DOI: 10.1534/genetics.106.063628

Duplication of Centromeric Histone H3 (*HTR12*) Gene in *Arabidopsis halleri* and *A. lyrata*, Plant Species With Multiple Centromeric Satellite Sequences

Akira Kawabe,*,[1] Shuhei Nasuda[†] and Deborah Charlesworth*

**Institute of Evolutionary Biology, Edinburgh University, Edinburgh EH9 3JT, United Kingdom and* [†]*Laboratory of Plant Genetics, Graduate School of Agriculture, Kyoto University, Sakyo-ku, Kyoto 606-8502, Japan*

Manuscript received July 18, 2006
Accepted for publication September 13, 2006

ABSTRACT

Arabidopsis halleri and *lyrata* have three different major centromeric satellite sequences, a unique finding for a diploid Arabidopsis species. Since centromeric histones coevolve with centromeric satellites, these proteins would be predicted to show signs of selection when new centromere satellites have recently arisen. We isolated centromeric protein genes from *A. halleri* and *lyrata* and found that one of them, *HTR12* (CENP-A), is duplicated, while CENP-C is not. Phylogenetic analysis indicates that the *HTR12* duplication occurred after these species diverged from *A. thaliana*. Genetic mapping shows that *HTR12* copy B has the same genomic location as the *A. thaliana* gene; the other copy (A, at the other end of the same chromosome) is probably the new copy. To test for selection since the duplication, we surveyed diversity at both *HTR12* loci within *A. lyrata*. Overall, there is no strong evidence for an "evolutionary arms race" causing multiple replacement substitutions. The *A. lyrata HTR12B* sequences fall into three classes of haplotypes, apparently maintained for a long time, but they all encode the same amino acid sequence. In contrast, *HTR12A* has low diversity, but many variants are amino acid replacements, possibly due to independent selective sweeps within populations of the species.

IT is thought that variation in the sequence-binding specificity of centromere-specific histone H3-like proteins and evolutionary changes in centromere sequences are causally related. When a centromere sequence mutates to a sequence with a competitive advantage over the existing one, conferring preferential segregation (Pardo-Manuel de Villena and Sapienza 2001), this could induce a selection pressure on the genes encoding binding proteins to resist preferential segregation (Henikoff *et al.* 2000, 2001; Malik and Henikoff 2002). Centromere evolution in Drosophila seems to involve adaptive evolution of one of the centromere-specific histone H3-like proteins, *Cid*, the Drosophila CENP-A (Henikoff *et al.* 2001; Malik and Henikoff 2001; Malik *et al.* 2002). An excess of amino-acid-changing substitutions is observed in the *Cid* sequence between different Drosophila species, suggesting adaptive evolution, consistent with these changes being driven by differences in centromeric satellite sequences between species (Malik and Henikoff 2001; Malik *et al.* 2002). The "arms race" theory of centromere evolution is, however, still controversial (*e.g.*, Sullivan 2002). First, it is not yet clear whether specific DNA sequences are essential for centromere functions. Second, CENP-A proteins of different species can sometimes replace one another (in transgenic experiments), proving that they are not specialized to function only with their own satellite sequences.

Recent work in the genus Arabidopsis has found evidence of adaptive evolution in the centromere-specific histone genes of these species also (Talbert *et al.* 2002). In *Arabidopsis thaliana*, the single-copy *HTR12* gene (At1g01370) is thought to encode the protein with the equivalent centromere-binding function and to represent the CENP-A gene of this species. The *A. thaliana HTR12* gene has nine exons and encodes a protein of 178 amino acids, which localizes to centromere regions (Talbert *et al.* 2002) and binds the *A. thaliana* centromeric satellite sequence (Nagaki *et al.* 2003). Like the centromeric histone H3 proteins of other organisms, the HTR12 protein has two domains, an N-terminal tail and a histone core domain. The histone core domain, which forms histone octamer nucleosomes, has ~60% amino acid identity to *A. thaliana* histone H3 proteins. *A. thaliana* and *A. arenosa* each have a single major satellite sequence (with some minor variants), which is highly repeated at the centromeres of all the chromosomes (Kamm *et al.* 1995); in *A. thaliana*, this is called pAL1, and in *A. arenosa* (an autotetraploid species), the centromere sequences differ and are called pAa (see Figure 1). Studies of *HTR12*

[1]*Corresponding author:* Institute of Evolutionary Biology, University of Edinburgh, Ashworth Laboratories, King's Bldgs., West Mains Rd., Edinburgh EH9 3JT, United Kingdom. E-mail: akira.kawabe@ed.ac.uk

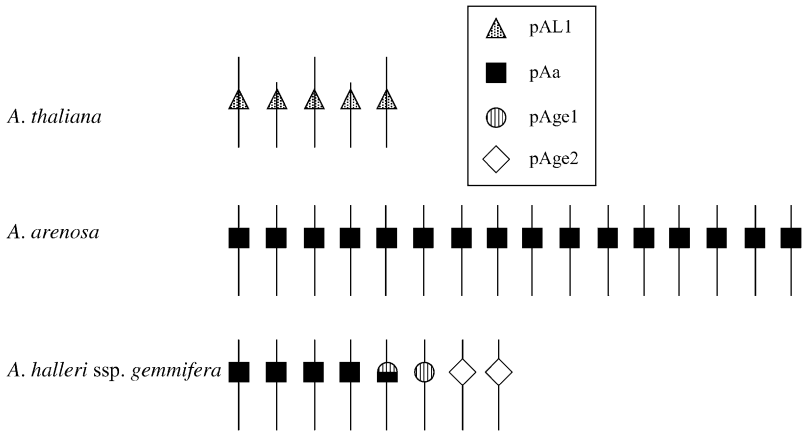

Figure 1.—Schematic of the centromeric satellite sequence families of *A. thaliana*, *A. arenosa*, and *A. halleri* ssp. *gemmifera* (from Kamm *et al.* 1995 and Kawabe and Nasuda 2005). The centromere satellite families present on each individual chromosome have not yet been identified in *A. halleri*.

suggest adaptive evolution between *A. thaliana* and *A. arenosa* (Talbert *et al.* 2002), and a recent study of seven more Arabidopsis species (including three tetraploids) inferred that two regions, the N-terminal tail and loop I of the histone core domain, show evidence of adaptive evolution (Cooper and Henikoff 2004).

In addition to CENP-A genes, CENP-C, another non-sequence-specific DNA-binding protein, has been reported to evolve adaptively. This has been inferred especially in mammals and grasses, whose CENP-A genes show no signatures of adaptive evolution (Talbert *et al.* 2004). In contrast, in Drosophila species, there is evidence of positive selection on CENP-A (*Cid*; Henikoff *et al.* 2001; Malik and Henikoff 2001; Malik *et al.* 2002), but not on CENP-C (Heeger *et al.* 2005). Comparisons between distant Drosophila species indicated adaptive evolution in CENP-C (Heeger *et al.* 2005; Richards *et al.* 2005), suggesting that different species may have experienced adaptive evolution in different centromere-binding protein genes. Comparisons of sequences of regions of the CENP-C from Arabidopsis species identified regions that appear to have undergone adaptive evolution on the basis of high K_a/K_s ratios between *A. thaliana* and *A. arenosa* (Talbert *et al.* 2004). These, however, are mainly caused by low synonymous divergence values (the same is true for the subregion studied here, see below). Furthermore, since it is based on comparison of sequences from only two species, it was not determined in which lineages positive selection occurred, since comparisons including additional species are needed to allow determination of the lineages in which substitutions occurred.

In *A. halleri* ssp. *gemmifera*, one of the closest relatives of *A. thaliana* (Miyashita *et al.* 1998; Kawabe and Miyashita 2002a), there are three major centromeric satellite families with different sequences in different chromosomes: the pAa satellite found in *A. arenosa*, plus two different sequences, pAge1 and -2 (Kawabe and Nasuda 2005; see also Figure 1). The other *A. halleri* subspecies (ssp. *halleri*) and the closely related *A. lyrata* (both subspecies *lyrata* and *petraea*) also have the same three major centromeric satellite families (A. Kawabe, unpublished observations).

