The allopolyploid Arabidopsis kamchatica originated from multiple individuals of Arabidopsis lyrata and Arabidopsis halleri

Shimizu-Inatsugi, Rie ; Lihova, Judita ; Iwanaga, Hiroko ; Kudoh, Hiroshi ; Marhold, Karol ; Savolainen, Outi ; Watanabe, Kuniaki ;…

(Citation)
Molecular Ecology,18(19):4024-4048

(Issue Date)
2009-10

(Resource Type)
journal article

(Version)
Accepted Manuscript

The allopolyploid *Arabidopsis kamchatica* originated from multiple individuals of *A. lyrata* and *A. halleri*

Rie Shimizu-Inatsugi*†‡, Judita Lihová§, Hiroko Iwanaga*†‡, Hiroshi Kudoh†‡, Karol Marhold§¶, Outi Savolainen||, Kuniaki Watanabe*‡, Valentin V. Yakubov** and Kentaro K. Shimizu*

*Institute of Plant Biology, University of Zurich, Zollikerstrasse 107, Zurich, CH-8008, Switzerland

†Center for Ecological Research, Kyoto University, Hirano 2-509-3, Otsu 520-2113, Japan

‡Department of Biology, Graduate School of Science, Kobe University, Nada, Kobe 657-8501, Japan

§Institute of Botany, Slovak Academy of Sciences, Dúbravská cesta 14, SK-845 23 Bratislava, Slovak Republic

¶Department of Botany, Charles University, Benátská 2, CZ-128 01 Praha 2, Czech Republic

||Department of Biology and Biocenter Oulu, University of Oulu, P.O. Box 3000, 90014, Oulu, Finland

**Institute of Biology and Soil Science, Far East Branch of the Russia Academy of Sciences, Prospect Stoletiya, 159, Vladivostok, 690022 Russia

Keywords: *Arabidopsis kamchatica* subspecies *kawasakiana,* independent origins, plant mating systems, Pleistocene glacial refugia, population genetics empirical, speciation

Corresponding author: Kentaro K. Shimizu
Institute of Plant Biology
University of Zurich
Zollikerstrasse 107, CH-8008 Zurich, Switzerland
Phone +41-44-634-8247, 8244
FAX +41-44-634-8204
shimizu@botinst.uzh.ch

Running title: Independent origins of *Arabidopsis kamchatica*

Manuscript information: 36 pages, 9 figures, 1 table
Supporting material: 5 figures, 3 tables, 1 text
Data deposition: Sequences are deposited in GenBank

Abstract

Polyploidization, or genome duplication, has played a critical role in the diversification of animals, fungi, and plants. Little is known about the population structure and multiple origins of polyploid species because of the difficulty identifying multiple homeologous nuclear genes. The allotetraploid species *Arabidopsis kamchatica* is closely related to the model species *A. thaliana* and is distributed in a broader climatic niche than its parental species. Here, we performed direct sequencing of homeologous pairs of the low-copy nuclear genes *WER* and *CHS* by designing homeolog-specific primers, and obtained also chloroplast and ribosomal internal transcribed spacer (ITS) sequences. Phylogenetic analysis showed that 50 individuals covering the distribution range including North America are allopolyploids derived from *A. lyrata* and *A. halleri*. Three major clusters within *A. kamchatica* were detected using Bayesian clustering. One cluster has widespread distribution. The other two are restricted to the southern part of the distribution range including Japan, where the parent *A. lyrata* is not currently distributed. This suggests that the mountains in Central Honshu and surrounding areas in Japan served as refugia during glacial–interglacial cycles and retained the diversity. We also found that multiple haplotypes of nuclear and chloroplast sequences of *A. kamchatica* are identical to those of their parental species. This indicates that multiple diploid individuals contributed to the origin of *A. kamchatica*. The haplotypes of low-copy nuclear genes in Japan suggest independent polyploidization events rather than introgression. Our findings suggest that self-compatibility and gene silencing occurred independently in different origins.

61 **Introduction**

62

63 Allopolyploidization, or genome-wide duplication with hybridization, has played an
64 important role in evolution and diversification in plants (Stebbins 1950, 1971; Ohno
65 1970; Levin 2002; Comai 2005; Marhold & Lihová 2006; Otto 2007). It was suggested
66 that nearly all angiosperms experienced polyploidization in the past and that 57% to
67 70% of them experienced polyploidy relatively recently (Otto 2007). Despite the
68 prevalence of polyploids, the identification of the parental species of polyploids has
69 been difficult. Common markers used in molecular phylogenetic studies include
70 chloroplast DNA (cpDNA) sequences with uniparental inheritance and nuclear
71 ribosomal internal transcribed spacer (ITS) sequences, which often retain only one
72 parental unit because of concerted evolution. When the cpDNA and ITS are derived
73 from different species, the incongruence can help identify the parents of an
74 allotetraploid. However, when concerted evolution of ITS has resulted in maintenance
75 of the haplotype from the same parent as that of cpDNA, only one of the parents can be
76 identified. In addition, it is technically difficult to separate homeologs derived from
77 multiple parental species because of their high degree of similarity with each other.
78 Cloning of PCR products would result in false sequences because of PCR errors and
79 artificial recombination between alleles and homeologs (Cronn *et al.* 2002; Lihová *et al.*
80 2006). To solve these problems, it is necessary to design homeolog-specific primers for
81 'low-copy nuclear genes'. Here, genes of a polyploid that are orthologous to a single-
82 copy nuclear gene in the parental species are referred to as low-copy nuclear genes, and
83 each copy is referred to as a homeolog.

84 We recently suggested that *Arabidopsis kamchatica* had an allotetraploid origin derived

85 from *A. lyrata* and *A. halleri* (Shimizu *et al.* 2005). *Arabidopsis kamchatica* has been
86 exploited as a model species to study the evolution of polyploids by using the extensive
87 genomic and genetic resources of *A. thaliana* and the ongoing whole genome
88 sequencing of its putative parent, *A. lyrata* (http://genome.jgi-
89 psf.org/Araly1/Araly1.home.html) (Shimizu 2002; Shimizu & Purugganan 2005). Self-
90 compatibility (Dart *et al.* 2004; Mable *et al.* 2004; Sugisaka & Kudoh 2008), flowering
91 time (under the nomenclature as *A. lyrata* from Alaska, Kuittinen *et al.* 2008), and the
92 epigenetically regulated *FWA* gene (Fujimoto *et al.* 2008) of *A. kamchatica* have been
93 studied recently. *Arabidopsis kamchatica* is distributed in East Asia and North America.
94 The overlap of *A. kamchatica*, *A. lyrata*, and *A. halleri* is limited only to Far East Russia
95 (Shimizu *et al.* 2005). While we suggested an allopolyploid origin based on a low-copy
96 nuclear gene in two individuals from Japan (Shimizu *et al.* 2005), an autopolyploid
97 origin of North American individuals was proposed based on cpDNA and ITS sequences
98 (Koch & Matschinger 2007).

99 Soltis *et al.* (2003) emphasized that recurrent origin of polyploid species is the
100 rule rather than the exception. The independent origin cases allow one to examine the
101 repeatability of evolution (Adams & Wendel 2005). Geographically independent origins
102 (or polytopic origins) have been documented in a few polyploid species that appeared
103 very recently. In *Tragopogon*, multiple origins during the 20th century were suggested
104 by a concordance between geographic variation patterns of the diploids and polyploids
105 (Tate *et al.* 2006). Levin (2002) noted that such a concordance is the best evidence for
106 multiple independent origins. However, unless the origin was very recent, geographic
107 variation may not be useful to identify independent origins of most polyploid species,
108 including *A. kamchatica*, because previous studies have suggested that the current

ranges of hybrids and parental species are poor predictors of the site of hybridization and that polyploid species tend to expand their distribution range by shifting to new ecological niches (Anderson & Stebbins 1954; Watanabe & Yahara 1984; Levin 2002; Beck *et al.* 2008). Apart from geographic data, independent origins have been supported strongly by multiple haplotypes (or alleles) shared by polyploid and parental species. The sharing of a polyploid with more than one chloroplast haplotype with a parental species, or more than two haplotypes of a nuclear homeolog indicates that multiple haplotypes of parental species contributed to the polyploid species, and suggests the independent origins of the polyploid species. A number of studies have shown the sharing of multiple polymorphic markers such as isozyme and cpDNA among polyploid and parental species, and have suggested that independent origins are widespread (reviewed by Soltis & Soltis 1993, 1999). However, it has also been suggested that haplotype sharing with parental species can result from both independent polyploidization events and introgression from diploid parental species (Ramsey & Schemske 1998; Husband 2004).

