Available online at www.sciencedirect.com
ScienceDirect
Molecular Phylogenetics and Evolution 42 (2007) 122–130
MOLECULAR PHYLOGENETICS AND EVOLUTION
www.elsevier.com/locate/ympev

Further insights into the phylogeny of *Arabidopsis* (Brassicaceae) from nuclear *Atmyb2* flanking sequence

James B. Beck [a,b,*], Ihsan A. Al-Shehbaz [b], Steve L. O'Kane Jr. [c], Barbara A. Schaal [a]

[a] *Department of Biology, Washington University, St. Louis, MO 63130, USA*
[b] *Missouri Botanical Garden, St. Louis, MO 63166, USA*
[c] *Department of Biology, University of Northern Iowa, Cedar Falls, IA 50614, USA*

Received 2 April 2006; revised 14 June 2006; accepted 16 June 2006
Available online 5 July 2006

Abstract

Arabidopsis thaliana is the preeminent plant model organism. However, significant advances in evolution and ecology are being made by expanding the scope of research beyond this single species into the broader genus *Arabidopsis*. Surprisingly, few studies have rigorously investigated phylogenetic relationships between the nine *Arabidopsis* species, and this study evaluates both these and hypotheses related to two instances of intra-generic hybridization. DNA sequences from the 5′ flanking region of the nuclear *Atmyb2* gene from 12 of the 14 *Arabidopsis* taxa were used to reconstruct the generic phylogeny. The strict consensus tree was highly concordant with previous studies, identifying lineages corresponding to widespread species but exhibiting a large basal polytomy. Our data indicates that the paternal parent of the allopolyploid *A. suecica* is *A. neglecta* rather than *A. arenosa* s.l., although the need for a detailed phylogeographical study of these three species is noted. Finally, our data provided additional phylogenetic evidence of hybridization between *Arabidopsis lyrata* s.l. and *A. halleri* s.l. Taken together, the well-defined lineages within the genus and the potential for hybridization between them highlight *Arabidopsis* as a promising group for comparative and experimental studies of hybridization.

Keywords: Arabidopsis; Atmyb2 gene; Hybridization

1. Introduction

It would be difficult to overstate the importance of *Arabidopsis thaliana* (L.) Heynh. as an experimental organism. A May 2006 search of the ISI Web of Knowledge "SCI-EXPANDED" database for "*Arabidopsis thaliana*" recovered over 20,000 articles, and more than one million *A. thaliana* sequences are currently archived in the GenBank database. Current work with this species is central to many research areas, including genetics (Caicedo et al., 2004; Edwards et al., 2005; Quesada et al., 2002), genomics (Nordborg et al., 2002; Schmid et al., 2005; Tian et al., 2002), ecology (Callahan and Pigliucci, 2002; Griffith et al., 2004; Mauricio, 1998), and evolution (Hoffman, 2005; Nasrallah et al., 2004; Shimizu et al., 2004).

As the scope of *A. thaliana* research broadens, significant advances are being made by incorporating *A. thaliana*'s close relatives. The genus *Arabidopsis* (DC.) Heynh. comprises 14 Eurasian taxa (nine species, five additional subspecies) that occupy a broad range of habitats from high-elevation rock outcrops to sandy, ocean-side localities (O'Kane and Al-Shehbaz, 1997), providing unique avenues for ecological, genetic, and evolutionary research. Mitchell-Olds (2001) noted several advantages of *Arabidopsis* as a model genus, including the broad adaptive range of the species and the impressive array of genetic markers optimized for the group. He stated (pg. 697) that these advantages make it possible "...to determine the ecological and evolutionary forces influencing quantitative genetic variation in the wild, and to examine the geographic distribution of non-neutral allelic variants." Nasrallah et al. (2000) produced fertile hybrids from a cross between *A. thaliana* and *A. lyrata* (L.) O'Kane and & Al-Shehbaz subsp. *lyrata*, and

* Corresponding author. Fax: +1 314 935 4432.
E-mail address: beck@biology.wustl.edu (J.B. Beck).

1055-7903/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2006.06.011

discussed the potential to understand genome evolution in newly formed polyploid species. Koch et al. (2001b) and Hong et al. (2003) used "phylogenetic shadowing," or the identification of conserved genetic regions by comparing sequence from several closely related species (Boffelli et al., 2003), to find functionally important motifs in *Arabidopis* and other Brassicaceae. Hoffman (2005) tracked the evolution of climatic niche in *Arabidopsis* by optimizing this character onto a genus phylogeny, and Ramos-Onsins et al. (2004) shed light on the process of speciation by examining the fate of eight unlinked genes in *A. halleri* (L.) O'Kane and & Al-Shehbaz subsp. *halleri* and *A. lyrata* (L.) O'Kane and Al-Shehbaz subsp. *petraea* (L.) O'Kane and & Al-Shehbaz.