If competition for binding to centromeric proteins with different DNA-binding specificity (or competition in relation to some other centromere function) caused the rapid change in centromere satellite sequences across all the chromosomes since the divergence from *A. thaliana*, we might expect these species to respond to the selection pressure outlined above, which will exist as long as multiple satellite DNA sequences are present at centromeres. The existence of multiple major centromere satellite sequences in *A. halleri* and *A. lyrata* suggests that these species may be in a transient stage, undergoing the proposed competitive centromere protein evolution, and it is thus interesting to study centromere-specific proteins (CENP-A and CENP-C) in these species to test whether they also vary, *i.e.*, whether there are different alleles at these loci that might interact differently with the different satellite DNA families, and whether there is evidence of recently arisen new alleles, suggesting recent selective sweeps caused by the spread of advantageous alleles. Here, we describe results from the CENP-A (*HTR12*) and CENP-C genes of species of plants related to *A. thaliana*.

The CENP-C gene is single copy, as in the related species, but the CENP-A gene is not just polymorphic with different *HTR12* alleles, but the *A. halleri* and *A. lyrata* genomes have more than one gene copy encoding CENP-A protein, *i.e.*, paralogous genes, and one of them is highly polymorphic. Among species of plants related to *A. thaliana*, duplication of centromeric histone H3 genes has previously been reported only in polyploids (Cooper and Henikoff 2004). It therefore seems significant that these species with clear signs of centromere satellite sequence evolution also have multiple centromeric histone H3 genes. We describe the evidence that the new copies arose in the lineage leading to *A. halleri* and *A. lyrata* and that both are functional genes, findings that are consistent with the new CENP-A sequence having evolved because it was able to associate with a new centromeric satellite sequence in these

species. The coexistence of both multiple centromere satellite families and centromeric histone H3 genes (*HTR12* in Arabidopsis) in *A. lyrata* and *A. halleri* suggests possible multiple satellite–protein interactions in the "arms race" of centromere evolution. A major purpose of this study therefore was to further test this, by testing for signatures of selection in the derived *HTR12* gene copy of *A. halleri* and *A. lyrata*, using both divergence between the species and diversity within them. Surprisingly, we failed to detect any evidence for fixation of advantage mutations in either locus in these species, or in their CENP-C.

MATERIALS AND METHODS

Plant materials and DNA isolation: To investigate the *HTR12* gene sequences of the study species and to estimate divergence between them, we used the following individual plants from the following species: an *A. halleri* ssp. *gemmifera* plant from Ohtani (Kasumi, Hyogo, Japan), an *A. halleri* ssp. *halleri* individual (AH-P1-13, from Pontresina, Switzerland), an *A. lyrata* ssp. *lyrata* individual (Ontario4, from Ontario, Canada, given to us by. B. K. Mable, University of Glasgow), an *A. lyrata* ssp. *petraea* plant (99R11-2, from Mount Esja, Iceland), and an *A. glabra* plant from Ohmi-Shirahama (Takashima, Shiga, Japan). The sample used for diversity estimates is described below. Total DNA was isolated from dried leaves by a modified CTAB method or using a FastDNA kit (Q-BIOgene) according to the manufacturer's instructions.

Isolation of centromeric histone H3 (*HTR12*) and CENP-C genes: Primers were designed on the basis of the sequence of the *HTR12* (At1g01370) gene of the *A. thaliana* strain *Col*-0. The primers ATC1672+ (5′-TAA AAA TCA ATG GCG AGA AC-3′) and ATC3536− (5′-CGA AAA GCA GAT AGA AAC AC-3′) were designed for the 5′ region, including the initiation codon, and the 3′ flanking region of the At1g01370 sequence, respectively. PCR reactions yielded two different sequences from genomic DNAs of *A. halleri* ssp. *gemmifera*. As described later, these two sequences appear to represent two loci, denoted in what follows by *HTR12A* and *HTR12B*. Primers specific for each putative locus were then designed (a 5′ primer ATC1672+ for both loci, and 3′ primers for loci A and B: HTR12a−: 5′-ATT CCG CTT TCC AGT TAT GTT T-3′ and HTR12b−: 5′-GGA TCC TAG ATA TTG TTA ACT ATT C-3′). These primers yielded two sequences from *A. halleri* ssp. *halleri*, *A. lyrata* ssp. *lyrata*, and *A. lyrata* ssp. *petraea* and from whole young shoot cDNA of *A. halleri* ssp. *gemmifera*. The exon–intron junctions were determined by comparing the *HTR12* sequences from cDNA and genomic DNA. For CENP-C amplification, two primers designed in exon 6 (5′-AAA AGG AAA AGA GGT AGA TGT GC-3′) and exon 11 (5′-ATG CCG ATA ACA GTA GTC AAA C-3′) were used. The sequences were deposited in the DDBJ and GenBank databases under accession nos. AB081500–AB081505, DQ450543–DQ450605, and DQ987606–DQ987610.

Genotyping of the mapping family: A set of 99 F_2 mapping family plants (Kuittinen *et al.* 2004; Hansson *et al.* 2006) was used to determine the chromosomal locations of the duplicated loci in *A. lyrata* ssp. *petraea*. Four F_2 plants were chosen to determine partial sequences of the two *HTR12* genes, using primers HTR1229+ (5′-GCC CCT CCC CAA ATC AAT C-3′) and either HTR12a− or HTR12b− (see above). Polymorphic restriction enzyme recognition sites found in the sequenced regions among the F_2 plants were then analyzed by PCR–RFLP (*Hin*fI was used for scoring the *HTR12A* genotypes, and *Hae*III for *HTR12B*).

TABLE 1

Sources and numbers of *A. lyrata* plants used in the diversity study

Population	Country	No. of plants: Sequence and PCR–RFLP	No. of plants: PCR–RFLP only
	Subspecies *petraea*		
Mount Esja	Iceland	7	18
Reykjanes	Iceland	0	2
Stubbsund	Sweden	4	3
Plech	Germany	4	4
Karhumaki	Russia	4	5
Spiterstulen	Norway	0	8
Lom	Norway	0	8
Clogwyn D'ur Arddu	Wales	2	2
Glaslyn	Wales	0	8
	Subspecies *lyrata*		
Ontario	Canada	4	0
Indiana	United States	0	4

Sequence diversity analysis: A total of 21 plants from five populations of *A. lyrata* ssp. *petraea* and 4 plants from a single population of *A. lyrata* ssp. *lyrata* were sampled and used to survey DNA polymorphism in the *HTR12* loci (Table 1). For these plants, sequences of at least three clones were determined for each sample and consensus sequences were used in the analyses below. If one clone differed from the others, more clones were sequenced to obtain both alleles of each plant. In total, 30 *HTR12A* and 30 *HTR12B* sequences were obtained from *A. lyrata*. Coding regions were assigned using cDNA information from *A. halleri* ssp. *gemmifera* obtained in this study and *A. thaliana* (Talbert *et al.* 2002).

Sequence analyses: The sequences were aligned manually. The DnaSP version 3.5 (Rozas and Rozas 1999) and MEGA2 (Kumar *et al.* 2000) programs were used to analyze divergence and diversity (using nucleotide diversity, π, as the estimate of intraspecific polymorphism) and population differentiation, to test linkage disequilibrium (LD), and to apply several tests for neutrality (see below), including McDonald–Kreitman tests (McDonald and Kreitman 1991). The other tests for neutrality were those of HKA (Hudson *et al.* 1987), Tajima's D (Tajima 1989), Fu and Li's D (Fu and Li 1993), Fu's F_s (Fu 1997), and Fay and Wu's H (Fay and Wu 2000). Wall's B and Q (Wall 1999), and Kelly's ZnS (Kelly 1997) were used to test for LD, using only parsimony informative sites. Recombination rates were estimated as R per site (R/site) and as R_m, the minimum number of recombination events, using DNAsp.