In this study, we address the population structure and polyploid origin of *A. kamchatica* by examining multiple populations distributed across its distribution range. In addition to cpDNA and ITS regions, we sequenced homeologous pairs of low-copy nuclear genes. We focused on the contribution of multiple parental individuals rather than the geographically independent origins. We discuss scenarios of independent polyploidization events vs. introgression based on low-copy nuclear genes.

Materials and Methods

Sampling

For *A. kamchatica*, two subspecies are recognized based on morphology, life history, and habitats. The first subspecies, *A. kamchatica* subsp. *kamchatica*, is a perennial, described originally from Kamchatka, Russia. It is reported from East Asia (Far East Russia, China, Korea, Japan, and Taiwan) and North America (Alaska, Canada, and Pacific Northwest of the United States). The second subspecies, *A. kamchatica* subsp. *kawasakiana*, is an annual found in sandy open habitats along seashores or lakeshores in lowlands in western Japan. Tetraploid chromosome number counts ($2n = 32$ and $n = 16_{II}$) were reported from samples in Japan, Far East Russia, Alaska, and Canada, and represent both subspecies (see references in Mulligan 1995; Shimizu *et al.* 2005; Warwick & Al-Shehbaz 2006). *Arabidopsis kamchatica* is morphologically similar to *A. lyrata*, and the taxon has been treated either as an infraspecific taxon of *A. lyrata* (O'Kane & Al-Shehbaz 1997) or as a distinct species (see references in Mulligan 1995; Shimizu *et al.* 2005).

Altogether 45 populations of the tetraploid *A. kamchatica* (both subspecies) were sampled, including one or two individuals per locality, giving a total of 50 individuals (Table 1). The sample locations ranged from the southwestern (Taiwan) to the northeastern (Alaska and Washington state in the USA, Canada) areas of the species range. The emphasis was on Kamchatka, from which *A. kamchatica* was described originally, and on Japan, where *A. kamchatica* subsp. *kawasakiana* has been recognized (see Fig. 1, Table 1) (Mulligan 1995; Shimizu *et al.* 2005). To facilitate the reference to the areas sampled, they are denoted by letters A–I (Fig. 1, Table 1).

For the potential parental species, diploids *A. lyrata* and *A. halleri,* we collected at least one sample from each subspecies described by O'Kane and Al-Shehbaz (1997) and Kolník and Marhold (2006). Within *A. halleri*, subspecies *gemmifera* is distributed in Eastern Asia, and the other four subspecies occur in Europe. Within *A. lyrata*, subspecies *petraea* is reported from Eurasia and subspecies *lyrata* from North America. Thus, the current distribution of *A. kamchatica* overlaps only partly with those of the diploids. Because *A. lyrata* is not found in Taiwan and Japan and *A. halleri* does not occur in North America, Far East Russia is the only area where the three species co-occur. Here we sampled the diploids mainly from Eastern Asia (Far East Russia, Japan) and from more remote areas (Europe and the USA). The samples represent 15 individuals of *A. halleri* from 13 populations and seven individuals of *A. lyrata* from four populations (Table 1). Although our sampling did not represent species-wide coverage (which was not the aim of the present study), we exploit here much of the sequence data on these diploids published previously (Ramos-Onsins *et al.* 2004; Koch & Matschinger 2007; Schmickl *et al.* 2008).

Chromosome number counts

Chromosome number was counted to check ploidy. Root tips were treated with cold water at 0°C for 24 hours, fixed in 3:1 (vol:vol) ethanol–acetic acid at 5°C for 1 hour, and stained in 1% acetic–orcein (see Shimizu *et al*. 2005). An individual from the population of kamC11, from which *CHS*-hal was not amplified, and two individuals from Kamchatka (the population of kamG39), where *A. kamchatica* was described originally, were assayed. Ihara (1976) reported triploid plants (as *Arabis* sp.) from the same site as kamC11 (Mt. Shikokutsurugi), but our chromosomal count showed a

tetraploid count.

Primer designs and strategies for separating homeologs and alleles

We sequenced two low-copy nuclear genes *WER (WEREWOLF)* and *CHS (CHALCONE SYNTHASE*), two cpDNA regions (the *trnL* intron and the *trnL–trnF* intergenic spacer region, see Koch *et al.* 2005; Ansell *et al.* 2007), and the ITS region of nuclear ribosomal DNA (ITS1-5.8S-ITS2) (Alvarez & Wendel 2003). The *WER* gene encodes a protein with an myb DNA-binding domain and is involved in the root hair development in *A. thaliana* (Lee & Schiefelbein 1999). A WU-BLAST search (www.arabidopsis.org) showed that the *WER* coding sequence has only 62%–80% identity with its closest homologs, *GL1* and *MYBRTF*, and is considered a single-copy gene. The *CHS* gene forms multigene families in some taxa but has been reported to be a single-copy gene in *A. thaliana* and other related diploid Brassicaceae taxa (Shimizu *et al*. 2005; Lihová *et al*. 2006).

To infer a phylogeny based on nuclear sequences, natural and artificial recombination between homeologs and between alleles should be excluded. To obtain sequences of two homeologs separately for both *WER* and *CHS* genes from the tetraploid *A. kamchatica*, we designed homeolog-specific primers using the methods described by Lihová *et al.* (2006) (Supporting Fig. S1a, Table S1, and Text S1). To obtain haplotypes (or allele sequences) within each homeolog, we applied the following strategies while avoiding cloning.

1. Plants were self-fertilized in a growth chamber to obtain homozygous individuals. This was feasible for *A. kamchatica* because selfing was possible in all 18 individuals we tried (Table 1). In several individuals of the self-incompatible diploids *A. halleri* and

A. lyrata (Castric & Vekemans 2004), pollination at the flower bud stage often avoided the self-incompatibility reaction and allowed self-fertilization.

2. When only one heterozygous site was found by direct sequencing, the haplotypes were resolved (e.g., *CHS* sequences named haltat1a1 and haltat1a2 from the individual haltat1, see Table 1).

3. When only one indel was found in direct sequencing, sequencing from both directions resolved the haplotypes (e.g., *WER* sequences kamA3Ha1 and kamA3Ha2; Table 1).

4. We sequenced multiple individuals of *A. halleri* subsp. *halleri* and subsp. *dacica* to find homozygotes.

Sequence alignments, copy numbers, phylogenetic and population genetic analysis, intrapopulation polymorphism, and Bayesian clustering

The details are given in Supporting Text S1 and Tables S1–S3. All sequences obtained in this study were deposited under GenBank accession numbers GQ303456–GQ303550. The sequence assemblies and alignments were performed in BioLign version 4.0.6.2 (http://www2.maizegenetics.net/index.php?page=bioinformatics/index.html) and edited manually using the program BioEdit version 7.0.4.1 (Hall 1999). After the alignments were assembled, identical sequences were detected using MacClade 4.0 PPC (Maddison & Maddison 2000) and merged, which reduced the alignments to comprise only the sets of unique sequences.