As research incorporating *A. thaliana* and its eight congeners expands, a well-resolved phylogeny will be essential. An accurate phylogenetic topology, complete with branch lengths, will allow researchers to pinpoint the origin of both genes and phenotypes, and accurately estimate how long these have persisted. Surprisingly, many questions remain regarding the phylogeny of this group. Although nine studies (Beilstein et al., 2006; Koch et al., 1999, 2000, 2001a,b; Miyashita et al., 1998; O'Kane and Al-Shehbaz, 2003; Price et al., 1994; Shimizu et al., 2005) have included multiple *Arabidopsis* species in phylogenetic analyses of sequence data, only one (O'Kane and Al-Shehbaz, 2003) included more than half of the 14 *Arabidopsis* taxa. The last authors examined nuclear internal transcribed spacer (ITS) data in 11 *Arabidopsis* taxa, recovering a well-supported, monophyletic *Arabidopsis*, with moderate support for several clades corresponding to species or species groups within the genus. Deeper nodes within *Arabidopsis* were unresolved. Beyond basic phylogenetic relationships, questions related to hybridization also remain. *Arabidopsis suecica* (Fries) Norrlin ($2n = 26$) is one of the most intensively studied hybrid species, the focus of nearly 50 years of taxonomic, cytogenetic, and molecular phylogenetic study (reviewed in O'Kane et al., 1997). Recent work has indicated that the species is of a single origin and that *A. thaliana* ($2n = 10$) is the maternal parent (Jakobsson et al., 2006; O'Kane et al., 1997; Price et al., 1994; Säll et al., 2003). O'Kane et al. (1997) cloned and sequenced individual nuclear rDNA internal transcribed spacer (ITS) repeats from *A. suecica*, some of which were identical to the maternal parent (*A. thaliana*), the remainder being nearly identical to repeats found in both *A. arenosa* (L.) Lawalrée subsp. *arenosa* (2*n* highly variable—discussed below) and *A. neglecta* (Schultes) O'Kane & Al-Shehbaz ($2n = 16, 32$). It should be noted that we use "*A. arenosa* s.l." to refer to the entire *A. arenosa* lineage, (both *A. arenosa* subsp. *arenosa* and *A. arenosa* (L.) Lawalrée subsp. *borbasii* (Zapalowicz) O'Kane and & Al-Shehbaz) and subspecific designations are used to indicate the two subspecies. Since *A. arenosa* s.l. is widespread in Europe and is frequent at low to moderate elevations, while *A. neglecta* is a high-elevation Carpathian endemic, the authors favored the former as the paternal parent, however they could not rule out *A. neglecta*. O'Kane and Al-Shehbaz (1997) note that both *A. neglecta* and the Croatian endemic *A. croatica* (Schott) O'Kane and Al-Shehbaz are morphologically similar to *A. arenosa* s.l., and these three species form a monophyletic group (along with certain clones from *A. suecica*) in the O'Kane and Al-Shehbaz (2003) ITS analysis. We will refer to these three taxa (*A. arenosa* s.l., *A. croatica*, and *A. neglecta*) as the "*A. arenosa* group." Further work is needed to evaluate the origin of *A. suecica* and general evolutionary relationships within this group. Instances of hybridization have also been reported between the *A. halleri* s.l. and *A. lyrata* s.l. lineages. Macnair et al. (1999) successfully produced experimental hybrids between *A. halleri* subsp. *halleri* and *A. lyrata* subsp. *petraea*, and Ramos-Onsins et al. (2004) provided evidence of natural hybridization between these two taxa. They observed a high number of shared ancestral polymorphisms between these two taxa over their eight-locus data set, and the sharing of nearly identical haplotypes at one locus. Shimizu et al. (2005) observed divergent haplotypes at the chalcone synthase (*Chs*) locus within two samples of *A. lyrata* (L.) O'Kane and Al-Shehbaz subsp. *kamchatica* (Fischer ex DC.), with some haplotypes forming a clade with haplotypes from *A. lyrata* subsp. *lyrata*, the others forming a clade with haplotypes from *A. halleri* (L.) O'Kane and Al-Shehbaz subsp. *gemmifera* (Matsumura) O'Kane and & Al-Shehbaz, a pattern similar to that observed recently by J. Steets, N. Takebayashi, and D. Wolf (unpublished data) at several nuclear loci. However, several other studies of *Arabidopsis* have failed to provide evidence of hybridization between *A. halleri* s.l. and *A. lyrata* s.l., each of which is comprised of three subspecies. Studies including at least one subspecies from each lineage could potentially provide evidence of hybridization. Such a pattern has not been observed in any other such study to date (Koch et al., 1999, 2000, 2001a,b—nuclear *Adh*, *Adh* promoter, *Apetala* promoter, *Chs*, ITS, chloroplast *matK*; O'Kane and Al-Shehbaz, 2003—ITS; Miyashita et al., 1998—*Adh*).

Given that further work with *Arabidopsis* will benefit greatly from a more detailed phylogenetic understanding of the genus and that the above issues remain unresolved, we will analyze sequence variation at a low copy nuclear region (the 5′ flanking sequence of the *Atmyb2* gene) in 12 of the 14 *Arabidopsis* taxa. *Atmyb2* is a regulatory gene located on the long arm of chromosome two in *A. thaliana* (Urao et al., 1993) and has been shown to be evolving in a neutral fashion in this species (Kamiya et al., 2002). Although this gene has not been used previously for phylogeny reconstruction, it exists as a single copy in *A. thaliana* and exhibits intra-specific variation within that species (Beck et al., unpublished data; Kamiya et al., 2002), making it an ideal candidate for this type of study. Specifically, we wish to answer the following questions: (1) What are the deep relationships between species and species complexes

in *Arabidopsis*? (2) Is *A. arenosa* s.l. the paternal parent of *A. suecica*? And (3), is there evidence for introgression between *A. halleri* s.l. and *A. lyrata* s.l.?

2. Materials and methods

2.1. Sampling, amplification, and sequencing

Sample information is given in Table 1. This study included 12 of the 14 currently recognized *Arabidopsis* taxa, *A. arenosa* subsp. *borbasii* and *A. pedemontana* (Boiss.) O'Kane & Al-Shehbaz were not included due to difficulty in obtaining material. The exclusion of these taxa is not likely to greatly impact our inferred phylogeny, as the morphology of both strongly suggest sister relationships to taxa included in our study. *A. arenosa* subsp. *borbasii* differs from the more widespread *A. arenosa* subsp. *arenosa* by only minor seed and leaf characters, while *A. pedemontana* is simply smaller and more glabrous relative to *A. cebennensis* (DC.) O'Kane and Al-Shehbaz (O'Kane and Al-Shehbaz, 1997). A portion of the 5′ flanking region of the nuclear *Atmyb2* gene was amplified with the primers "ATM-1" (5′-ccctaaactcgcctaactcc-3′) and "ATM-4" (5′-attcgctcgtaatcttcc-3′). All regions were PCR-amplified under standard conditions. Reactions produced one or two distinct fragments, which were cut from a 1.5% agarose gel, cleaned with a Geneaid kit (Geneaid), and cloned into a pGEM-T vector system (Promega). One to six clones per fragment were amplified, dye-labeled using a Big Dye Terminator Kit (Applied Biosystems), and sequenced on an Applied Biosystems 3130xl Genetic Analyzer. Due to difficulty encountered in aligning the entire ATM-1 to ATM-4 region the primer ATM-1 was used with "ATM out3" (5′-gaactctcttaccaaagaaag-3′) to amplify, dye-label, and sequence an approximately 360 bp portion of 3′ end of the ATM-1 to ATM-4 region.

2.2. Phylogenetic analyses

Identical haplotypes from the same sample were not analyzed. Sequences were manually aligned in Se-Al (Rambaut, 1996) and the aligned matrix was exported as a NEXUS file. Insertion/deletion events were coded as additional characters and added to the end of the NEXUS file as 1/0 characters. All insertion/deletion events, both autapomorphic and synapomorphic, were scored except in the case of multi-state repeats (such as a T/A repeat) or poor alignment. In the case of overlapping insertion/deletion events, the "simple gap coding" method of Simmons and Ochoterena (2000) was used. Because the monophyly of *Arabidopsis* is not in question, and in order to avoid alignment problems due to the inclusion of more divergent taxa, we chose to root trees with the *A. thaliana Atmyb2* flanking sequence from the published genome. *A. thaliana* has repeatedly been shown to be sister to the remaining *Arabidopsis* species (Koch et al., 2001a; Miyashita et al., 1998; O'Kane and Al-Shehbaz, 2003). A heuristic parsimony search with 100 random addition replicates was performed using PAUP* 4.0b10 (Swofford, 2002) with the following parameters: starting trees obtained by stepwise addition, TBR branch swapping, "MulTrees" turned on, and steepest descent not in effect. One thousand bootstrap replicates were conducted with PAUP* 4.0b10 using identical parameters. Decay indices were calculated in PAUP* 4.0b10 using

