetics* **161**: 1697–1711.
- Durrett, R., Nielsen, R., and York, T.L. 2004. Bayesian estimation of genomic distance. *Genetics* **166**: 621–629.
- Ermolaeva, M.D., Wu, M., Eisen, J.A., and Salzberg, S.L. 2003. The age of *Arabidopsis thaliana* genome duplication. *Plant Mol. Biol.* **51**: 859–866.
- Fulton, T.M., Chunwongse, J., and Tanksley, S.D. 1995. Microprep protocol for extraction of DNA from tomato and other herbaceous plants. *Plant Mol. Biol. Report.* **13**: 207–209.
- Gale, M. and Devos, K.M. 1998. Plant comparative genetics after 10 years. *Science* **282**: 656–659.
- Hall, A.E., Fiebig, A., and Preuss, D. 2002. Beyond the *Arabidopsis* genome: Opportunities for comparative genomics. *Plant Physiol.* **129**: 1439–1447.
- Kianian, S.F. and Quiros, C.F. 1992. Generation of a *Brassica oleracea* composite RFLP map: Linkage arrangements among various populations and evolutionary implications. *Theor. Appl. Genet.* **84**: 544–554.
- Koch, M. 2003. Molecular phylogenetics, evolution and population biology in the *Brassicaceae*. In *Plant Genome biodiversity and evolution* (eds. A.K. Sharma and A. Sharma), pp. 1–35. Science Publishers, Enfield, NH.
- Koch, M., Bishop, J., and Mitchell-Olds, T. 1999. Molecular systematics and evolution *Arabidopsis* and *Arabis*. *Plant Biol.* **1**: 529–537.
- Koch, M., Haubold, B., and Mitchell-Olds, T. 2000. Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis*, *Arabis* and related genera (*Brassicaceae*). *Mol. Biol. Evol.* **17**: 1483–1498.
- Koch, M., Haubold, B., and Mitchell-Olds, T. 2001. Molecular systematics of the *Brassicaceae*: Evidence from coding plastidic *matK* and nuclear *Chs* sequences. *Am. J. Bot.* **88**: 534–544.
- Kowalski, S.P., Lan, T.-H., Feldmann, K.A., and Paterson, A.H. 1994. Comparative mapping of *Arabidopsis thaliana* and *Brassica oleracea* chromosomes reveals islands of conserved organization. *Genetics* **138**: 499–510.
- Kuittinen, H. and Aguade, M. 2000. Nucleotide variation at the *CHALCONE ISOMERASE* locus in *Arabidopsis thaliana*. *Genetics* **155**: 863–872.
- Kuittinen, H., de Haan, A.A., Vogl, C., Oikarinen, S., Leppala, J., Koch, M., Mitchell-Olds, T., Langley, C., and Savolainen, O. 2004. Comparing the linkage maps of the close relatives *Arabidopsis lyrata* and *Arabidopsis thaliana*. *Genetics* **168**: 1575–1584.
- Kurata, N., Moore, G., Nagamura, Y., Foote, T., Yano, M., Minobe, Y., and Gale, M. 1994. Conservation of genome structure between rice and wheat. *Biotechnol.* **12**: 276–278.
- Lagercrantz, U. 1998. Comparative mapping between *Arabidopsis thaliana* and *Brassica nigra* indicates that *Brassica* genomes have evolved through extensive genome replication accompanied by chromosome fusions and frequent rearrangements. *Genetics* **150**: 1217–1228.
- Lagercrantz, U. and Lydiate, D.A. 1996. Comparative genome mapping in *Brassica*. *Genetics* **144**: 1903–1910.
- Lander, E.S., Green, P., Abrahamson, J., Barlow, A., Daly, M.J., Lincoln, S.E., and Newburg, L. 1987. Mapmaker: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* **1**: 174–181.