The final alignments of the nuclear regions (*WER*, *CHS*, *CHS*-lyr and ITS) were subjected to maximum-parsimony (MP) analysis (Swofford 2001) and to Bayesian inference based on a Markov chain Monte Carlo algorithm (MCMC; Huelsenbeck & Ronquist 2001). Bootstrap analyses (Felsenstein 1985) were performed using 100,000

228 resamplings with the fast-heuristic search as implemented in PAUP* (Swofford 2001).
229 Gaps were used as additional characters in the MP analyses (see also Text S1). Except
230 for short 1- or 2-bp gaps that appeared to be caused by slipped-strand mispairing, each
231 gap was scored, and the scoring was appended to the alignment. In the case of simple,
232 nonoverlapping gaps, these were coded as binary characters using the "simple gap-
233 coding" approach as suggested by Simmons and Ochoterena (2000). More complex
234 gaps (i.e., of different lengths and overlapping) were coded as multistate (up to four
235 states) characters.

236 The haplotype network of the *trnL* intron region was constructed by a minimum-
237 spanning network using the NETWORK program v. 4.5.1.0 (Bandelt *et al*. 1999; freely
238 available at www.fluxus-engineering.com). To survey the extent of polymorphism
239 within populations, sequences of the *trnL–trnF* region were obtained from additional
240 individuals of three populations (see Text S1 and Table S3). MEGA4 (Tamura *et al*.
241 2007) was used for neighbor-joining analysis. The minimum numbers of recombination
242 events were detected using the four-gamete test (Hudson & Kaplan 1985), and the levels
243 of silent-site nucleotide diversity of each homeolog were estimated as π (Tajima 1983)
244 (Table S2) as implemented in DnaSP version 4.10.7 (Rozas *et al*. 2003). To survey the
245 associations between different loci, the gametic disequilibrium D' (Hedrick 1987) was
246 calculated.

247 To infer the population structure of *A. kamchatica*, the Bayesian clustering
248 algorithm implemented in the program *structure* version 2.2
249 (http://pritch.bsd.uchicago.edu/structure.html) (Pritchard *et al*. 2000) was used. The data
250 were treated as haploid data as recommended for complete-selfing species (Gao *et al*.

2007) and as commonly done for predominantly-selfing species (e.g., Nordborg *et al*. 2005; Beck *et al*. 2008). The programs *CLUMPP* (Jakobsson & Rosenberg 2007), *distruct* (http://rosenberglab.bioinformatics.med.umich.edu/distruct.html) (Rosenberg 2004), and Δ*K* statistic (Evanno *et al*. 2005) were used to summarize and interpret the outputs.

Results

Chromosome counts

We counted chromosome numbers of two individuals from the population of kamG39 in Kamchatka and one individual from the population of kamC11 in Japan (see Table 1 for population origins). All were tetraploids with $2n = 32$ (Fig. 2).

Homeologous pairs of WER *and* CHS

Amplification of the nuclear genes *WER* and *CHS* in the tetraploid *A. kamchatica* using homeolog-specific primers resulted in two homeologs in each gene, named here as *WER*-hal, *WER*-lyr, *CHS*-hal, and *CHS*-lyr (Table 1). Among the 50 individuals analyzed, *CHS*-hal was not amplified from three individuals from Central Honshu and Shikoku (kamC11, kamC12, and kamD23), which suggests a large deletion or a rearrangement (see Text S1 for details). The tetraploid count of one of them (kamC11) indicates that it was not caused by the change in ploidy. These results were also supported by the survey of copy numbers using PCR and restriction patterns (Fig. S1) (following Lihová *et al.* 2006).

No natural recombination was detected in either *WER* or *CHS* homeologs of *A. kamchatica* using the four-gamete test (Hudson & Kaplan 1985). However, natural recombination was detected in three of eight *CHS* sequences obtained from European diploids (lyrpet2, haltat1a2, and halovi1) (Fig. S2a). We conducted a phylogenetic analysis without these three sequences.

Homeologous pairs of *A. kamchatica* were aligned with those of the diploid species *A. lyrata* and *A. halleri,* and with *A. pedemontana* and *A. thaliana* as outgroups. The phylogenetic trees of all *WER* and *CHS* sequences are shown in Figs 3 and 4. The phylogenetic tree of the *lyrata*-originated *CHS* homeolog alone was also inferred because we obtained longer sequences than for the *halleri*-originated homeolog, and several unique haplotypes were identified (Fig. S3a). Maximum parsimony analysis and Bayesian inference resulted in very similar tree topologies with slight differences only in weakly supported clades (see also Fig. S3, Text S1 and Table S2 for more details).

Three major clades with high bootstrap supports are resolved in the phylogenetic tree of *WER* (Fig. 3): one clade comprised all individuals of *A. halleri* and the corresponding homeolog from the tetraploids (*WER*-hal), another clade comprised three individuals of *A. lyrata* and the other homeolog from the tetraploids (*WER*-lyr), and the third clade included three individuals of *A. lyrata*. In the analysis of *CHS*, three major clades were resolved similarly (Fig. 4). These results strongly support the allopolyploid origin of *A. kamchatica* (including both subspecies) from the diploids *A. lyrata* and *A. halleri.*

Nucleotide diversity (π) of silent sites of the tetraploid *A. kamchatica* is in the range of 0.0006–0.0026 among the four nuclear loci (average 0.0013) (Table S2). The number of haplotypes resolved in the tetraploid is in the range of 8–11 among the four

299 loci when indel polymorphisms are included.

300 The geographic distribution of the nuclear haplotypes of *A. kamchatica* was not
301 random. We found a widespread and common haplotype within each locus (shown with
302 one or two black stars in Figs 3, 4; for *CHS*-lyr, see also Fig. S3a based on longer
303 sequences), which was observed mainly from the broad area F–I (Figs 1 and S4), along
304 with several geographically more restricted haplotypes. In particular, two regions
305 harbored one or more haplotypes that were geographically restricted in all four loci:
306 lowlands in Western Honshu (B), represented by *A. kamchatica* subsp. *kawasakiana*,
307 and mountains in Central Honshu together with mountains in Western Honshu and
308 Shikoku (C and D) (Figs 1 and S4). The division of these areas suggested by the
309 association between homeologs is also supported by the Bayesian clustering analysis
310 (see below).

311 Although our sampling of the parental diploid species was limited, we identified
312 one or two haplotype sequences that were identical to those of *A. kamchatica* in each of
313 the four loci (Figs 3 and 4, Text S1). First, in *WER*-hal and *CHS*-hal, two haplotypes
314 observed in *A. kamchatica* are identical to those of *A. halleri*. Second, in *WER*-lyr and
315 *CHS*-lyr, a haplotype found in North America is identical to a haplotype of *A. lyrata*
316 individuals from Far East Russia (lyrpet4 and 5). In addition, in *CHS*-lyr, an
317 intermediate frequency haplotype (kamCL,DL in Fig. S2a and S3a) is identical to a
318 haplotype of *A. lyrata* individuals from western Russia (lyrpet2) over more than 1 kb to
319 the left of the recombination breakpoint (Fig. S2a).

320 To increase the sequences of diploid taxa, 22 *CHS* sequences from *A. lyrata* and *A.*
321 *halleri* reported by Ramos-Onsins *et al.* (2004) were combined with our data. Critically,
322 Ramos-Onsins *et al.* (2004) used cloning, and the possibility of artificial recombination

cannot be excluded. Thus, we estimated the gene genealogy, in which the identity of the haplotypes could be revealed but the branch pattern might not reflect the historical phylogenetic relationship of the entire region (Fig. S5). We observed the same pattern of haplotype sharing between polyploid and diploid parents with or without the sequences reported by Ramos-Onsins *et al.* (2004); two haplotypes of *CHS*-hal were shared with *A. halleri*, and a haplotype of *CHS*-lyr was shared with *A. lyrata* (Figs 4 and S5). The haplotype of lyrpet2 mentioned above was identical to that from the same population sequenced by Ramos-Onsins *et al.* (2004).

These results confirm further that *A. kamchatica* is allopolyploid derived from *A. halleri* and *A. lyrata*, and that at least two distinct haplotypes of both *A. halleri* and *A. lyrata* were incorporated into *A. kamchatica*.