Table 1
Sample information

Sample	Collector	Source	GenBank Accession No.
Arabidopsis arenosa subsp. *arenosa* 1	Beck 802 (MO)	Germany, Berlin	**(9) AM293682 (11) AM293683**
A. arenosa subsp. *arenosa* 2	Beck 876 (MO)	Hungary, Heves Co.	**(1) AM293684 (2) AM293685 (4) AM293686**
A. cebennensis 1	Coste s.n. (BUC)	France	(3) AM293687
A. croatica 1	Beck 872 (MO)	Croatia, Zavižan Mountain	**(1) AM293688 (3) AM293689**
A. halleri subsp. *halleri* 1	O'Kane 3641 (MO)	Czech Republic, Nove Mesto	**(S1) AM293690** (L4) AM293691
A. halleri subsp. *gemmifera* 1	Murakami s.n.	Japan, Kyoto	(1) AM293692 (5) AM293693 (8) AM293694
A. halleri subsp. *gemmifera* 2	O'Kane 3687 (MO)	Japan, Shiga Prefecture	(10) AM293695 (12) AM293696
A. halleri subsp. *gemmifera* 3	O'Kane 3690 (MO)	Japan, Shiga Prefecture	(1) AM293697 (2) AM293698 (4) AM293699 (13) AM293700
A. halleri subsp. *ovirensis* 1	O'Kane 3656 (MO)	Poland, Tatra National Park	(10) AM293701 (14) AM293702
A. lyrata subsp. *lyrata* 1	Beck 811 (MO)	USA, Michigan	(1) AM293703 (2) AM293704
A. lyrata subsp. *lyrata* 2	Olds s.n.	USA, Massachusetts	(10) AM293705
A. lyrata subsp. *kamchatica* 1	Hsu 11052	Taiwan	(S1) AM293706 (L6) AM293707
A. lyrata subsp. *kamchatica* 2	O'Kane 3684 (MO)	Japan, Nikko Bot. Garden	(S1) AM293708 (S4) AM293709 (L1) AM293710 (L3) AM293711 (L11) AM293712
A. lyrata subsp. *petraea* 1	Koch s.n.	Germany, Plech	(1) AM293713
A. neglecta 1	O'Kane 3658 (MO)	Poland, Tatra National Park	**(3) AM293714 (5) AM293715 (7) AM293716** (12) AM293717 (17) AM293718
A. neglecta 2	O'Kane 3663a (MO)	Poland, Tatra National Park	**(5) AM293719** (14) AM293720
A. suecica 1	Schmuths s.n.	Finland, Helsinki	(S1) AM293721 **(L8) AM293722**
A. suecica 2	Beck 894 (MO)	Laboratory strain "Sue1"	(S1) AM293723 **(L1) AM293724 (L3) AM293725**
A. suecica 3	Mummenhof s.n. (H)	Finland	(S1) AM293726 (S2) AM293727 **(L1) AM293728 (L4) AM293729**
A. thaliana	None	Laboratory strain "Columbia"	NC_003071

Numbers in parentheses indicate sequences from individual clones, accession numbers in bold represent full ATM-1 to ATM-4 sequences.

a command file generated in MacClade 4.0 (Maddison and Maddison, 2000). The strict consensus tree was drawn using WinClada (Nixon, 2002).

In order to investigate evolutionary relationships within the *A. arenosa* group the entire ATM-1 to ATM-4 region was sequenced for two *A. arenosa* subsp. *arenosa* individuals, one of *A. croatica*, two of *A. neglecta*, and three of *A. suecica*. *A. halleri* subsp. *halleri* was used as an outgroup. Alignment and tree building were identical to the above genus-wide analysis.

The status of certain alternative topologies were assessed by comparing the shortest trees conforming to these topologies to the most parsimonious trees (MPTs) obtained in the original, unconstrained search. Alternative topologies were constructed in MacClade 4.0 (Maddison and Maddison, 2000) and were used to limit subsequent parsimony searches, instructing the algorithm to find the shortest trees conforming to the constraint. Parameters for constrained searches were otherwise identical to the unconstrained search. All MPT's for each constrained scenario were compared to each unconstrained MPT using a one-tailed Wilcoxon signed-ranks test (Templeton, 1983) employed in PAUP* 4.0b10. Although use of the Wilcoxon signed-ranks test for comparison of a posteriori (unconstrained MPT) and a priori (constrained MPT) topologies has been questioned (Goldman et al., 2000), this approach has been shown to be equally or more conservative relative to alternative tests (Melville et al., 2001; Townsend and Larson, 2002). Hypotheses embodied by these alternative topologies were rejected if trees conforming to them were significantly longer than any of the MPTs.

3. Results

3.1. Genus-wide analysis

Amplification produced one or two bands in all samples, and homology assessment was relatively straightforward (see below). The entire ATM-1 to ATM-4 region varied in size from 708 bp (*A. suecica* maternal copy) to approximately 1800 bp (*A. halleri* s.l.). This region is 712 bp in the *A. thaliana* genome sequence. Sequences of the analyzed ATM out3 to ATM-4 region varied from 353 bp (*A. suecica* paternal copy) to 419 bp in *A. halleri* subsp. *gemmifera*. This region is 365 bp in the *A. thaliana* genome sequence. The ATM out3 to ATM-4 region was easily alignable except in a few cases. The genus-wide ATM out3 to ATM-4 aligned data matrix consisted of 484 characters, including 15 insertion/deletion events. A total of 159 characters were excluded due to gaps or uncertain alignment. One haplotype from the long fragment of *A. lyrata* subsp. *kamchatica* sample 2 exhibited synapormorphies defining the *A. lyrata* s.l. clade (see below) at its 5′ end and synapomorphies defining the *A. halleri* s.l. clade at its 3′ end. Haplotypes unequivocally placed in these two clades were also present in this sample and the haplotype in question was viewed as a recombinant and excluded from the analysis. The analyzed matrix of 325 characters yielded 56 (17%) variable and 29 (9%) parsimony-informative characters. Eleven (73%) of the 15 insertion/deletion events were parsimony-informative. Each of 100 random addition replicates recovered the same island of 12 most parsimonious trees (MPTs) (length = 62, consistency index = 0.95, and retention index = 0.98). The strict consensus tree (Fig. 1) exhibited a moderately supported (62% bootstrap) clade comprising *A. lyrata* s.l., with *A. lyrata* subsp. *petraea* sister to a clade containing the remaining two subspecies. Also recovered was a strongly supported (97%) clade including *A. halleri* s.l. and haplotypes obtained from the longer of two fragments amplified in both samples of *A. lyrata* subsp. *kamchatica*. A moderately supported (68%) clade was recovered comprising *A. arenosa* subsp. *arenosa* and certain *A. neglecta* haplotypes. Also recovered was a moderately supported (63%) clade including certain *A. neglecta* haplotypes and haplotypes recovered from the longer of two fragments in all three *A. suecica* samples. *A. croatica* and *A. cebennensis* were part of a large basal polytomy formed by these two species and the clades mentioned above. Haplotypes from the shorter of two fragments in all three *A. suecica* samples were identical at all analyzed characters, and differed by one character (0.003%) from the *A. thaliana* genome sequence.