Trees were constructed from the sequences by the neighbor-joining method with Jukes–Cantor distances, using the MEGA2 program (Kumar *et al.* 2000). Synonymous and replacement divergence values were estimated by Nei and Gojobori's method with Jukes–Cantor correction, using MEGA2. The directions of mutations in *A. halleri* and *lyrata* were determined by parsimony with the *A. thaliana* sequence as the outgroup. When the nucleotides in both species differed from that in *A. thaliana*, or when the site was in a sequence gap in the *A. thaliana* sequence, the site was excluded from this analysis. For the CENP-C gene, mutations were assigned to individual branches by parsimony and then D_N, D_S, and D_{intron} values

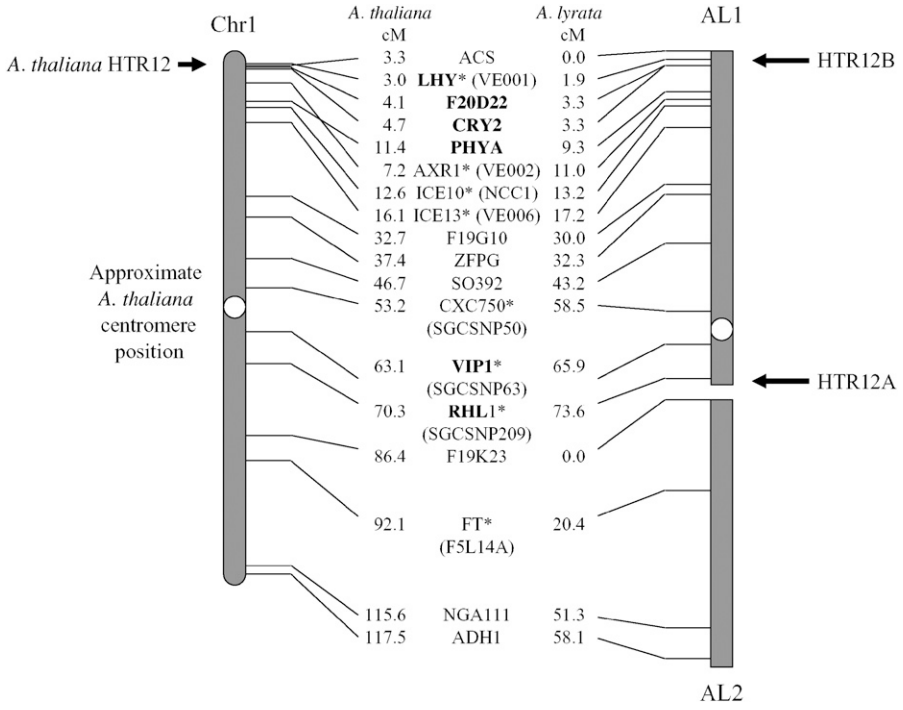

FIGURE 2.—Genetic maps of *A. thaliana* chromosome 1 and the homologous *A. lyrata* linkage groups 1 and 2 (AL1 and -2), showing the map positions of the *A. thaliana HTR12* gene and the duplicated *HTR12* loci in *A. lyrata* ssp. *petraea*. The names of markers showing close linkage to the duplicated *HTR12* loci in *A. lyrata* are in boldface type. Map positions of markers in the two species are centimorgan values in the *A. thaliana* recombinant inbred (RI) genetic map in the The Arabidopsis Information Resource database (http://www.arabidopsis.org/) and the *A. lyrata* map of KUITTINEN *et al.* (2004). Seven markers in the *A. lyrata* map (indicated by asterisks after the marker names) correspond to markers that were not mapped in the *A. thaliana* RI genetic map. The homologous *A. thaliana* markers were assumed to be located close to the positions of mapped markers that are <100 kb away in the complete genome sequence. These marker names are shown in parentheses. More map details are described in HANSSON *et al.* (2006).

were calculated for each branch by dividing the number of mutations by the numbers of sites using Jukes–Cantor correction.

PCR–RFLP: Three sequence types were found at the *HTR12B* locus (see below). To estimate the haplotype frequencies in samples from natural populations, we used PCR–RFLP. Two HTR12B primers (HTR1229+: 5′-GCC CCT CCC CAA ATC AAT C-3′ and HTR12b−) were used to obtain a PCR product of ~550 bp, which was digested by either *Eco*RV or *Pst*I and separated on 1.5% agarose gels. The *Pst*I recognition site is specific to sequence type 1 and the *Eco*RV recognition site is present in both types 1 and 2. Thus, the three sequence types can be distinguished.

RESULTS

Duplicated *HTR12* genes: Unlike other diploid animals and plants, two apparently distinct *HTR12* loci, *HTR12A* and *HTR12B*, were found in the genomes of all four *A. halleri* and *lyrata* subspecies, and in cDNA. The CENP-C gene is single copy, as in other related species, and will be discussed below. The exon–intron structures of the *HTR12A* and *HTR12B* genes of *A. lyrata* are the same as for the previously reported Arabidopsis *HTR12* genes (TALBERT *et al.* 2002). In both *A. halleri* subspecies, however, the *HTR12A* gene has a 16-bp deletion relative to either the *A. halleri* ssp. *gemmifera HTR12B* or the *A. thaliana HTR12* sequence. This deletion is located near the 3′-end of intron 1 and includes the putative 3′ AG splice site. There is a new functional splice site 22 bp upstream of the usual splice site, so that the *A. halleri HTR12A* exon 2 is two amino acids longer than the other *HTR12* loci. Both duplicates are probably functional, since this product was obtained from cDNA from *A. halleri* ssp. *gemmifera* whole young shoots. Furthermore, mean site divergence values per replacement site (K_a) are low between the A and B copies, compared with silent-site divergence values ($K_a = 0.020$, $K_s = 0.078$, $K_{silent} = 0.069$), and between each of them and the *A. thaliana HTR12* sequence ($K_a = 0.064$, $K_s = 0.132$, $K_{silent} = 0.102$).

Chromosomal locations of the duplicated *HTR12* genes: To test the interpretation that the two types of *HTR12* sequences represent a duplication, we used genetic mapping using polymorphisms in a mapping family of *A. lyrata* ssp. *petraea* (KUITTINEN *et al.* 2004). Because one parent was homozygous for *HTR12B*, and the other for *HTR12A*, linkage of the duplicated loci could not be tested directly, but we infer that they are unlinked from their positions in the genetic map (Figure 2). In *A. thaliana*, the marker closest to the single *HTR12* locus (At1g01370) is *PVV4* (At1g01480), at 3.2 cM, near the tip of the short arm of chromosome 1. *HTR12B* is probably located at the same position as the *A. thaliana HTR12*, as it is closely linked (in the *A. lyrata* linkage group 1) to several genes that, in *A. thaliana*, are near *HTR12*: LHY (MIPS code At1g01060, one recombinant/82 meioses scored), F20D22 (At1g04120, one recombinant/83 meioses), CRY2 (At1g04400, two recombinants/80 meioses), and PHYA (At1g09570, four recombinants/87 meioses). In contrast, *HTR12A* is closely linked to the following genes in the long arm of *A. thaliana* chromosome 1, also in the *A. lyrata* linkage group 1: VIP1 (At1g43700, eight recombinants/81 meioses) and RHL1 (At1g48380, one recombinant/75 meioses). The locations of closely linked genes in *A. thaliana*, together with the general conservation of gene order between the *A. thaliana* and *A. lyrata* maps of this chromosome (KUITTINEN *et al.* 2004; KOCH and KIEFER 2005;