Chloroplast and ITS sequences

Once allopolyploidy is confirmed by low-copy nuclear genes, cpDNA is highly suitable to infer the independent origin of a polyploid because a unique uniparental haplotype is transmitted in each hybridization event without recombination. Sequencing two cpDNA regions (*trnL* intron and the *trnL-trnF* intergenic spacer region) resulted in 18 cpDNA haplotypes in *A. kamchatica*, *A. lyrata*, and *A. halleri*. In 50 individuals of *A. kamchatica*, we identified seven cpDNA haplotypes (cpHap1–7) (Table 1, Fig. S2c). Four of them (cpHap1, 2, 3, 5) were also found in diploid *A. halleri* subsp. *gemmifera* but not in any other diploid taxa. These results suggest that at least four individuals of *A. halleri* contributed to the origin of the allopolyploid species *A. kamchatica*.

We further analyzed our data in the context of previously published data of the two cpDNA regions (582 individuals from Koch & Matschinger 2007, Schmickl *et al.*

2008). Whereas the *trnL* intron region was alignable, the alignment of the *trnL–trnF* intergenic spacer region was uncertain because of frequent and possibly parallel mutations in tandemly duplicated copies (Fig. S2c). Thus, as the first step, the *trnL* intron region was used to construct a haplotype network. In our 50 samples of *A. kamchatica*, three haplotypes of the *trnL* intron (one in cpHap2, 3, 4, and 5, one in cpHap6, one in cpHap1 and 7) were identified (Fig. 5a). The resolution of the *trnL* intron region alone was not enough to infer the parental species because the same haplotypes have been observed in the published data of *A. arenosa*, *A. halleri*, and *A. lyrata*. Therefore, the identities of the *trnL–trnF* intergenic spacer region were considered to distinguish the cpDNA haplotypes from different species (Figs 5b and 6). Among seven haplotypes of *A. kamchatica* (cpHap1–7), we found that cpHap1, 2, 3, and 5 were shared with *A. halleri* subsp. *gemmifera* and that they were not found in any other diploid taxa. In turn, cpHap4, 6, and 7 were found exclusively in *A. kamchatica* analyzed here. These results agree with our conclusions reported above, suggesting that at least four cpDNA haplotypes of *A. halleri* subsp. *gemmifera* were incorporated into *A. kamchatica*.

In the tree based on the ITS region (ITS1–5.8*S*–ITS2), *A. kamchatica*, except for one individual (kamD17), formed a clade with diploid *A. lyrata*. In contrast, the ITS of the single individual from Japan (kamD17) was distinct from all other allotetraploids and formed a clade with *A. halleri* (Figs 7, S2b, and S3d).

Population structure of Arabidopsis kamchatica

We used a model-based Bayesian clustering method (Pritchard *et al.* 2000) to infer the population structure of *A. kamchatica* integrating the information from the four nuclear

loci (*WER*-hal, *WER*-lyr, *CHS*-hal, *CHS*-lyr) and the cpDNA (Fig. 8 and Text S1). The high values of the mean posterior probability of data Ln $P(X|K)$, ΔK and the symmetric similarity coefficient (*SSC*) supported the clustering of $K = 3$ (Fig. 8b–d). The three clusters correspond to those described above in the section of phylogeny. Cluster 1 (light green in Fig. 8a) covers a wide range of distribution including northern Japan, Kamchatka, Alaska, Canada, and the Pacific Northwest of the USA (areas F–I). Cluster 2 (orange) includes *A. kamchatica* subsp. *kawasakiana* from lowlands in Western Honshu (B) and three individuals from Taiwan (A). Cluster 3 (blue) comprises the individuals from mountains in Central Honshu, Western Honshu, and Shikoku (C and D) (Figs 1 and 8a).

Genetic admixtures between different clusters were suggested mostly in geographic border regions. For example, one individual from the lowlands of Northern Honshu (kamE26 in area E) had haplotypes characteristic of subsp. *kawasakiana* in *WER*-hal and *CHS*-hal, whereas haplotypes typical for subsp. *kamchatica* were seen in *CHS*-lyr and cpDNA, suggesting that it is a hybrid between these two subspecies. In addition, individuals in area F had common haplotypes of low-copy nuclear genes and belonged to cluster 1, but its cpDNA (cpHap3) was mainly found in cluster 3. Such 'plastid capture' is observed often in plant species (Okuyama *et al.* 2005).

Discussion

Allopolyploid origin of Arabidopsis kamchatica *from* A. lyrata *and* A. halleri

In contrast to the previous studies on *Arabidopsis* species, exploring mainly cpDNA and

395 ITS nuclear ribosomal data (Koch & Matschinger 2007), we focused the present study
396 on biparentally inherited low-copy nuclear genes. By designing homeolog-specific
397 primers, we targeted two genes (*WER* and *CHS*) in the tetraploid *A. kamchatica*, for
398 which both allo- and autopolyploidy were argued previously (Shimizu *et al.* 2005, Koch
399 & Matschinger 2007). Confounding factors in the analyses of low-copy nuclear genes
400 are artificial and natural recombination. In the present study, the former problem was
401 excluded by avoiding cloning procedure, and the latter was not detected by the four-
402 gamete test in our Asian and American materials. The subsequent phylogenetic analysis
403 of both *WER* and *CHS* genes revealed that one of the homeologs retrieved from *A.*
404 *kamchatica* clustered with *A. lyrata*, whereas the other clustered with *A. halleri*. We
405 obtained congruent results for the recently studied *FWA* genes for two individuals of *A.*
406 *kamchatica* (representing both subspecies), which showed homeologs corresponding to
407 *A. lyrata* and *A. halleri*, respectively (Fujimoto *et al.* 2008). These results provide strong
408 evidence that *A. kamchatica* (both subspecies recognized by Shimizu *et al.* 2005) is an
409 allopolyploid derived from the diploids *A. lyrata* and *A. halleri*.

410 Bayesian cluster analysis that integrated the nuclear and cpDNA haplotype data
411 identified three geographically defined clusters. Cluster 1 covered a broad range from
412 Northern Japan and Kamchatka to North America. Cluster 2 comprised mainly *A.*
413 *kamchatica* subsp. *kawasakiana* in lowlands of Western Honshu, Japan. Cluster 3
414 included the individuals from mountains in Central Honshu, in which a number of rare
415 haplotypes were found. In the next sections, we discuss the origin of polyploidy as a
416 possible explanation of this population structure.

417 Hybrid and allopolyploid origins have quite often been inferred from the
418 incongruence between cpDNA and ITS data. However, this approach can fail to detect

419 the hybrid origin in cases when both cpDNA and ITS represent only one of the parents
420 or when the sampling and/or the resolution is not adequate (Kim *et al.* 2008). Thus, it
421 may not be suitable for species-wide analysis of polyploid species. The cpDNA and ITS
422 of *A. kamchatica* (Koch & Matschinger 2007) did not suggest the hybrid origin of
423 American individuals, although these support the hybrid origin of Japanese individuals
424 that was shown previously using the low-copy nuclear gene *CHS* (Shimizu *et al.* 2005).
425 We suggest that, provided that the effects of natural and artificial recombination are
426 avoided, low-copy nuclear genes provide critical information for the study of hybrid and
427 allopolyploid origins, as well as the geographic organization of their genetic variation.
428
429 *The origin of the allopolyploid* Arabidopsis kamchatica *from multiple individuals of its*
430 *diploid parents*
431 Critical data to support the independent origins of polyploid species include the sharing
432 of multiple haplotypes between diploid and polyploid species (Soltis *et al.* 2003). Here
433 we found ample evidence in both cpDNA and low-copy nuclear DNA that multiple
434 haplotypes of the parental species (*A. lyrata* and *A. halleri*) contributed to the polyploid
435 *A. kamchatica*. Because ITS sequences displayed a low level of variation, we discuss
436 these only briefly.
437 Here we identified seven cpDNA haplotypes in the allopolyploid *A. kamchatica*,
438 and four of them were shared with the Asian diploid taxon, *A. halleri* subsp. *gemmifera*.
439 Even when we increased the sample size by incorporating the large-scale surveys of
440 cpDNA in the genus *Arabidopsis* by Koch and Matschinger (2007) and Schmickl *et al.*
441 (2008), the same pattern of haplotype sharing was found. These results suggest strongly
442 that at least four individuals of *A. halleri* subsp. *gemmifera* contributed to the origin of *A.*

443 *kamchatica* because a unique uniparental haplotype of cpDNA is usually transmitted at
444 each generation without recombination. This also indicates that *A. halleri* subsp.
445 *gemmifera* was always the maternal parent, although the possibility of shared
446 polymorphism caused by rare introgression or by incomplete lineage sorting between *A.*
447 *lyrata* and *A. halleri* cannot be excluded (Ramos-Onsins *et al.* 2004).