3.2. Arabidopsis arenosa group analysis

The *A. arenosa* group aligned data matrix consisted of 1158 characters, including 18 insertion/deletion events. A total of 404 characters were excluded due to gaps or uncertain alignment. The analyzed matrix of 754 characters yielded 69 (9%) variable and 48 (6%) parsimony-informative characters. Fourteen (78%) of the 18 insertion/deletion events were parsimony-informative. Ninety-seven of 100 random addition replicates recovered the same MPT (length = 82, consistency index = 0.90, and retention index = 0.95). The MPT is shown in Fig. 2. *A. croatica* was sister to the remainder of the ingroup, which comprised two well-supported clades. The first (96% bootstrap) included all haplotypes recovered from *A. arenosa* subsp. *arenosa* and a haplotype from *A. neglecta* sample 1. It should be noted that a haplotype exhibiting synapomorphies defining this clade was also recovered from *A. neglecta* sample 2, but was not included in the analysis due to a stretch of poor sequence. The second clade (100% bootstrap) comprised haplotypes from both *A. neglecta* samples and haplotypes recovered from the long fragment in all *A. suecica* samples.

4. Discussion

4.1. Homology assessment

Amplification using the ATM-1 and ATM-4 primers produced single fragments in nine samples. All samples of the hybrid species *A. suecica* exhibited double fragments,

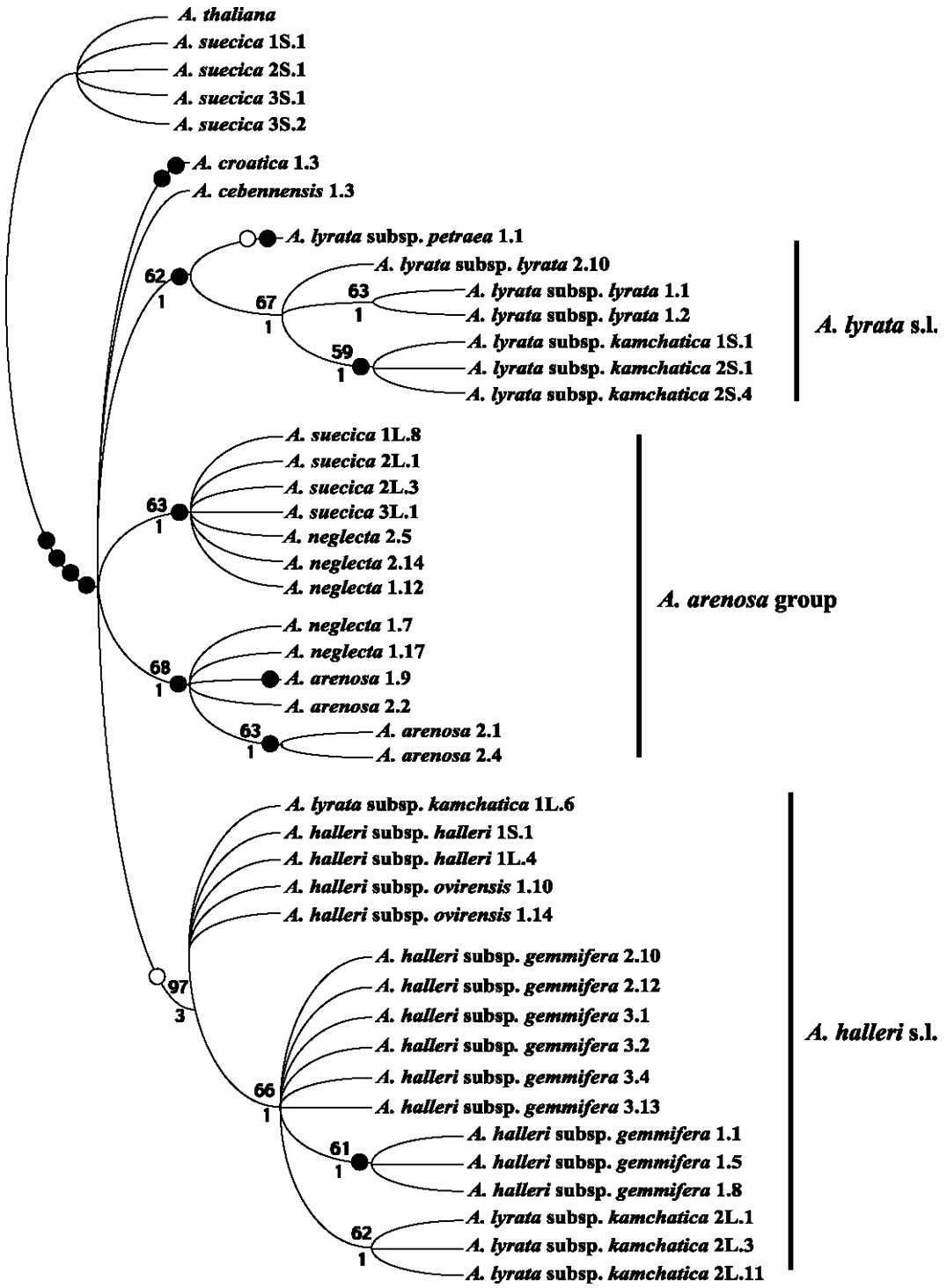

Fig. 1. Strict consensus of 12 MPTs recovered in the genus-wide analysis. Bootstrap values appear above nodes, decay values below. Filled circles indicate unique insertion/deletion events, open circles indicate insertion/deletion events that have occurred in parallel. Codes following taxon names indicate sample number, short or long fragment (where applicable), and clone number. For example, "*A. suecica* 2S.1" refers to the first clone from the short fragment of *A. suecica* sample number two.