- Lukens, L., Zou, F., Lydiate, D.A., Parkin, I., and Osborn, T. 2003. Comparison of a *Brassica oleracea* genetic map with the genome of *Arabidopsis thaliana*. *Genetics* **164**: 359–372.
- Lukowitz, W., Gillmor, S., and Scheible, W.-R. 2000. Positional cloning in *Arabidopsis*. Why it feels so good to have a genome initiative working for you. *Plant Physiol.* **123**: 795–805.
- Lysak, M.A., Pecinka, A., and Schubert, I. 2003. Recent progress in chromosome painting of *Arabidopsis* and related species. *Chromosome Res.* **11**: 195–204.
- McClintock, B. 1984. The significance of responses of the genome to challenge. *Science* **226**: 792–801.
- Nasrallah, M.E., Yogeewaran, K., Snyder, S., and Nasrallah, J.B. 2000. *Arabidopsis* species hybrids in the study of species differences and evolution of amphiploidy in plants. *Plant Physiol.* **124**: 1605–1614.
- Paterson, A.H. and Bennetzen, J.L. 2001. Comparative mapping of plant chromosomes. In *DNA-based markers in plants* (eds. R.L. Phillips and I.K. Vasil), pp. 101–114. Kulwer Academic Press, Boston.
- Periera, M.G., Lee, M., Bramel-Cox, P., Woodman, W., Doebley, J., and Whitkus, R. 1994. Construction of an RFLP map in sorghum and comparative mapping in maize. *Genome* **37**: 236–243.
- Prince, J.P., Pochard, E., and Tanksley, S.D. 1993. Construction of a molecular linkage map of pepper, and a comparison of synteny with tomato. *Genome* **36**: 404–417.
- Raes, J., Vandepoel, K., Simillion, C., Saeys, Y., and Van de Peer, Y. 2003. Investigating ancient duplication events in the *Arabidopsis* genome. *J. Struct. Funct. Genomics* **3**: 117–129.
- Ramos-Onsins, S.E., Stranger, B.E., Mitchell-Olds, T., and Aguade, M. 2004. Multilocus analysis of variation and speciation in the closely related species *Arabidopsis halleri* and *A. lyrata*. *Genetics* **166**: 373–388.
- Rieseberg, L.H. 2001. Chromosomal rearrangements and speciation. *Trends Ecol. Evol.* **16**: 351–358.
- Schmuths, H., Meister, A., Horres, R., and Bachmann, K. 2004. Genome size variation among accessions of *Arabidopsis thaliana*. *Ann. Bot.* **93**: 317–321.
- Sharbel, T.F., Haubold, B., and Mitchell-Olds, T. 2000. Genetic isolation by distance in *Arabidopsis thaliana*: Biogeography and postglacial colonization of Europe. *Mol. Ecol.* **9**: 2109–2118.
- Temnykh, S., DeClerk, G., Lukashova, A., Lipovich, L., Cartinhour, S., and McCouch, S. 2001. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): Frequency, length, variation, transposon associations, and genetic marker potential. *Genome Res.* **11**: 1441–1452.
- van Treuren, R., Kuitteinen, H., Karkkainen, K., Baena-Gonzalez, E., and Savolainen, O. 1997. Evolution of microsatellites in *Arabis petraea* and *Arabis lyrata*, outcrossing relatives of *Arabidopsis thaliana*. *Mol. Biol. Evol.* **14**: 220–229.
- Vision, T.J., Brown, D.G., and Tanksley, S.D. 2000. The origins of genome duplications in *Arabidopsis*. *Science* **290**: 2114–2117.
- Yang, Y.-W., Lai, K.-N., Tai, P.-Y., and Li, W.-H. 1999. Rates of nucleotide substitution in angiosperm mitochondrial DNA sequences and dates of divergence between *Brassica* and other angiosperm lineages. *J. Mol. Evol.* **48**: 597–604.

Received November 5, 2005; accepted in revised form February 14, 2005.