448 In *WER*-hal and *CHS*-hal homeologs, two haplotypes observed in *A. kamchatica*
449 were identical to those of *A. halleri*. In addition, two haplotypes of *CHS*-lyr were found
450 in *A. lyrata* over more than 1 kb. We increased the sample size by incorporating 22 *CHS*
451 sequences of *A. lyrata* and *A. halleri* reported by Ramos-Onsins *et al.* (2004). Again, the
452 same pattern of haplotype sharing was found. These results are consistent with
453 independent origins, although it is also possible that two haplotypes may have entered
454 the tetraploid *A. kamchatica* through an unreduced diploid gamete of a parent in a single
455 polyploidization event.

456 It is expected that many haplotypes of polyploid species cannot be identified from
457 diploid parents, possibly because the sampling may not be dense enough or because the
458 haplotypes may be derived in polyploids or lost in diploids. The loss in diploid would be
459 pronounced and complicated in low-copy nuclear genes because they are subjected to
460 recombination. Recombination was detected in our relatively long sequence length (~1
461 kb) in diploid parental species, which are self-incompatible (Castric & Vekemans 2004).
462 For example, over 1 kb of *CHS* was shared between diploid (lyrpet2) and polyploid
463 individuals (*CHS*-lyr of kamCL,DL in Fig. S2a), but a recombination breakpoint was
464 identified. It is possible that the haplotype sharing of low-copy nuclear genes is limited
465 only if gene flow occurred very recently or if a haplotype was common and maintained
466 for a long time. We suggest that cpDNA is useful for studying independent origins

because of the uniparental inheritance and the absence of recombination between haplotypes, once allopolyploidy is confirmed by low-copy nuclear genes. Our data suggest strongly that multiple individuals of parental species contributed to the origin of the allopolyploid *A. kamchatica*.

Independent polyploidization events vs. introgression

Although a common interpretation of the sharing of multiple haplotypes has been independent polyploidization events, it has been also noted that introgression from diploid into polyploid can also result in haplotype sharing (Ramsey & Schemske 1998; Husband 2004). Introgression is possible through a triploid bridge or through the hybridization of polyploids with unreduced gametes or with autopolyploids of a parental diploid, although the fertility of such hybrid individuals tends to be low. Distinguishing these two scenarios is challenging because their effects would be similar. However, we propose that the two scenarios would have different effects on low-copy nuclear genes when combined with cpDNA (Fig. 9), although they cannot be distinguished with certainty. In the introgression scenario, only the introgression parent would contribute additional sequence diversity to the homeologous loci, whereas the loci from the other parent would not receive any new haplotypes. Thus, homeologs derived from only one parent should have distinct haplotypes compared with other polyploid individuals, whereas the homeologs from other parents would be maintained. In contrast, in the scenario of independent polyploidization events, homeologs derived from both parents could be distinct from those of other polyploid individuals.

In our data, a few individuals in North America among the cluster 1 may represent the scenario of introgression. Most of the individuals in cluster 1 (areas F–I) displayed a

491 single haplotype (most common and widespread among the tetraploids) in all four
492 nuclear loci as well as cpDNA (Fig. 9, represented by kamG34 individual). Its cpDNA
493 haplotype (cpHap1) was also found in diploid *A. halleri* subsp. *gemmifera*. Nevertheless,
494 a few individuals (including kamH46 and kamI48 in Fig. 9) showed different *lyrata*-
495 homeolog sequences (in both *WER* and *CHS* datasets; Figs 3, 4, and 9). These
496 homeologs are shared with two individuals of *A. lyrata* from Far East Russia (lyrpet4,
497 lyrpet5, see Table 1), suggesting a recent gene flow. Similarly, the ITS sequence of a
498 few individuals was consistent with the overlap of a common haplotype of *A.*
499 *kamchatica* and a haplotype of *A. lyrata* (lyrlyr1,2, lyrpet4,5; Fig. S2b), suggesting
500 again a recent gene flow. These data might suggest introgression from *A. lyrata* rather
501 than independent polyploidization events. We note that it is difficult to exclude the
502 possibility that it represents another independent polyploidization event and that the
503 *halleri*-parent was nearly identical.

504 On the other hand, populations of *A. kamchatica* subsp. *kawasakiana* (area B in
505 cluster 2) are suggested to represent an independent polyploidization event. In both
506 *lyrata*- and *halleri*-originated homeologs (*WER*-lyr, *WER*-hal, *CHS*-lyr and *CHS*-hal),
507 subsp. *kawasakiana* exhibited rare (mostly unique to this group) haplotypes that are
508 distinct from other individuals (Fig. 9, represented by kamkwsB8). This is not consistent
509 with introgression from a parent. Its cpDNA haplotype (cpHap2) was shared between
510 the polyploid subsp. *kawasakiana* and the diploid *A. halleri* subsp. *gemmifera* (Fig. 5).
511 These results suggest strongly that *A. kamchatica* subsp. *kawasakiana* originated by an
512 independent polyploidization event compared with other polyploid individuals.
513 Although we cannot exclude formally the possibility of introgression from both parents
514 or of recurrent derived mutations, they would be much less parsimonious.

515 The most complex pattern appeared in cluster 3, which had a number of unique
516 haplotypes along with the widespread haplotype. Two cases in area D (mountains in
517 Central Honshu) fulfill the same criterion of independent origins as subspecies
518 *kawasakiana*. First, in many individuals with cpHap3 (represented by kamD16 in Fig.
519 9), both *lyrata*- and *halleri*-originated homeologs were distinct, and their cpDNA
520 haplotype was also found in *A. halleri* subsp. *gemmifera*. Second, a single individual
521 (kamD18 in Fig. 9) had unique haplotypes in both *lyrata*- and *halleri*-originated
522 homeologs (*CHS*-lyr and *CHS*-hal), and its cpDNA haplotype (cpHap5) was shared with
523 *A. halleri* subsp. *gemmifera*. In addition, kamD18 had a distinct ITS haplotype, which is
524 consistent with its unique history. These data suggest that those individuals had
525 independent origins from the other tetraploids analyzed, although more data from the
526 same population are needed to provide more details.
527 In summary, our data suggest that *A. kamchatica* comprises individuals with
528 independent origins. The independent origins of subspecies *kawasakiana* from other
529 individuals were strongly suggested, and two more independent origins of individuals in
530 the mountains in Central Honshu are also suggested. In the Bayesian clustering (Fig. 8),
531 three of the four suggested independent origins appeared as distinct clusters, except for
532 one represented by a single individual kamD18. These results suggest strongly that the
533 independent origins had a profound effect on the population structure of *A. kamchatica*.
534 In addition, introgression from *A. lyrata* into *A. kamchatica* would explain the distinct
535 haplotypes found in some North American populations.
536
537 *Genetic diversity in* Arabidopsis kamchatica
538 The nucleotide diversity was lower in the self-compatible species *A. kamchatica*

539 (average 0.0013) than in the outcrossing parental species (0.0150 in *A. halleri*, 0.0230 in
540 *A. lyrata* subsp. *petraea*, and 0.0031 in *A. lyrata* subsp. *lyrata*) (Ramos-Onsins *et al.*
541 2004). It was also lower than in the self-compatible *A. thaliana* (~0.0035–0.0055 at
542 synonymous sites and ~0.007 at intronic regions) (Nordborg *et al.* 2005). This low
543 nucleotide diversity of *A. kamchatica* might reflect its self-compatibility, and it might
544 also reflect a bottleneck in which only a few haplotypes of the parental species were
545 incorporated into *A. kamchatica* and a relatively recent origin of this species.
546 Consistently, many haplotypes of the *WER* and *CHS* genes and cpDNA of *A.*
547 *kamchatica* are identical to those of the parental species, supporting the idea that the
548 origin of *A. kamchatica* is relatively recent.