the smaller of which corresponded to that of the maternal parent *A. thaliana*, the larger of which corresponded to the paternal parent (discussed below). Both samples of *A. lyrata* subsp. *kamchatica* exhibited double fragments, the smaller of which were part of a clade with the remaining sequences from *A. lyrata* s.l. and the larger of which were part of a clade with sequences from *A. halleri* s.l. (Fig. 1). The presence of two fragments is likely due to hybridization between *A. lyrata* subsp. *kamchatica* and a member of *A. halleri* s.l. (discussed below). *Arabidopsis lyrata* subsp. *lyrata* (samples 1 and 2), *A. lyrata* subsp. *petraea*, and *A. arenosa* subsp. *arenosa* sample 2 exhibited minor

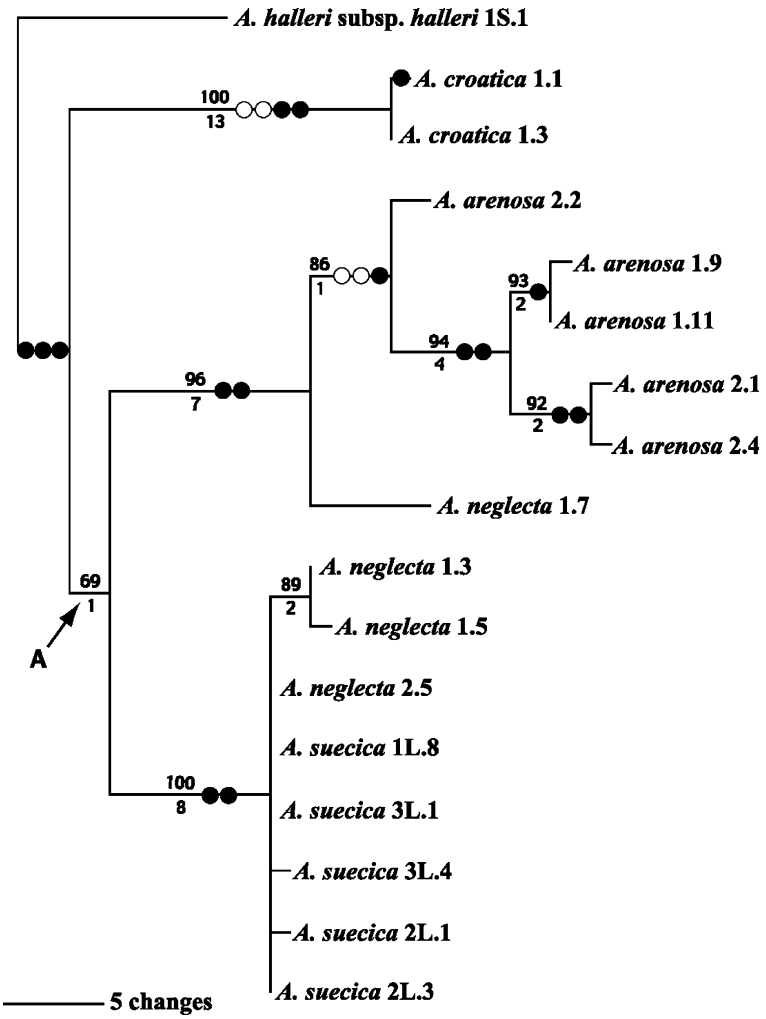

Fig. 2. Phylogram of the single MPT from the *A. arenosa* group analysis. Bootstrap values appear above nodes, decay values below. Filled circles indicate unique insertion/deletion events, open circles indicate insertion/deletion events that have occurred in parallel. Codes following taxon names indicate sample number, short or long fragment (where applicable), and clone number. For example, "*A. suecica* 1L.8" refers to the eighth clone from the long fragment of *A. suecica* sample number one. Clade A is discussed in the text.

fragments that were of considerably different size than the main fragment exhibiting typical clones. Sequences from these minor fragments were unalignable over a large portion of the sequence, and were, therefore, deemed non-homologous and were not further analyzed. A fragment of approximately 1800 bp was amplified in all five samples from *A. halleri* s.l., while a shorter fragment (1084 bp) and a faint, considerably longer fragment (>2.5 kb) were also amplified in *A. halleri* subsp. *halleri* and *A. halleri* subsp. *gemmifera* sample 3, respectively. The 1084 bp fragment in *A. halleri* subsp. *halleri* was completely alignable across its entire length. The long fragment in *A. halleri* subsp. *gemmifera* sample 3 was too weak for cycle sequencing. The typical (1800 bp) fragment was completely alignable over the analyzed ATM out3 to ATM-4 region in all samples, but was unalignable approximately 250 bp into the 5′ end of the entire ATM-1 to ATM-4 region in the two (*A. halleri* subsp. *halleri* sample 1 and *A. halleri* subsp. *gemmifera* sample 3) samples for which longer sequencing was performed. Sequences from this typical fragment were included because sequences from the typical fragment in *A. halleri* subsp. *halleri* were identical or differed by a single bp from the clearly homologous 1084 bp fragment of the same sample. This similarity and the common ancestry of the *A. halleri* subsp. *gemmifera* sequences (rendered paraphyletic by the sequences likely introgressed into *A. lyrata* subsp. *kamchatica*) make it unlikely that the 1800 bp fragment is a recently duplicated, paralogous locus.

4.2. Phylogeny of Arabidopsis

The *Atmyb2* flanking sequence phylogeny presented here (Fig. 1) is almost entirely concordant with the ITS phylogeny presented by O'Kane and Al-Shehbaz (2003). Both phylogenies include clades corresponding to *A. lyrata* s.l., *A. halleri* s.l., and a large basal polytomy. The ITS phylogeny exhibits a clade corresponding to the *A. arenosa* group, which is consistent with the morphological predictions of O'Kane and Al-Shehbaz (1997). Although *A. croatica* and two clades containing *A. arenosa* subsp. *arenosa* and

A. neglecta were part of the large basal polytomy in this study, a clade uniting all three elements was recovered in eight of the 12 MPTs, suggesting that the hypothesized relationship between these three species is a good working hypothesis. The failure to recover deeper relationships (those between *A. cebennensis*, *A. halleri* s.l., *A. lyrata* s.l., and the *A. arenosa* group) in both studies could be due to a "hard" polytomy resulting from simultaneous or rapidly successive cladogenesis, or alternatively from substitutional saturation or a lack of informative characters (Jackman et al., 1999). The overall lack of homoplasy (consistency index = 0.95) and the consistent distribution of characters along branches in trees resulting from two rather different DNA sequences (*Atmyb2*, which is a single copy gene flanking sequence and ITS, which is a tandemly repeated intergenic spacer) suggests that rapid lineage diversification early in the history of this group is a distinct possibility. Future work should evaluate this hypothesis by employing additional non-coding nuclear sequences, which have been shown to provide substantially more phylogenetic information relative to ITS (Howarth and Baum, 2005; Small et al., 1998).