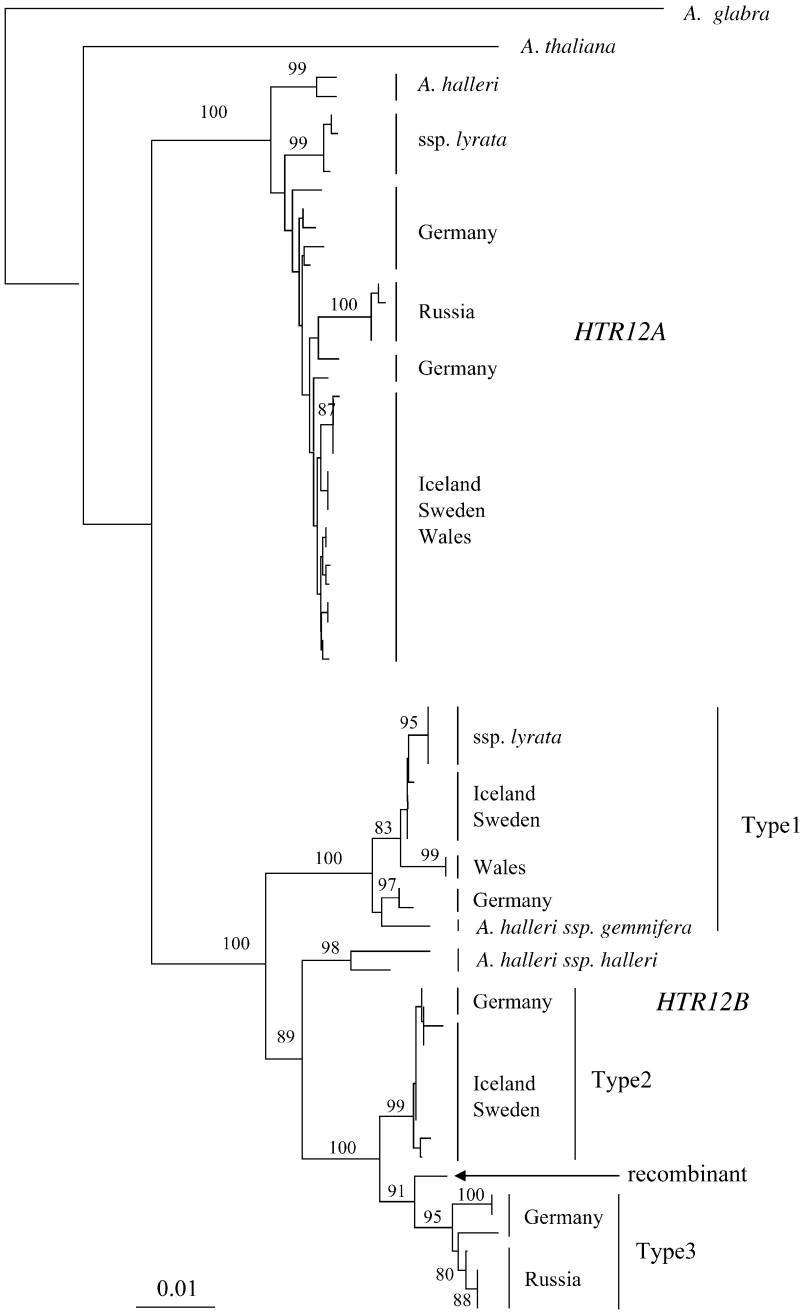

FIGURE 3.—Neighbor-joining tree based on the nucleotide differences between the *HTR12* sequences, based on the entire region. Bootstrap values >80% are shown by the relevant nodes. A distance bar is shown below the trees.

YOGEESWARAN *et al.* 2005; HANSSON *et al.* 2006), indicate that the duplicate *HTR12* genes are probably at opposite ends of the chromosome corresponding to the *A. lyrata* linkage group 1.

The origin of the duplication: The divergence between the *A. thaliana* sequence and the duplicated *HTR12* loci is much larger than between the duplicated loci in *A. halleri* and *A. lyrata*, suggesting that the duplication occurred after the ancestor of *A. halleri* and *A. lyrata* split from that of *A. thaliana*, assuming that mutations accumulated in the two lineages at similar rates. The respective mean divergence values for all sites are 0.0916 *vs.* 0.0557; the silent-site divergence from the *A. thaliana* sequence is also about twice that between the duplicated loci (see above). In *A. glabra*, moreover, we found only a single *HTR12* gene, supporting the conclusion that the duplication occurred after the split of *A. thaliana* from the lineage leading to *A. halleri* and *A. lyrata*. The phylogeny, with *A. glabra* as an outgroup species, is shown in Figure 3 (the diversity results within *A. lyrata* are described below).

These findings seem to rule out the alternative possibility, suggested by the location of the *HTR12A* locus, that the duplication is an ancient one that was lost from *A. thaliana* as part of the fusion event that created its chromosome 1, which is the derived state (KOCH and KIEFER 2005). *HTR12B* is probably the ancestral copy, because of its identical location to *A. thaliana HTR12*

TABLE 2

Nucleotide variants and diversity estimates (π) in the *HTR12* loci

		Coding region											
		Entire region			Synonymous			Nonsynonymous			Noncoding		
Taxon	N	No. of sites	S_n	π	No. of sites	S_n	π	No. of sites	S_n	π	No. of sites	S_n	π
						HTR12A							
All	30	1600	57 (60)	0.0085	129.32	3	0.0072	377.68	8	0.0062	1087	46 (49)	0.0095
ssp. *petraea*	26	1626	50 (53)	0.0076	129.27	3	0.0056	377.73	6	0.0048	1113	41 (44)	0.0088
ssp. *lyrata*	4	1767	3	0.0009	139.67	0	0	406.33	1	0.0012	1218	2	0.0009
						HTR12B							
All	30	1546	114 (115)	0.0262	124.33	12	0.0374	361.67	3	0.0033	1055	98 (99)	0.0328
ssp. *petraea*	26	1546	111 (112)	0.0247	124.33	11	0.0340	361.67	2	0.0024	1055	97 (98)	0.0312
ssp. *lyrata*	4	1776	0	0	136.67	0	0	391.33	0	0	1245	0	0

Numbers of mutations are shown in parentheses. S_n, number of polymorphic sites.

and *HTR12A*, a recently duplicated copy in the lineage leading to *A. halleri* and *A. lyrata*. Since *HTR12A* has an exon–intron structure in *A. lyrata* identical to that of other *HTR12* genes, the duplication event was not via a messenger RNA intermediate, but probably involved a transposition of genomic sequence.

Diversity in the duplicated *HTR12* loci: To test for the action of selection, we also obtained data on diversity within species for the duplicate loci.

Within-population diversity: *A. lyrata HTR12A* is considerably less polymorphic than *HTR12B*. For the entire region sequenced, the nucleotide diversity estimates, π, are 0.8 and 2.6%, respectively (Table 2; similar values of 0.7 and 2.5%, respectively, are found in *A. lyrata* ssp. *petraea*), but the difference is nonsignificant by a HKA test ($P = 0.187$). In addition to 57 single nucleotide polymorphisms, there are 41 length variants in the *HTR12A* locus, 4 of these indels being in the coding region. In the *HTR12B* locus, there are 114 polymorphic sites and 29 length variants (only 1 in the coding region). Two plants from the Plech population have an identical large deletion in *HTR12B* (101 bp), starting in intron 2; this causes loss of the intron–exon junction and part of exon 3 and shifts the reading frame, so this sequence is probably an inactive allele. Alternatively, since this deletion causes loss of the splicesome recognition site in intron 2, part of the intron 2 sequence may be included in the mRNA. The transcript of this allele would then include 10 bp of intron 2 sequence, but would have lost 28 bp of exon 3, forming a protein 6 amino acids shorter than that encoded by the nondeletion alleles, and with no homology between the 7 amino acids changed in the deletion alleles and the presumably wild-type sequence (13 amino acids), although we could not test this hypothesis by analyzing cDNAs because only dried leaves were available, from which mRNA could not be extracted.

The diversity estimates (Table 2) are similar to those for other nuclear genes in *A. lyrata* ssp. *petraea* (Wright *et al.* 2003; Wright 2003; Ramos-Onsins *et al.* 2004). The mean silent site π-value based on 30 non-*HTR12* loci is 1.4%, excluding *PgiC*, whose diversity is very high, but may be duplicated (see Kawabe and Miyashita 2002b).