549 Interestingly, diverse haplotypes arising from independent origins of *A.*
550 *kamchatica* were found in the mountains in Central Honshu and surrounding areas in
551 Japan, where *A. lyrata* is not found currently. Thus, this diversity cannot be explained
552 solely by contemporary or very recent polyploidization and introgression. Because the
553 mountains of Central Honshu, Japan, are known to have acted as refugia for many plant
554 species (Fujii & Senni 2006), we suggest that *A. kamchatica* or its parental species
555 might have survived there during Pleistocene glacial periods. The independent origins
556 of *A. kamchatica* contrast with the single origin of *A. suecica* (Jakobsson *et al.* 2006),
557 which is distributed in Northern Europe and might not have originated in refugial areas.

558

559 Arabidopsis kamchatica *as a model to study the molecular basis of polyploid evolution*
560 Although polyploid species with very recent independent origins such as *Tragopogon*
561 and artificial polyploids provide insights into the immediate responses of

562 polyploidization (Comai 2005; Tate *et al.* 2006; Otto 2007), *A. kamchatica* offers a
563 different case, in which evolutionary changes occurred over a longer timescale. Our
564 previous report on the epigenetically regulated *FWA* gene showed that the homeolog
565 derived from *A. halleri* is silenced in both subspecies. In conjunction with the finding of
566 independent origins, this case represents an example of repeatable gene silencing after
567 polyploidization (Adams & Wendel 2005). We have also shown that *A. kamchatica* is
568 self-compatible (Table 1; Sugisaka & Kudoh 2008), whereas *A. halleri* and most *A.*
569 *lyrata* have been reported to be predominantly self-incompatible (Castric & Vekemans
570 2004; Mable *et al.* 2004). This suggests either that self-incompatibility was lost
571 independently, as reported in a few species (Okamoto *et al.* 2007; Shimizu *et al.* 2008),
572 or that self-compatible haplotypes spread beyond different populations. A biogeographic
573 study of the genus *Arabidopsis* (Hoffmann 2005) reported that *A. kamchatica* (as *A.*
574 *lyrata* subsp. *kamchatica*, most of which should correspond to *A. kamchatica*) grows in
575 a broader range of climate in terms of temperature and precipitation than other
576 subspecies of *A. lyrata* and *A. halleri*. *Arabidopsis kamchatica* will be a unique model to
577 understand the molecular basis of parallel evolution and habitat exploitation in
578 polyploid species.
579
580
581 **Acknowledgements**
582
583 The authors thank Ana Caicedo, Maria Clauss, Shinji Fujii, Matthias Helling, Masahiro
584 Kanaoka, Akira Kawabe, Tetsuhiro Kawagoe, Masaki Kobayashi, Martin Kolník,
585 Masako Komaki, Masumi Kono, Akiyo Naiki, Hideki Nakai, Sachiko Nishida,

Hanspeter Schöb, Jiro Sugisaka, Ryohei Terauchi, Takashi Tsuchimatsu, Hirokazu Tsukaya, Nobuko Yahata, and Masaki Yamaguchi for sample collections, and Kiyotaka Okada, Michael Purugganan, Hajime Ikeda, and Dusan Senko for experimental and analytical support. This work was supported by the Swiss National Foundation, the University Research Priority Program of Systems Biology / Functional Genomics of the University of Zurich, Grants-in-Aid for Special Research on Priority Areas (No. 119043010) from MEXT, Japan, the Kyoto University gCOE program, Forschungskredit of the University of Zurich to R.S.-I. and H.I., and Ministry of Education, Youth and Sports of the Czech Republic (grant no. 0021620828) to K.M.

References

Adams KL, Wendel JF (2005) Novel patterns of gene expression in polyploid plants. *Trends in Genetics* **21**, 539–543.

Alvarez I, Wendel JF (2003) Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution* **29**, 417–434.

Anderson E, Stebbins GL (1954) Hybridization as an evolutionary stimulus. *Evolution* **8**, 378–388.

Ansell SW, Schneider H, Pedersen N, *et al.* (2007) Recombination diversifies chloroplast *trnF* pseudogenes in *Arabidopsis lyrata*. *Journal of Evolutionary Biology* **20**, 2400–2411.

Bandelt HJ, Forster P, Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* **16**, 37–48.

610 Beck JB, Schmuths H, Schaal BA (2008) Native range genetic variation in *Arabidopsis*
611 *thaliana* is strongly geographically structured and reflects Pleistocene glacial
612 dynamics. *Molecular Ecology* **17**, 902–915.
613 Castric V, Vekemans X (2004) Plant self-incompatibility in natural populations: a
614 critical assessment of recent theoretical and empirical advances. *Molecular*
615 *Ecology* **13**, 2873–2889.
616 Comai L (2005) The advantages and disadvantages of being polyploid. *Nature Review*
617 *Genetics* **6**, 836–846.
618 Cronn R, Cedroni M, Haselkorn T, Grover C, Wendel JF (2002) PCR-mediated
619 recombination in amplification products derived from polyploid cotton.
620 *Theoretical and Applied Genetics* **104**, 482–489.
621 Dart S, Kron P, Mable BK (2004) Characterizing polyploidy in *Arabidopsis lyrata* using
622 chromosome counts and flow cytometry. *Canadian Journal of Botany* **82**, 185–
623 197.
624 Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals
625 using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**,
626 2611–2620.
627 Felsenstein J (1985) Confidence limits on phylogenies — an approach using the
628 bootstrap. *Evolution* **39**, 783–791.
629 Fujii N, Senni K (2006) Phylogeography of Japanese alpine plants: biogeographic
630 importance of alpine region of Central Honshu in Japan. *Taxon* **55**, 43–52.
631 Fujimoto R, Kinoshita Y, Kawabe A*, et al.* (2008) Evolution and control of imprinted
632 *FWA* genes in the genus *Arabidopsis*. *PLoS Genetics* **4**, e1000048.
633 Gao H, Williamson S, Bustamante CD (2007) A Markov chain Monte Carlo approach

for joint inference of population structure and inbreeding rates from multilocus genotype data. *Genetics* **176**, 1635–1651.

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acid Symposium Series* **41**, 95–98.

Hedrick PW (1987) Gametic disequilibrium measures: proceed with caution. *Genetics* **117**, 331–341.

Hoffmann MH (2005) Evolution of the realized climatic niche in the genus *Arabidopsis* (Brassicaceae). *Evolution* **59**, 1425–1436.

Hudson RR, Kaplan NL (1985) Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* **111**, 147–164.

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754–755.

Husband BC (2004) The role of triploid hybrids in the evolutionary dynamics of mixed-ploidy populations. *Biological Journal of the Linnean Society* **82**, 537–546.

Ihara K (1976) Mode of local differentiation in *Arabis lyrata* and *A. gemmifera* (Cruciferae) in Japan. *Journal of the Faculty of Science, University of Tokyo, III* **12**, 1–36.

Jakobsson M, Hagenblad J, Tavare S, *et al.* (2006) A unique recent origin of the allotetraploid species *Arabidopsis suecica*: Evidence from nuclear DNA markers. *Molecular Biology and Evolution* **23**, 1217–1231.

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of

population structure. *Bioinformatics* **23**, 1801–1806.

Kim ST, Sultan SE, Donoghue MJ (2008) Allopolyploid speciation in *Persicaria* (Polygonaceae): insights from a low-copy nuclear region. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 12370–12375.

Koch MA, Dobes C, Matschinger M*, et al.* (2005) Evolution of the *trnF*(GAA) gene in *Arabidopsis* relatives and the Brassicaceae family: monophyletic origin and subsequent diversification of a plastidic pseudogene. *Molecular Biology and Evolution* **22**, 1032–1043.