4.3. The paternal parent of A. suecica

The *Atmyb2* flanking sequence was consistent with previous studies that identify *A. thaliana* as one of the parental species and that indicate the hybridization event was relatively recent. Säll et al. (2003) surveyed over 4.2 kb of chloroplast sequence in 48 *A. suecica* samples, finding only a single indel polymorphic among them. Likewise, Jakobsson et al. (2006) sequenced four nuclear fragments (over 1800 bp in total) and found that at each fragment putative maternal or paternal *A. suecica* sequences were invariant and putative maternal sequences were always identical to one of the observed *A. thaliana* sequences. We also find low sequence variation within *A. suecica*, consistent with a recent origin of the hybrid taxon. The entire ATM-1 to ATM-4 region is 840 aligned bp in our putative paternal *A. suecica* haplotypes (10 clones from three samples) and only two singleton substitutions and variation at a mono- and a tri-nucleotide repeat were observed within this region. The ATM-1 to ATM-4 region is 712 aligned bp in our putative maternal (from *A. thaliana*) *A. suecica* haplotypes (nine clones from three species), and only a single substitution and one instance of length variation at a di-nucleotide repeat were observed within this region. Interestingly, the most frequent putative maternal haplotype found within *A. suecica* is identical to one of the 20 *A. thaliana* haplotypes recovered in a global survey of *Atmyb2* sequence variation (J. Beck, unpublished data). This haplotype is common from approximately 20° E longitude to the eastern limits of *A. thaliana*'s range in Central Asia. The range of this putative maternal donor haplotype overlaps with the range of putative maternal donor genotypes identified in Jakobsson et al. (2006) in Eastern Europe, highlighting this area as the potential site of origin for *A. suecica*. Regarding the paternal parent, patterns of relatedness among haplotypes in the *A. arenosa* group suggest that *A. neglecta* is the more likely paternal parent of *A. suecica*. In both topologies (Figs. 1 and 2) all putative paternal haplotypes from *A. suecica* are most closely related to certain *A. neglecta* haplotypes. One *A. neglecta* clone differed from an *A. suecica* clone by only a single tri-nucleotide repeat. Another strongly supported clade contains the remaining *A. neglecta* haplotypes and all *A. arenosa* subsp. *arenosa* haplotypes. Our data strongly reject the hypothesis that putative paternal *A. suecica* haplotypes share a most recent common ancestor with *A. arenosa* subsp. *arenosa* haplotypes, as trees constrained to exhibit a clade comprising only these haplotypes were significantly longer than the MPT (length = 99, $P < 0.0001$). These results are consistent with those of Jakobsson et al. (2006), in which haplotypes from an *A. arenosa* subsp. *arenosa* sample differed from *A. suecica* putative paternal haplotypes (eight *A. suecica* samples) by approximately 0.9, 1.0, and 2.0% at three loci. Given the recent origin of *A. suecica*, haplotype identity or near-identity at many loci would be expected if their *A. arenosa* subsp. *arenosa* sample was closely related to the paternal donor.

Deep divergences between haplotypes observed in the same individual (*A. neglecta* sample 1, *A. arenosa* subsp. *arenosa* sample 2) could be due to comparing divergent alleles at the same locus or haplotypes at recently duplicated, and thus paralogous, loci. A gene duplication event on the stem lineage leading to clade "A" (Fig. 2) is unlikely, as *A. arenosa* subsp. *arenosa* and *A. suecica* only exhibit haplotypes from one of the two daughter clades. It is more reasonable to assume that haplotypes at a single *Atmyb2* locus are still undergoing sorting in highly variable and poorly defined *A. neglecta* and *A. arenosa* lineages. Previous studies of *A. arenosa* s.l. and *A. neglecta* strongly support this view. *A. arenosa* is a morphologically variable, widespread species common in Europe from central France east to approximately 30° E longitude (Hoffman, 2005; O'Kane and Al-Shehbaz, 1997). Měsíček (1970) investigated chromosome number and morphology only in Czechoslovakian populations of *A. arenosa* s.l., and identified diploid ($2n = 16$), triploid ($2n = 24$), tetraploid ($2n = 32$), pentaploid ($2n = 39,40$), and aneuploid ($2n = 18, 19, 30, 31, 34$) individuals. Měsíček was also able to successfully cross *A. arenosa* (subspecies unclear) to the diploid ($2n = 16$) cytotype of *A. neglecta*, which also occurs as a tetraploid. Finally, in a recent RAPD study Lind-Halldén et al. (2002) found *A. arenosa* subsp. *arenosa* to be a relatively variable species at the genome level, exhibiting substantially more intraspecific variation when compared to *A. suecica* or *A. thaliana*. Indeed, rather than viewing the paternal parent of *A. suecica* as either *A. arenosa* s.l. or *A. neglecta*, it may be better to view this parent as a particular genotype within a larger syngameon. A dense, range-wide phylogeographic and cytological study of all three species is needed to evaluate the status of the *A. arenosa* s.l. and *A. neglecta* lineages, and the proposition that the paternal genome of *A. suecica* came from *A. neglecta*.

4.4. Hybridization between A. halleri s.l. and A. lyrata s.l.

The presence of haplotypes from both samples of *A. lyrata* subsp. *kamchatica* within the well-supported *A. halleri* s.l. clade is strong evidence of hybridization involving members of these two taxa, which is consistent with the *Chs* data of Shimizu et al. (2005). Our data strongly support this placement, as trees constrained to exhibit a clade consisting of all *A. lyrata* s.l. sequences were significantly longer than any of the MPTs (length = 66, $P = 0.023$). The hybridization scenario is also supported by the presence in each *A. lyrata* subsp. *kamchatica* sample of "correct" haplotypes clearly descended from the common ancestor of *A. lyrata* s.l. The presence of these multiple, and therefore paralogous (Wendel and Doyle, 1998), loci would not be expected under a scenario of widespread sharing of ancestral polymorphisms. The clade comprised of *A. lyrata* subsp. *kamchatica* sample 2 haplotypes and all haplotypes from *A. halleri* subsp. *gemmifera* indicates hybridization between these two subspecies, although haplotypes from *A. lyrata* subsp. *kamchatica* sample 1 are unresolved within the larger *A. halleri* s.l. clade, suggesting multiple hybridization events, possibly with another *A. halleri* subspecies. We did not detect evidence of hybridization between *A. halleri* subsp. *halleri* and *A. lyrata* subsp. *petraea* as was reported in Ramos-Onsins et al. (2004), although our study only included one individual from each taxon. These two separate hybridization events are certainly plausible biogeographically, as the ranges of *A. lyrata* subsp. *kamchatica* and *A. halleri* subsp. *gemmifera* overlap in east Asia and the ranges of *A. halleri* subsp. *halleri* and *A. lyrata* subsp. *petraea* overlap in central and eastern Europe (Hoffman, 2005). Furthermore, the observation of polyploidy in *A. lyrata* subsp. *petraea* and *A. lyrata* subsp. *kamchatica* is consistent with the hybridization scenario. Although only diploids ($2n = 16$) are known from the three subspecies of *A. halleri* s.l. and *A. lyrata* subsp. *lyrata* (Al-Shehbaz and O'Kane, 2002; Kolník and Marhold, 2006), both diploids and tetraploids ($2n = 32$) are known from *A. lyrata* subsp. *petraea* and *A. lyrata* subsp. *kamchatica* (Al-Shehbaz and O'Kane, 2002; Dart et al., 2004). The observation of both ploidy levels in *A. lyrata* subsp. *kamchatica* calls for a range-wide phylogeographic study of this species and taxa potentially involved in the hybridization event to determine both the genotypes involved and the number if times hybridization occurred. Taken together, these hybridization events indicate rather weak reproductive barriers between these two species complexes, at least historically. Weak reproductive barriers are a common feature in *Arabidopsis* as a whole, considering the documented examples of natural and artificial hybridization. The relatively recent hybridization between *A. thaliana* ($2n = 10$) and a member of the *A. arenosa* group ($2n = 16$) to form *A. suecica* ($2n = 26$) took place between taxa with seemingly incompatible cytotypes that diverged approximately 5 mya (Koch et al., 2000). Artificial hybridization is also common, with many successful laboratory crosses between major lineages within the genus (Měsíček, 1970; reviewed in Nasrallah et al., 2000). These low reproductive barriers and the wealth of genetic tools available for *Arabidopsis* should provide future researchers with the framework for investigating the genetic, ecological, and evolutionary consequences of hybridization.