The higher polymorphism in *HTR12B* is due mainly to the presence of three divergent types of sequences with high synonymous and noncoding site divergence (6/12 synonymous and 47/99 noncoding variations were fixed between sequence types, although, with our small sample size, the numbers of fixed sites will be overestimated), but there are only three nonsynonymous variants, and these are not fixed between the sequence types, so they appear not to represent functionally different alleles. Two of the three types have diversity similar to that of the entire set of *HTR12A* sequences. The three types are not simply local variants, since all are widely distributed geographically, and some populations include several (the Plech population from Germany has all three). Type 1 is found in the North American subspecies, *A. lyrata* ssp. *lyrata*, and in four European *A. lyrata* ssp. *petraea* populations, as well as in *A. halleri* ssp. *gemmifera* (Figure 3). Sequence type 2 is much less polymorphic than the other types (or than *HTR12A*) and was found in three geographically distant European ssp. *petraea* populations: Mount Esja (Iceland), Stubbsund (Sweden), and Plech (Figure 3). All the sequences from the Karhumaki population (Russia) are of type 3, together with some from Plech. Further evidence for long-term persistence of the three different *HTR12B* sequence types is that there are many shared polymorphisms between *A. lyrata* and *A. halleri*. The sequences of an *A. halleri* ssp. *halleri* plant from Switzerland cluster separately from the other sequences, but are most similar to sequence types 2 and 3, whereas the *A. halleri* ssp. *gemmifera* sequence is similar to sequence type 1. These results suggest that the different *HTR12B* sequence types originated before *A. halleri* and *A. lyrata* split from one another.

TABLE 3

Frequencies of the three major *HTR12B* sequence types determined by PCR–RFLP analyses

			Genotypes											
			Homozygotes			Heterozygotes			Sequence types			Frequencies of sequence types		
Population	Country	No. of plants	1	2	3	1/2	1/3	2/3	1	2	3	1	2	3
Mount Esja	Iceland	25	8	12		5			21	29	0	0.420	0.580	0
Reykjanes	Iceland	2	2						4	0	0	1	0	0
Clogwyn D'ur Arddu	Wales	4	4						8	0	0	1	0	0
Glaslyn	Wales	8	8						16	0	0	1	0	0
Plech	Germany	14	1	2		3	3	5	8	10	8	0.308	0.385	0.308
Lom	Norway	8	8						16	0	0	1	0	0
Spiterstulen	Norway	8	8						16	0	0	1	0	0
Stubbsund	Sweden	13	10	3					20	6	0	0.769	0.231	0
Karhumaki	Russia	9			9				0	0	18	0	0	1
Ontario	Canada	4	4						8	0	0	1	0	0
Indiana	United States	4	4						8	0	0	1	0	0
Total		99	57	17	9	8	3	5	125	45	26	0.638	0.230	0.133

See MATERIALS AND METHODS.

In contrast, the *HTR12A* sequences of *A. lyrata* and *A. halleri* cluster separately from one another, and even sequences from the two *A. lyrata* subspecies, *lyrata* and *petraea*, can be distinguished in the phylogenetic tree (with 99% bootstrap support; Figure 3). In both *HTR12A* and *HTR12B* trees, sequences from the populations from Wales, Karhumaki, and Ontario form clusters, supporting previous suggestions of some isolation between these geographically distant *A. lyrata* populations (WRIGHT *et al.* 2003; RAMOS-ONSINS *et al.* 2004). However, all these populations have very low within-population polymorphism that could cause separate clustering.

F_{ST} for *HTR12A* exceeds 70% for 10 of 15 between-population comparisons, and the mean for all populations is 0.71 (average F_{ST} in subspecies *petraea* is 0.62, similar to the value based on 18 reference loci studied in the same populations, which yielded a mean of 0.58, including ssp. *lyrata* and 0.54 for the European ssp. *petraea* populations; A. KAWABE, S. I. WRIGHT, A. FORREST and D. CHARLESWORTH, unpublished data).

Frequencies of the different *HTR12B* sequence types: Because our samples are small, they only poorly estimate the frequency distribution of the three *HTR12B* sequence types. To estimate frequencies, we therefore also did PCR–RFLP analyses for 99 plants from 11 populations from seven countries (Table 3). Eight of the populations were found to be monomorphic for one sequence type, mostly type 1 (the only exception is the Karhumaki population, fixed for sequence type 3). However, three populations showed polymorphism by PCR–RFLP, where two or three sequence types were found by sequence analyses. All three polymorphic populations have intermediate frequencies of each sequence type. In the Mount Esja population, the genotype frequencies do not deviate from Hardy–Weinberg equilibrium frequencies ($\chi^2 = 0.877$). However, Plech has an excess of heterozygotes and deviates significantly (11 heterozygotes of 14 plants, $\chi^2 = 7.942$, $P < 0.01$), while none of the 13 plants in the Stubbsund sample was heterozygous ($\chi^2 = 10.692$, $P < 0.01$).

Patterns of polymorphism at the *HTR12* loci: Despite the highly divergent *HTR12B* sequences, only Fu's F_s statistic (FU 1997) for the *HTR12A* locus suggests a significant deviation from neutrality (Table 4). This statistic is highly sensitive to hitchhiking events and population growth. Thus this locus may have experienced a recent bottleneck event. This result remains significant for just the *A. lyrata* ssp. *petraea* sequences.

We found a high proportion of significant linkage disequilibria between pairs of sites in *HTR12B* data sets (27% of site pairs are significant after Bonferroni correction). Wall's *B* and *Q* statistics are also both significant for the *HTR12B* sequences (Table 4). Nevertheless, many recombination events are detectable in both *HTR12A* and *HTR12B* (Table 4).

In the *A. lyrata* sequences, *HTR12A* has eight replacement polymorphisms and three synonymous ones, while for *HTR12B* the ratio of replacement-to-synonymous polymorphisms is 3:12, a significant difference. A McDonald–Kreitman test using divergence between the duplicated loci in *A. lyrata* gave a significant excess of replacement polymorphisms in *HTR12A*, whereas *HTR12B* showed no significant difference from neutrality. The only form of selection suggested by these results is purifying selection in *HTR12B* since the gene duplication. Using divergence from the *A. thaliana* sequence, a McDonald–Kreitman test was not significant for *HTR12A*, but was significant ($P < 0.05$) for the *HTR12B*

TABLE 4

Tests of neutrality and results of estimates of linkage disequilibrium and recombination

	HTR12A			*HTR12B*	
Locus and taxa	All	ssp. *petraea*	ssp. *lyrata*	All	ssp. *petraea*
Sample size (alleles)	30	26	4	30	26
Tajima's *D*	−0.389	−0.444	0.168	1.518	1.172
Fu and Li's D^a	0.063	0.009	−0.368	0.997	0.986
Fu's F_s	−7.962**	−5.664*	−2.181	4.271	2.987
Fay's H^a	−0.055	−2.055	0.667	−1.113	−5.772
LD					
No. of parsimony informative sites	38	31	1	98	95
Significant pairs ($P < 0.05$ by Fisher's exact test)	215/703	169/465	NA	2831/4753	2326/4465
Significant pairs after Bonferroni correction	44 (6.26%)	32 (6.88%)	NA	1296 (27.27%)	707 (15.83%)
Kelly's *ZnS*	0.173	0.229	NA	0.374	0.376
Wall's *B*	0.162	0.167	NA	0.433*	0.457*
Wall's *Q*	0.263	0.29	NA	0.602*	0.632*
R/site	0.0090	0.0045	NA	0.0025	0.0021
R_m	11	10	NA	7	7

HTR12B was not analyzed separately for *A. lyrata* ssp. *lyrata* because there were no polymorphic sites. *$P < 5\%$; **$P < 0.01$.
[a] Estimated using *A. thaliana* sequence as the outgroup.

locus, which shows an excess of fixed replacement variants, but few replacement polymorphisms (Table 5), suggesting the action of directional selection. Previous studies of *HTR12* also found unusually high replacement divergence between several Arabidopsis species (Talbert *et al.* 2002; Cooper and Henikoff 2004). To infer when these substitutions (and thus the selection) occurred, we determined the direction of each mutation by parsimony using the sequence of the single copy in *A. glabra*. This analysis placed seven synonymous and 13 replacement substitutions in the *A. thaliana* lineage, while four synonymous and 5 replacement substitutions are assigned to the *A. lyrata* lineage before the duplication event (for two of the replacement sites that differ between the Arabidopsis species, we cannot determine the lineage, because there is a sequence gap in *A. glabra*). These differences between substitutions before and after the duplication are not significant by a *G*-test. The replacement fixations are therefore probably not specifically related to the recent duplication. Overall, there is little evidence for selectively driven substitutions in either *A. lyrata* gene.