Koch MA, Matschinger M (2007) Evolution and genetic differentiation among relatives of *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 6272–6277.

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

Kuittinen H, Niittyvuopio A, Rinne P, Savolainen O (2008) Natural variation in *Arabidopsis lyrata* vernalization requirement conferred by a *FRIGIDA* indel polymorphism. *Molecular Biology and Evolution* **25**, 319–329.

Lee MM, Schiefelbein J (1999) WEREWOLF, a MYB-related protein in *Arabidopsis*, is a position-dependent regulator of epidermal cell patterning. *Cell* **99**, 473–483.

Levin AD (2002) *The Role of Chromosomal Change in Plant Evolution.* Oxford University Press, New York.

Lihová J, Shimizu KK, Marhold K (2006) Allopolyploid origin of *Cardamine asarifolia* (Brassicaceae): incongruence between plastid and nuclear ribosomal DNA

681 sequences solved by a single-copy nuclear gene. *Molecular Phylogenetics and*
682 *Evolution* **39**, 759–786.
683 Mable BK, Beland J, Di Berardo C (2004) Inheritance and dominance of self-
684 incompatibility alleles in polyploid *Arabidopsis lyrata*. *Heredity* **93**, 476–486.
685 Maddison DR, Maddison WP (2000) *MacClade4: analysis of phylogeny and character*
686 *evolution, version 4.0*. Sinauer, Sunderland, MA, USA.
687 Marhold K, Lihová J (2006) Polyploidy, hybridization and reticulate evolution: lessons
688 from the Brassicaceae. *Plant Systematics and Evolution* **259**, 143–174.
689 Mulligan GA (1995) Synopsis of the genus *Arabis* (Brassicaceae) in Canada, Alaska and
690 Greenland. *Rhodora* **97**, 109–163.
691 Nordborg M, Hu TT, Ishino Y*, et al.* (2005) The pattern of polymorphism in
692 *Arabidopsis thaliana*. *PLoS Biology* **3**, 1289–1299.
693 O'Kane SLJ, Al-Shehbaz IA (1997) A synopsis of *Arabidopsis* (Brassicaceae). *Novon* **7**,
694 323–327.
695 Ohno S (1970) *Evolution by gene duplication.* Springer-Verlag, New York.
696 Okamoto S, Odashima M, Fujimoto R*, et al.* (2007) Self-compatibility in *Brassica*
697 *napus* is caused by independent mutations in *S*-locus genes. *Plant Journal* **50**,
698 391–400.
699 Okuyama Y, Fujii N, Wakabayashi M*, et al.* (2005) Nonuniform concerted evolution and
700 chloroplast capture: heterogeneity of observed introgression patterns in three
701 molecular data partition phylogenies of Asian *Mitella* (Saxifragaceae).
702 *Molecular Biology and Evolution* **22**, 285–296.
703 Otto SP (2007) The evolutionary consequences of polyploidy. *Cell* **131**, 452–462.
704 Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using

multilocus genotype data. *Genetics* **155**, 945–959.

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

Ramsey J, Schemske DW (1998) Pathways, mechanisms and rates of polyploid formation in flowering plants. *Annual Review of Ecology and Systematics* **29**, 477–501.

Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes* **4**, 137–138.

Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**, 2496–2497.

Schmickl R, Jørgensen MH, Brysting AK, Koch MA (2008) Phylogeographic implications for the North American boreal-arctic *Arabidopsis lyrata* complex. *Plant Ecology & Diversity* **1**, 245–254.

Shimizu KK (2002) Ecology meets molecular genetics in *Arabidopsis*. *Population Ecology* **44**, 221–233.

Shimizu KK, 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 Phytotaxonomica et Geobotanica* **56**, 165–174.

Shimizu KK, Purugganan MD (2005) Evolutionary and ecological genomics of Arabidopsis. *Plant Physiology* **138**, 578–584.

Shimizu KK, Shimizu-Inatsugi R, Tsuchimatsu T, Purugganan MD (2008) Independent

origins of self-compatibility in *Arabidopsis thaliana*. *Molecular Ecology* **17**, 704–714.

Simmons MP, Ochoterena H (2000) Gaps as characters in sequence-based phylogenetic analyses. *Systematic Biology* **49**, 369–381.

Soltis DE, Soltis PS (1993) Molecular-Data and the Dynamic Nature of Polyploidy. *Critical Reviews in Plant Sciences* **12**, 243–273.

Soltis DE, Soltis PS (1999) Polyploidy: recurrent formation and genome evolution. *Trends in Ecology & Evolution* **14**, 348–352.

Soltis DE, Soltis PS, Tate JA (2003) Advances in the study of polyploidy since Plant speciation. *New Phytologist* **161**, 173–191.

Stebbins GL (1950) *Variation and evolution in plants.* Columbia Univ, New York.

Stebbins GL (1971) *Chromosomal evolution in higher plants.* Edward Arnold, London.

Sugisaka J, Kudoh H (2008) Breeding system of the annual Cruciferae, *Arabidopsis kamchatica* subsp. *kawasakiana*. *Journal of Plant Research* **121**, 65–68.

Swofford DL (2001) *PAUP* phylogenetic analysis using parsimony (and other methods)*. Sinauer, Sunderland, MA, USA.

Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics* **105**, 437–460.

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* **24**, 1596–1599.

Tate JA, Ni Z, Scheen AC*, et al.* (2006) Evolution and expression of homeologous loci in *Tragopogon miscellus* (Asteraceae), a recent and reciprocally formed allopolyploid. *Genetics* **173**, 1599–1611.

753 Warwick SI, Al-Shehbaz IA (2006) Brassicaceae: Chromosome number index and
754 database on CD-Rom. *Plant Systematics and Evolution* **259**, 237–248.
755 Watanabe K, Yahara T (1984) Studies on the Asian Eupatoria. 2. Cytogeography of
756 *Eupatorium chinense* subsp. *sachalinense* var. *oppositifolium*. *Botanical*
757 *Magazine-Tokyo* **97**, 87–105.
758
759 Research interest of the authors
760 R.S.-I. is interested in the molecular and physiological basis of plant polyploidization.
761 J.L. and K.M. are interested in the taxonomy, phylogeny and phylogeography of
762 Brassicaceae. H.I. is interested in the nucleotide variation of a variety of organisms
763 to infer their histories. H.K. is working on the ecological genetics of adaptation in plants.
764 O.S. studies the genetics of adaptation in plants, in *Arabidopsis* species and Scots pine.
765 K.W. is interested in polyploid evolution. V.V.Y. studies the flora of the Far East. K.K.S.
766 is interested in evolutionary and ecological functional genomics of *Arabidopsis* relatives.
767
768

768 **Figure legends**

769

770 Fig. 1. Map showing the sample sites and geographic distribution of haplotypes of the
771 *CHS*-hal homeolog. Circles indicate *Arabidopsis kamchatica* subsp. *kamchatica* and
772 asterisks indicate subsp. *kawasakiana*. Eight haplotypes are depicted in different colors,
773 as shown also in Fig. 4. Heterozygotes are shown as half circles. Populations with
774 missing *CHS*-hal data are indicated by white circles. The Honshu and Shikoku islands
775 of the Japanese archipelago are magnified.

776

777 Fig. 2. Chromosome number count in *Arabidopsis kamchatica*. Mitotic metaphase
778 chromosomes ($2n = 32$) from the population named kamG32 in Kamchatka, Russia (A),
779 and from the population kamC11 in mountains in Western Honshu and Shikoku, Japan
780 (B). For population origin, see Table 1.

781

782 Fig. 3. Strict consensus tree of the 78 most-parsimonious trees based on nuclear *WER*
783 sequence data (955 aligned positions plus additional coding of 14 indels). Bootstrap
784 values above 50% are shown along the branches. The tree displays 24 unique sequences,
785 representing 120 sequences obtained from 72 individuals. Haplotype names follow
786 those in Table 1, and the numbers preceding the names indicate the number of
787 sequences represented by a given branch. *WER*-lyr of E26L and A1-3L are not included
788 in the tree because of the existence of long indels.