Acknowledgments

The authors thank Doug Schemske and Heike Schmuths for sharing seeds; Yu-Chung Chiang and Hui-Fen Kuo for sharing DNA samples; the Ministry of Culture of the Republic of Croatia for permission to collect *A. croatica*; and the Graduate Studies Program at the Missouri Botanical Garden for logistical support. This study was funded by the National Science Foundation DDIG DEB-0411875 (J.B.B.) and the Division of Biology and Biomedical Sciences at WU.

References

Al-Shehbaz, I.A., O'Kane Jr., S.L., 2002. Taxonomy and phylogeny of *Arabidopsis* (Brassicaceae) (22 August, 2002). In: Somerville, C.R., Meyerowitz, E.M. (Eds.), The *Arabidopsis* Book. American Society of Plant Biologists, Rockville, MD, doi: 10.1199/tab.0009, <www.aspb.org/publications/arabidopsis/>.

Beilstein, M.A., Al-Shehbaz, I.A., Kellogg, E.A., 2006. Brassicaceae phylogeny and trichome evolution. Am. J. Bot. 93, 607–619.

Boffelli, D., McAuliffe, J., Ovcharenko, D., Lewis, K.D., Ovcharenko, I., Pachter, L., Rubin, E.M., 2003. Phylogenetic shadowing of primate sequences to find functional regions of the human genome. Science 299, 1391–1394.

Caicedo, A.L., Stinchcombe, J.R., Olsen, K.M., Schmitt, J., Purugganan, M., 2004. Epistatic interaction between *Arabidopsis FRI* and *FLC* flowering time genes generates a latitudinal cline in a life history trait. Proc. Natl. Acad. Sci. USA 101, 15670–15675.

Callahan, H.S., Pigliucci, M., 2002. Shade-induced plasticity and its ecological significance in wild populations of *Arabidopsis thaliana*. Ecology 83, 1965–1980.

Dart, S., Kron, P., Mable, B.K., 2004. Characterizing polyploidy in *Arabidopsis lyrata* using chromosome counts and flow cytometry. Can. J. Bot. 82, 185–197.

Edwards, K.D., Lynn, J.R., Gyula, P., Nagy, F., Millar, A.J., 2005. Natural allelic variation in the temperature-compensation mechanism of the *Arabidopsis thaliana* circadian clock. Genetics 170, 387–400.

Goldman, N., Anderson, J.P., Rodrigo, A.G., 2000. Likelihood-based tests of topologies in phylogenetics. Syst. Biol. 49, 652–670.

Griffith, C., Kim, E., Donohue, K., 2004. Life-history variation and adaptation in the historically mobile plant *Arabidopsis thaliana* (Brassicaceae) in North America. Am. J. Bot. 91, 837–849.

Hoffman, M.H., 2005. Evolution of the realized climate niche in the genus *Arabidopsis* (Brassicaceae). Evolution 59, 1425–1436.

Hong, R.L., Hamaguchi, L., Busch, M.A., Weigel, D., 2003. Regulatory elements of the floral homeotic gene *AGAMOUS* identified by phylogenetic footprinting and shadowing. Plant Cell 15, 1296–1309.

Howarth, D.G., Baum, D.A., 2005. Genealogical evidence of homoploid hybrid speciation in an adaptive radiation of *Scaevola* (Goodeniaceae) in the Hawaiian Islands. Evolution 59, 948–961.

Jackman, T.R., Larson, A., De Queiroz, K., Losos, J.B., 1999. Phylogenetic relationships and tempo of early diversification in *Anolis* lizards. Syst. Biol. 48, 254–285.

Jakobsson, M., Hagenblad, J., Tavaré, S., Säll, T., Halldén, C., Lind-Halldén, C., Nordborg, M., 2006. A unique recent origin of the

allotetraploid species *Arabidopsis suecica*: evidence from nuclear DNA markers. Mol. Biol. Evol. 23, 1217–1231.

Kamiya, T., Kawabe, A., Miyashita, N.T., 2002. Nucleotide polymorphism at the *Atmyb2* locus of the wild plant *Arabidopsis thaliana*. Genet. Res. 80, 89–98.

Koch, M., Bishop, J., Mitchell-Olds, T., 1999. Molecular systematics and evolution of *Arabidopsis* and *Arabis*. Plant Biol. 1, 529–537.

Koch, M., Haubold, B., Mitchell-Olds, T., 2000. Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis*, *Arabis*, and related genera. Mol. Biol. Evol. 17, 1483–1498.

Koch, M., Haubold, B., Mitchell-Olds, T., 2001a. Molecular systematics of the Brassicaceae: evidence from coding plastidic *matK* and nuclear *Chs* sequences. Am. J. Bot. 88, 534–544.

Koch, M., Weisshaar, B., Kroymann, J., Haubold, B., Mitchell-Olds, T., 2001b. Comparative regulatory evolution: conservation and function of the *Chs* and *Apetala3* promoters. Mol. Biol. Evol. 18, 1882–1891.

Kolník, M., Marhold, K., 2006. Distribution, chromosome numbers and nomenclature conspect of *Arabidopsis halleri* (Brassicaceae) in the Carpathians. Biologia Bratislava 61, 41–50.