TABLE 5

McDonald–Kreitman tests

	Polymorphisms within *A. lyrata*		Fixed differences from *A. thaliana*			Fixed differences from other locus		
Locus	Synonymous	Replacement	Synonymous	Replacement	Significance	Synonymous	Replacement	Significance
HTR12A	3	8	15	21	NS	6	2	
HTR12B	12	3	12	19	*	6	2	NS

	Polymorphisms within *A. lyrata*	
Lineage[b]	Synonymous	Replacement
A. thaliana	7	13
A. lyrata		
Before duplication	4	5
HTR12A	4	1
HTR12B	2	1

Note that the number of polymorphic sites differs between the comparisons because of alignment gaps. *$P < 0.05$.
[a] One replacement polymorphic site in *HTR12A* is found in polymorphic indel variation in *HTR12B*. If this mutation is excluded, the comparison between 3:7 *vs.* 6:2 is not significant ($P = 0.063$).
[b] Directions of mutations are determined by parsimony with *A. glabra* sequence. Two replacements found only in *A. thaliana* sequence locate in a sequence gap in *A. glabra*. One synonymous and two replacement changes in the *A. thaliana* lineage are in the polymorphic indel variation in *HTR12B*.

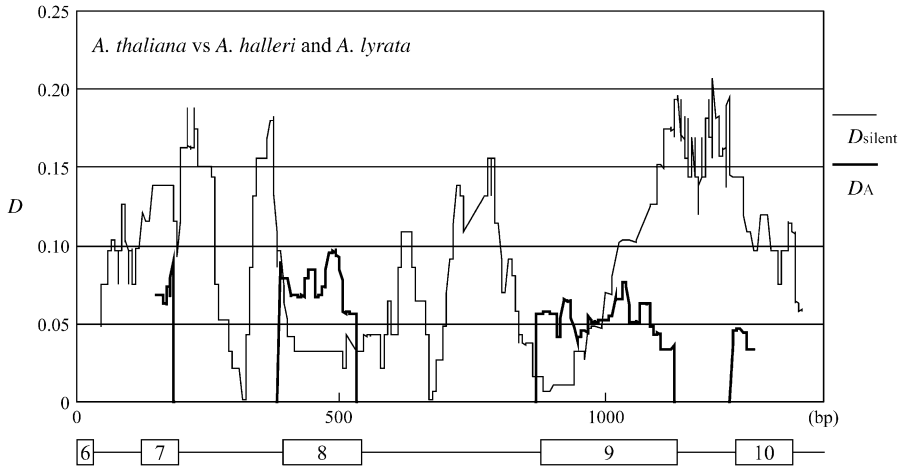

FIGURE 4.—Sliding-window analyses for the CENP-C gene. Sliding-window plot for divergence between *A. thaliana* and *A. halleri–A. lyrata* is shown. The exon–intron structure is indicated below the plot. Divergence values are plotted in 1-bp steps in 50- and 100-bp windows for silent (thin line) and replacement (thick line) sites, respectively.

CENP-C gene divergence: To test for selection acting on the other centromere protein gene, CENP-C, we sequenced the gene in *A. lyrata* and *A. halleri* and analyzed divergence from *A. thaliana*. We determined sequences for one plant from each *A. halleri* and *A. lyrata* subspecies for this genomic region for which adaptive evolution was suggested between *A. thaliana* and *A. arenosa* (TALBERT *et al.* 2004); the region extends from exon 6 to intron 10 and its length is ~1.4 kb. Using the *A. glabra* sequence as an outgroup, most mutations could be assigned to individual branches by parsimony. In the coding region analyzed (~600 bp), the high D_N/D_S values tend to be found for branches that have high D_S. The highest D_N/D_S ratio is inferred in the *A. glabra* lineage (1.423), and the lowest in the *A. halleri* lineage (0.184). Except in the *A. glabra* and *A. thaliana* lineages, D_N values are lower than D_S or D_{intron} values.

Although relatively high D_N/D_S ratios are seen in CENP-C, in almost all species comparisons this is mainly due to low synonymous divergence. Exons 8 and 9, which are a putative region undergoing adaptive evolution (TALBERT *et al.* 2004), have no synonymous substitutions, even in comparison with the *A. glabra* sequence. On the other hand, there are many substitutions at replacement and intron sites (Figure 4).

Among the taxa that have multiple centromere satellite families (*A. halleri* and *A. lyrata*), divergence estimates for this gene are not unusually low ($D_S = 0.042$, $D_N = 0.018$, and $D_{intron} = 0.032$ between the species), ruling out a recent selective sweep before the split of the two species, such as could be due to selection associated with the fixation of centromere satellite sequences shared between these two species. The sequences from the two subspecies within both *A. lyrata* and *A. halleri* have very few differences (Figure 5; note that no intron sequence is available for *A. arenosa*). Thus an interpretation involving independent selective sweeps in each species after their split is possible, but it is evident that neither species has maintained different CENP-C alleles that could have different DNA–protein interactions. The gene trees for synonymous and replacement sites also suggest a divergence time for CENP-C sequences of *A. lyrata* and *A. halleri* very close to these species' split from the *A. arenosa* sequence. Two major centromere satellite families, pAge1 and pAge2, are shared by *A. lyrata* and *A. halleri* but not found in *A. arenosa*. If their appearance has led to replacement of CENP-C by new, functionally different alleles since these species split, one would expect the *A. lyrata* and *A. halleri* sequences to be much more similar to each other than to *A. arenosa*. Our results do not, therefore, support a conclusion of adaptive changes in CENP-C.

DISCUSSION

Duplication of the *HTR12* gene in *A. halleri* and *A. lyrata*: This is the first known case of duplication of the centromeric histone H3 (CenpA) gene in a diploid Arabidopsis species. TALBERT *et al.* (2002) suggested that, in allopolyploid species, the two CENP-A proteins from the different parent species might form heterodimers, sometimes allowing coexistence of both species' centromeric satellite sequences. However, this state probably does not persist in the long term. In *Zea mays*, which is a fairly recent allotetraploid (due to an event estimated to be ~11 MYA; GAUT and DOEBLEY 1997), there is now only one centromeric histone H3 gene copy (ZHONG *et al.* 2002). The only other known case of duplication is in the nematode *Caenorhabditis elegans*, which has two CENP-A loci, while the related species *C. briggsae* has just one copy, but one copy is not expressed, or is expressed very weakly (MONEN *et al.* 2005). Duplication and maintenance of two or more copies of CENP-A thus appears to be rare in both plant and animal species. *A. halleri* and *A. lyrata* may be in a situation similar to that in the early stage of allopolyploid formation; the HTR12 proteins encoded by the duplicate genes have diverged in sequence, and it remains to be tested in the future whether they now colocalize at

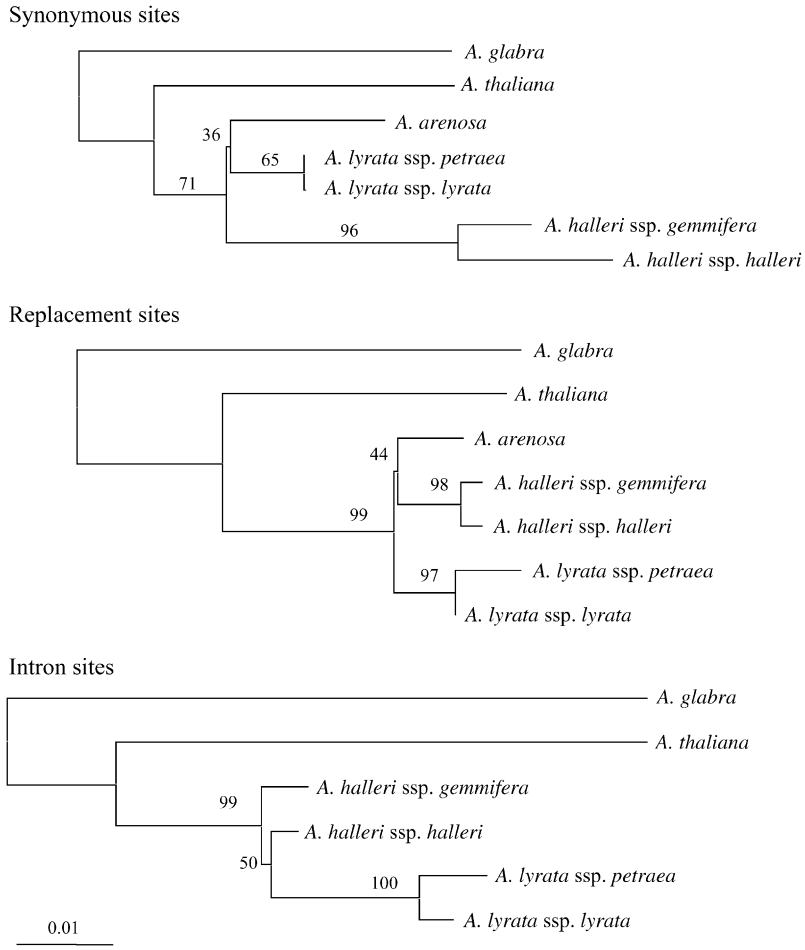

FIGURE 5.—Neighbor-joining tree based on the nucleotide differences between the CENP-C sequences. Trees based on the synonymous sites, replacement sites, and intronic region are shown. Bootstrap values are shown by the relevant nodes. All trees have the same scale (see the distance bar below the trees).

different centromere regions and bind different centromere repeats. Since there are three major satellite families, but only two *HTR12* genes, there is not a simple one-to-one relationship between HTR12 and centromeric satellite families. *A. halleri* and *A. lyrata* may thus represent a transient stage with protein–DNA interactions whose specificity is not stringent.

Evidence for selection: If centromere evolution drives the evolution of differences in centromeric histone H3 genes (*HTR12* in Arabidopsis) to acquire altered binding specificities to prevent distorted segregation of different centromere satellite variants (HENIKOFF *et al.* 2001; MALIK and HENIKOFF 2001), the two *HTR12* loci may be expected to have experienced different selective pressures. For example, one of the duplicates might experience an evolutionary "arms race" with a centromere and should then have many fixed replacement substitutions (and low within-locus diversity may be found, if a selective sweep has occurred recently). Neither of the two duplicated *HTR12* loci in *A. lyrata* shows clear evidence of the expected multiple replacement substitutions fixed in their lineages even though relaxed selective constraint might be expected for duplicated genes.

The *HTR12A* locus: Polymorphism in *A. lyrata* is low for the duplicated *HTR12A* locus, which might suggest a recent selective sweep, and we find a significant result from one test that can detect such events (Fu's F_s test, which detects an unusually low frequency of old alleles, *i.e.*, an excess of variants at frequencies lower than predicted by a constant population size and neutral model, indicating recent mutations on the external branches). The low *HTR12A* locus polymorphism, and also the high relative number of replacement polymorphisms, could be due to a recent bottleneck affecting either this species or just this gene, since a prolonged bottleneck of low population effective size can hinder the elimination of slightly deleterious replacement mutations. The significant F_s value is seen only for *HTR12A*, whereas a recent strong bottleneck should affect all loci (*e.g.*, the one inferred in *A. halleri* ssp. *gemmifera*, which was detected at four loci; see KAWABE and MIYASHITA 2003). However, data from other *A. lyrata* loci do not suggest a bottleneck in population size (only 2 of 18 reference loci showed significantly negative F_s values, A. KAWABE, S. I. WRIGHT, A. FORREST and D. CHARLESWORTH, unpublished results). In the future, it will be interesting to analyze more loci, including genes near the *HTR12A* locus.

If bottlenecks are ruled out, a selective sweep in the *HTR12A* locus in *A. lyrata* may be a possible

explanation for our results for this locus, although there is certainly no evidence for an "arms race"-driven substitution process and no significant excess of replacement substitutions.

The *HTR12B* locus: An excess of replacement substitutions was found for *HTR12B*, which is probably the ancestral copy, but this was due to substitutions in the *A. thaliana* lineage, not in *HTR12B*. Moreover, this gene has high diversity in *A. lyrata* and shows evidence of long-term maintenance of several different haplotypes, clearly dating to before the split between *A. lyrata* and *A. halleri*. Our segregation analyses establish that the different sequence types are allelic. *HTR12B* has therefore not undergone the repeated selective sweeps expected under an "arms race" scenario.

It is possible that persistence of these sequence types is related to the presence of the different major centromere sequences in our study species. The linkage disequilibrium that we observe is consistent with long-term balancing selection, but none of the tests that we applied detects significant deviations from neutrality. A low recombination frequency (or a high degree of selfing, implying a low effective recombination rate) can also cause linkage disequilibria (Nordborg *et al.* 2000). However, *HTR12B* is not in a genomic region of low recombination rate (Hansson *et al.* 2006), and *A. lyrata* is an outcrossing plant in most populations; of the populations studies here, only the Ontario population (population TC in Mable *et al.* 2005) may be partially inbreeding. Larger samples from individual populations may be needed to detect a signature of balancing selection by Tajima's test, since pooled data from different populations can be misleading (see Schierup *et al.* 2000). However, the fact that no replacement sites are fixed between the three *HTR12B* sequence types suggests that they may not represent functionally different types. We also cannot exclude the possibility that a gene very closely linked to *HTR12B* experiences balancing selection and influences the *HTR12B* locus polymorphism. The surrounding genome region should then show linkage disequilibrium with *HTR12B* and should also have high diversity (Charlesworth *et al.* 1997; Takahata and Satta 1998; Schierup *et al.* 2000; Navarro and Barton 2002); these are testable predictions, but have not yet been tested.

Although the reasons for the maintenance of the divergent *HTR12B* alleles are thus unclear, it is clear that this locus has not been the target of a centromere satellite–protein coevolutionary arms race. The only clear evidence that we have found for selection suggests that the *HTR12B* locus is under purifying selection. The *HTR12B* locus results suggest purifying selection, while *HTR12A* has low polymorphism and might possibly have undergone a selective sweep. HTR12B might have maintained interactions with the ancestral satellite sequence families, while HTR12A evolved to interact with new satellite families. However, the evidence is weak that HTR12A has been repeatedly selected for functional changes involving interaction with the new centromere sequences, because there is no evidence of an "arms race" since we do not observe multiple amino acid substitutions at this locus.

If these species have undergone a centromere "evolutionary arms race," it might perhaps have involved the other centromere protein, CENP-C, which, as mentioned above, has been reported as acting as a player of the centromere "arms race" evolution in both plants and animals (Talbert *et al.* 2004), including Arabidopsis species. Our finding of high divergence between *A. halleri* and *A. lyrata*, and low divergence within these species, does not, however, support CENP-C evolution associated with the multiple centromere satellite families shared between *A. halleri* and *A. lyrata*. The high D_N/D_S ratios between many species appear to be due to low synonymous changes, whose cause is not understood, and not to adaptive evolution or relaxed selective constraints. The cause of the reduced rate of synonymous change in the CENP-C gene of Arabidopsis is not understood, but constraints affecting synonymous sites are reported in other organisms and can cause local heterogeneity of synonymous divergence, and consequently high D_N/D_S ratios in certain regions of some genes (*e.g.*, Pond and Muse 2005; Charmary *et al.* 2006).

We thank H. Kuittinen and O. Savolainen for DNA samples from the mapping family, B. K. Mable for plant material, S. Preuss for isolation of genomic DNAs, and the Institute of Evolutionary Biology, Edinburgh University Sequencing service for sequencing.

LITERATURE CITED

Chamary, J. V., J. L. Parmley and L. D. Hurst, 2006 Hearing silence: non-neutral evolution at synonymous sites in mammals. Nat. Rev. Genet. **7:** 98–108.

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

Cooper, J. L., and S. Henikoff, 2004 Adaptive evolution of the histone fold domain in centromeric histones. Mol. Biol. Evol. **21:** 1712–1718.

Fay, J. C., and C.-I Wu, 2000 Hitchhiking under positive Darwinian selection. Genetics **155:** 1405–1413.

Fu, Y.-X., 1997 Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics **147:** 915–925.

Fu, Y.-X., and W.-H. Li, 1993 Statistical tests of neutrality of mutations. Genetics **133:** 693–709.

Gaut, B. S., and J. F. Doebley, 1997 DNA sequence evidence for the segmental allotetraploid origin of maize. Proc. Natl. Acad. Sci. USA **94:** 6809–6814.