Lind-Halldén, C., Halldén, C., Säll, T., 2002. Genetic variation in *Arabidopsis suecica* and it parental species *A. arenosa* and *A. thaliana*. Hereditas 136, 45–50.

Macnair, M.R., Bert, V., Huitson, S.B., Saumitou-Laprade, P., Petit, D., 1999. Zinc tolerance and hyperaccumulation are genetically independent characters. Proc. R. Soc. Lond. B 266, 2175–2179.

Maddison, D.R., Maddison, W.P., 2000. MacClade 4: Analysis of phylogeny and character evolution. v. 4.0, Sinauer Associates, Sunderland, MA.

Mauricio, R., 1998. Costs of resistance to natural enemies in field populations of the annual plant *Arabidopsis thaliana*. Am. Nat. 151, 20–28.

Melville, J., Schulte, J.A.I., Larson, A., 2001. A molecular phylogenetic study of ecological diversification in the Australian lizard genus *Ctenophorus*. J. Exp. Zool. 291, 339–353.

Měsíček, J., 1970. Chromosome counts in *Cardaminopsis arenosa* Agg (Cruciferae). Preslia 42, 225–248.

Mitchell-Olds, T., 2001. *Arabidopsis thaliana* and its wild relatives: a model system for ecology and evolution. Trends Ecol. Evol. 16, 693–700.

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

Nasrallah, M., Yogeeswaran, K., Snyder, S., Nasrallah, J.B., 2000. *Arabidopsis* species hybrids in the study of species differences and evolution of amphiploidy in plants. Plant Physiol. 124, 1605–1614.

Nasrallah, M.E., Liu, P., Sherman-Broyles, S., Boggs, N.A., Nasrallah, J.B., 2004. Natural variation in expression of self-incompatibility in *Arabidopsis thaliana*: implications for the evolution of selfing. Proc. Natl. Acad. Sci. USA 101, 16070–16074.

Nixon, K., 1999–2002. WinClada ver. 1.00.08. Ithaca, NY. Distributed by the author.

Nordborg, M., Borevitz, J.O., Bergelson, J., Berry, C.C., Chory, J., Hagenblad, J., Kreitman, M., Maloof, J.N., Noyes, T., Oefner, P.J., Stahl, E., Weigel, D., 2002. The extent of linkage disequilibrium in *Arabidopsis thaliana*. Nat. Genet. 30, 190–193.

O'Kane Jr., S.L., Al-Shehbaz, I.A., 1997. A synopsis of *Arabidopsis* (Brassicaceae). Novon 7, 323–327.

O'Kane Jr., S.L., Al-Shehbaz, I.A., 2003. Phylogenetic position and generic limits of *Arabidopsis* (Brassicaceae) based on sequences of nuclear ribosomal DNA. Ann. Mo. Bot. Gard 90, 603–612.

O'Kane Jr., S.L., Schaal, B.A., Al-Shehbaz, I.A., 1997. The origins of *Arabidopsis suecica* (Brassicaceae) as indicated by nuclear rDNA sequences. Syst. Bot. 21, 559–566.

Price, R.A., Palmer, J.D., Al-Shehbaz, I.A., 1994. Systematic relationships of *Arabidopsis*: a molecular and morphological perspective. In: Somerville, C.R., Meyerowitz, E.M. (Eds.), Arabidopsis. Cold Spring Harbor Laboratory Press, New York, pp. 7–19.

Quesada, V., García-Martínez, S., Piqueras, P., Ponce, M.R., Micol, J.L., 2002. Genetic architecture of NaCl tolerance in *Arabidopsis*. Plant Physiol. 130, 951–963.

Rambaut, A., 1996. Se-Al: Sequence Alignment Editor, version 2.0a11. <http://evolve.zoo.ox.ac.uk/>.

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

Säll, T., Jakobsson, M., Lind-Halldén, C., Halldén, C., 2003. Chloroplast DNA indicates a single origin of the allotetraploid *Arabidopsis suecica*. J. Evol. Biol. 16, 1019–1029.

Schmid, K.J., Ramos-Onsins, S., Ringys-Beckstein, H., Weisshaar, B., Mitchell-Olds, T., 2005. A multilocus sequence survey in *Arabidopsis thaliana* reveals a genome-wide departure from a neutral model of DNA sequence polymorphism. Genetics 169, 1601–1615.

Shimizu, K.K., Cork, J.M., Caicedo, A.L., Mays, C.A., Moore, R.C., Olsen, K.M., Ruzsa, S., Coop, G., Bustamante, C.D., Awadalla, P., Purugganan, M., 2004. Darwinian selection on a selfing locus. Proc. Natl. Acad. Sci. USA 306, 2081–2084.

Shimizu, K.K., Fujii, S., Marhold, K., Watanabe, K., Kudoh, H., 2005. *Arabidopsis kamchatica* (Fisch. ex DC.) K. Shimizu & Kudoh and *A. kamchatica* subsp. *kawasakiana* (Makino) K. Shimizu & Kudoh, new combinations. Acta Phytotax. Geobot. 56, 163–172.

Simmons, M.P., Ochoterena, H., 2000. Gaps as characters in sequence-based phylogenetic analyses. Syst. Biol. 49, 369–381.

Small, R.L., Ryburn, J.A., Cronn, R.C., Seelanan, T., Wendel, J.F., 1998. The tortise and the hare: choosing between noncoding plastome and nuclear *Adh* sequences for phylogeny reconstruction in a recently diverged plant group. Am. J. Bot. 85, 1301–1315.

Swofford, D.L., 2002. PAUP*. Phylogenetic Analysis Using Parsimony (* and Other Methods). v. 4.0b10, Sinauer Associates, Sunderland, MA.

Templeton, A.R., 1983. Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and the apes. Evolution 37, 221–244.

Tian, D., Araki, H., Stahl, E., Bergelson, J., Krietman, M., 2002. Signature of balancing selection in *Arabidopsis*. Proc. Natl. Acad. Sci. USA 99, 11525–11530.

Townsend, T., Larson, A., 2002. Molecular phylogenetics and mitochondrial genomic evolution in the Chamaeleonidae (Reptilia, Squamata). Mol. Phylogenet. Evol. 23, 22–36.

Urao, T., Yamaguchi-Shinozaki, K., Urao, S., Shinozaki, K., 1993. An *Arabidopsis myb* homolog is induced by dehydration stress and its gene product binds to the conserved MYB recognition sequence. Plant Cell 5, 1529–1539.

Wendel, J., Doyle, J.J., 1998. Phylogenetic incongruence: window into genome history and molecular evolution. In: Soltis, D.E., Soltis, P.S., Doyle, J.J. (Eds.), Molecular Systematics of Plants II. Kluwer, Dordrecht, pp. 265–296.