Hansson, B., A. Kawabe, S. Preuss, H. Kuittinen and D. Charlesworth, 2006 Comparative gene mapping in *Arabidopsis lyrata* chromosomes 1 and 2 and the corresponding *A. thaliana* chromosome 1: recombination rates, rearrangements and centromere location. Genet. Res. **87:** 75–85.

Heeger, S., O. Leismann, R. Schittenhelm, O. Schraidt, S. Heidmann *et al.*, 2005 Genetic interactions of separase regulatory subunits reveal the diverged Drosophila Cenp-C homolog. Genes Dev. **19:** 2041–2053.

Henikoff, S., K. Ahmad, J. S. Platero and B. Vansteensel, 2000 Heterochromatic deposition of centromeric histone H3-like proteins. Proc. Natl. Acad. Sci. USA **97:** 716–721.

Henikoff, S., K. Ahmad and H. S. Malik, 2001 The centromere paradox: stable inheritance with rapidly evolving DNA. Science **293:** 1098–1102.

Hudson, R. R., M. Kreitman and M. Aguade, 1987 A test of neutral molecular evolution based on nucleotide data. Genetics **116:** 153–159.

Kamm, A., I. Glasso, T. Schmidt and J. S. Heslop-Harrison, 1995 Analysis of a repetitive DNA family from *Arabidopsis arenosa* and relationship between *Arabidopsis* species. Plant Mol. Biol. **27:** 853–862.

Kawabe, A., and N. T. Miyashita, 2002a DNA variation in the acidic chitinase locus (*ChiA*) region in *Arabis gemmifera* and among its related species. Genes Genet. Syst. **77:** 167–175.

Kawabe, A., and N. T. Miyashita, 2002b Characterization of duplicated two cytosolic phosphoglucose isomerase (*PgiC*) loci in *Arabidopsis halleri* ssp. gemmifera. Genes Genet. Syst. **77:** 159–165.

Kawabe, A., and N. T. Miyashita, 2003 DNA polymorphism in active gene and pseudogene of the cytosolic phosphoglucose isomerase (*PgiC*) loci in *Arabidopsis halleri* ssp. *gemmifera*. Mol. Biol. Evol. **20:** 1043–1050.

Kawabe, A., and S. Nasuda, 2005 Structure and genomic organization of centromeric repeats in *Arabidopsis* species. Mol. Genet. Genomics **272:** 593–602.

Kelly, J. K., 1997 A test of neutrality based on interlocus associations. Genetics **146:** 1197–1206.

Koch, M., and M. Kiefer, 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*. Am. J. Bot. **92:** 761–767.

Kuittinen, H., A. A. De Haan, C. Vogl, S. Oikarinen, J. Leppälä *et al.*, 2004 Comparing the linkage map of the close relatives *Arabidopsis lyrata* and *A. thaliana*. Genetics **168:** 1575–1584.

Kumar, S., K. Tamura, I. Jacobsen and M. Nei, 2000 MEGA2: Molecular Evolutionary Genetics Analysis, Version 2.0. Pennsylvania State University/Arizona State University, University Park, PA/Tempe, AZ.

Mable, B. K., A. V. Robertson, S. Dart, C. D. Berardo and L. Witham, 2005 Breakdown of self-incompatibility in the perennial *Arabidopsis lyrata* (Brassicaceae) and its genetic consequences. Evolution **59:** 1437–1448.

Malik, H. S., and S. Henikoff, 2001 Adaptive evolution of Cid, a centromere-specific histone in Drosophila. Genetics **157:** 1293–1298.

Malik, H. S., and S. Henikoff, 2002 Conflict begets complexity: the evolution of centromeres. Curr. Opin. Genet. Dev. **12:** 711–718.

Malik, H. S., D. Vermaak and S. Henikoff, 2002 Recurrent evolution of DNA-binding motif in the Drosophila centromeric histone. Proc. Natl. Acad. Sci. USA **99:** 1449–1454.

McDonald, J. H., and M. Kreitman, 1991 Adaptive protein evolution at the *Adh* locus in *Drosophila*. Nature **351:** 652–654.

Miyashita, N. T., A. Kawabe, H. Innan and R. Terauchi, 1998 Intra- and interspecific DNA variation and codon bias of alcohol dehydrogenase (*Adh*) locus in *Arabis* and *Arabidopsis* species. Mol. Biol. Evol. **15:** 1420–1429.

Monen, J., P. S. Maddox, F. Hyndman, K. Oegema and A. Desai, 2005 Differential role of CENP-A in the segregation of holocentric C. elegans chromosomes during meiosis and mitosis. Nat. Cell Biol. **7:** 1248–1255.

Nagaki, K., P. B. Talbert, C. X. Zhong, R. K. Dawe, S. Henikoff *et al.*, 2003 Chromatin immunoprecipitation reveals that the 180-bp satellite repeat is the key functional DNA element of *Arabidopsis thaliana* centromeres. Genetics **163:** 1221–1225.

Navarro, A., and N. H. Barton, 2002 The effects of multilocus balancing selection on neutral variability. Genetics **161:** 849–863.

Nordborg, M., 2000 Linkage disequilibrium, gene trees, and selfing: an ancestral recombination graph with partial self-fertilization. Genetics **154:** 923–929.

Pardo-Manuel de Villena, F., and C. Sapienza, 2001 Female meiosis drives karyotypic evolution in mammals. Genetics **159:** 1179–1189.

Pond, S. K., and S. V. Muse, 2005 Site-to-site variation of synonymous substitution rates. Mol. Biol. Evol. **22:** 2375–2385.

Ramos-Onsins, S. E., B. E. Stranger, T. Mitchell-Olds and M. Aguade, 2004 Multilocus analysis of variation and speciation in the closely related species *Arabidopsis halleri* and *A. lyrata*. Genetics **166:** 373–388.

Richards, S., Y. Liu, B. R. Betterncourt, P. Hradecky, S. Letovsky *et al.*, 2005 Comparative genome sequencing of Drosophila pseudoobscura chromosomal, gene and cis-element evolution. Genome Res. **15:** 1–18.

Rozas, J., and R. Rozas, 1999 *DnaSP* version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics **15:** 174–175.

Schierup, M. H., X. Vekemans and D. Charlesworth, 2000 The effect of hitch-hiking on genes linked to a balanced polymorphism in a subdivided population. Genet. Res. **76:** 63–73.

Sullivan, B. A., 2002 Centromere round-up at the heterochromatin corral. Trends Biotechnol. **3:** 89–92.

Tajima, F., 1989 Statistical test for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123:** 585–595.

Takahata, N., and Y. Satta, 1998 Footprints of intragenic recombination at *HLA* loci. Immunogenetics **47:** 430–441.

Talbert, P. B., R. Masuelli, A. P. Tyagi, L. Comai and S. Henikoff, 2002 Centromeric localization and adaptive evolution of an Arabidopsis histone H3 variant. Plant Cell **14:** 1053–1066.

Talbert, P. B., T. D. Bryson and S. Henikoff, 2004 Adaptive evolution of centromere protein in plants and animals. J. Biol. **3:** 18.

Wall, J. D., 1999 Recombination and the power of statistical tests of neutrality. Genet. Res. **74:** 65–79.

Wright, S. I., 2003 Effects of recombination rate and mating system on genome evolution and diversity in *Arabidopsis lyrata*. Ph. D. Thesis, University of Edinburgh, Edinburgh, United Kingdom.

Wright, S. I., B. Lauga and D. Charlesworth, 2003 Subdivision and haplotype structure in natural populations of *Arabidopsis lyrata*. Mol. Ecol. **12:** 1247–1263.

Yogeeswaran, K., A. Frary, T. L. York, A. Amenta, A. H. Lesser *et al.*, 2005 Comparative genome analyses of Arabidopsis spp.: inferring chromosomal rearrangement events in the evolutionary history of *A. thaliana*. Genome Res. **15:** 505–515.

Zhong, C. X., J. B. Marshall, C. Topp, R. Mroczek, A. Kato *et al.*, 2002 Centromeric retroelement and satellites interact with maize kinetochore protein CENH3. Plant Cell **14:** 2825–2836.

Communicating editor: S. W. Schaeffer