789

790 Fig. 4. Strict consensus tree of the four most-parsimonious trees based on nuclear *CHS*
791 sequence data (1314 aligned positions plus additional coding of six indels). Bootstrap

values above 50% are shown along the branches. The tree displays 22 unique sequences, representing 118 sequences obtained from 68 individuals. Haplotype names follow those in Table 1, and the numbers preceding the names indicate the number of sequences represented by a given branch. The colored symbols correspond to those in Fig. 1.

Fig. 5. Two cpDNA regions (*trnL* intron and *trnL–trnF* intergenic spacer) of *Arabidopsis kamchatica* and other *Arabidopsis* species. **a.** Minimum spanning network based on the *trnL* intron region of *Arabidopsis* species. The *trnL* intron sequences obtained in this study and obtained from GenBank were used to construct the network. GenBank data (Koch & Matschinger 2007; Schmickl *et al.* 2008) are marked with stars and are categorized as explained in Text S1. The circle size indicates the number of individuals. Crossbars on a branch represent unsampled or extinct haplotypes. **b.** Inset: cpDNA haplotypes incorporating data from the *trnL–trnF* intergenic spacer region as well as *trnL* intron region. cpHap2–5 have the same haplotype at the *trnL* intron and are distinguishable in their deletion pattern in the *trnL–trnF* intergenic spacer region. Likewise, cpHap1 and cpHap7 have the same haplotype at the *trnL* intron. The haplotype cpHap6 represents another haplotype at the *trnL* intron. The haplotypes that are not identical to those of *A. kamchatica* are represented as "others". The number below each circle represents the number of individuals included.

Fig. 6. Map showing the geographic distribution of cpDNA haplotypes. Circles indicate *Arabidopsis kamchatica* subsp. *kamchatica*, asterisks subsp. *kawasakiana*, triangles *A. halleri*, and squares *A. lyrata*. Haplotypes are depicted in different colors: haplotype 1,

816 pink; 2, green; 3 orange; 4, sky blue; 5, dark blue; 6, yellow; and 11, red. Haplotype 7 is
817 not visible because of overlapping with 6, and the distributions of 8 to 10 and 12 to 18
818 are outside of the range of this map (see Table 1). The Honshu and Shikoku islands of
819 the Japanese archipelago are magnified. Seven haplotypes were found in *A. kamchatica*
820 (including both subspecies). Four of them (cpHap1, 2, 3, 5) were found both in diploid
821 *A. halleri*. subsp. *gemmifera* and *A. kamchatica*; cpHap1 was common and widespread
822 throughout *A. kamchatica*, cpHap2 was restricted mostly to subsp. *kawasakiana* (area
823 B), cpHap3 was found in Honshu and Hokkaido (areas C, D, and F), and cpHap5 was
824 found in one individual in Central Honshu (kamD18). Three other haplotypes of *A.*
825 *kamchatica* (cpHap4, 6, 7) were not shared with diploid taxa; cpHap4 was short and
826 found in *A. kamchatica* subsp. *kawasakiana* and may represent a deletion derivative,
827 and cpHap6 and cpHap7 were found in Taiwan. See also Fig. 8a.
828
829 Fig. 7. Fifty-percent majority-rule consensus tree of 74,944 most-parsimonious trees
830 based on nuclear ribosomal ITS sequence data. Values above the branches indicate the
831 percentage of the most-parsimonious trees bearing their respective clades. The values in
832 brackets below the branches are bootstrap values (above 50%). The tree displays 25
833 unique sequences, representing 75 individuals.
834
835 Fig. 8. Genetic structure of *Arabidopsis kamchatica* based on cpDNA and nuclear *WER*
836 and *CHS* data, as inferred by the Bayesian clustering algorithm implemented in
837 *structure* software.
838 **a.** Population structure of *A. kamchatica*. Each individual is shown as a thin vertical
839 column partitioned into *K* colored components representing inferred membership in *K*

genetic clusters. The regional origin of the individuals (A–I) is shown on the top. The individual name (ID 1–50) and the cpDNA haplotype (cpHap1–7) of each individual are shown below. **b.** Mean symmetric similarity coefficient (*SSC*) ± SD over 190 pairs of 20 runs for each *K* value. **c.** Mean posterior probability of data Ln $P(X|K)$ ± S over 20 runs for each *K* value. **d.** Plot of ΔK for each *K*.

Fig. 9. Schematic diagram of haplotypes of *A. kamchatica* and two scenarios, introgression from a diploid vs. independent polyploidization events.

The four haplotypes of cpDNA (cpHap1, 2, 3, and 5) are shared with diploid *A. halleri* subsp. *gemmifera*. The haplotypes of six loci (cpDNA, ITS, *WER*-hal, *CHS*-hal, *WER*-lyr, and *CHS*-lyr) of six representative individuals are shown. In each locus, different haplotypes are shown by different shapes, and also by colors that correspond to those in Figs 1, 4, 6, S3 and S4. The ITS sequence of the kamI48 individual was heterogeneous. See text for details.

Table 1. Samples

Name of taxon	Area		Sample name	Population
Arabidopsis kamchatica subsp. *kamchatica*	A	Taiwan	kamA1	Taiwan, Taroko N.P., close to the entrance of the park, 2930 m
Arabidopsis kamchatica subsp. *kamchatica*	A	Taiwan	kamA2	Taiwan, Taroko N.P., close to the entrance of the park, 2930 m
Arabidopsis kamchatica subsp. *kamchatica*	A	Taiwan	kamA3	Taiwan, Taroko N.P., close to the high altitude experimental station, 3000 m
Arabidopsis kamchatica subsp. *kawasakiana*	B	Lowland in Western Honshu, Japan	kamkwsB4	Japan, Mie, Meiwa, Fukiiura, 2 m
Arabidopsis kamchatica subsp. *kawasakiana*	B	Lowland in Western Honshu, Japan	kamkwsB5	Japan, Mie, Meiwa, Fukiiura, 2 m
Arabidopsis kamchatica subsp. *kawasakiana*	B	Lowland in Western Honshu, Japan	kamkwsB6	Japan, Shiga, Takashima, 85m
Arabidopsis kamchatica subsp. *kawasakiana*	B	Lowland in Western Honshu, Japan	kamkwsB7	Japan, Shiga, Takashima, 85m
Arabidopsis kamchatica subsp. *kawasakiana*	B	Lowland in Western Honshu, Japan	kamkwsB8	Japan, Shiga, Ohtsu, Ohmimaiko, 85 m
Arabidopsis kamchatica subsp. *kawasakiana*	B	Lowland in Western Honshu, Japan	kamkwsB9	Japan, Shiga, Hikone, 85 m
Arabidopsis kamchatica subsp. *kawasakiana*	B	Lowland in Western Honshu, Japan	kamkwsB10	Japan, Toyama, Hamakurosaki, 2 m
Arabidopsis kamchatica subsp. *kamchatica*	C	Mountains in Western Honshu and Shikoku, Japan	kamC11	Japan, Tokushima, Mt. Shikokutsurugi, 1740 m
Arabidopsis kamchatica subsp. *kamchatica*	C	Mountains in Western Honshu and Shikoku, Japan	kamC12	Japan, Tokushima, Mt. Shikokutsurugi, 1740 m
Arabidopsis kamchatica subsp. *kamchatica*	C	Mountains in Western Honshu and Shikoku, Japan	kamC13	Japan, Tottori, Daisenji, 580 m
Arabidopsis kamchatica subsp. *kamchatica*	C	Mountains in Western Honshu and Shikoku, Japan	kamC14	Japan, Tottori, Mt. Daisen, 1600 m
Arabidopsis kamchatica subsp. *kamchatica*	D	Mountains in Central Honshu, Japan†	kamD15	Japan, Toyama, along Jintsu River at Toyama airport, 30 m
Arabidopsis kamchatica subsp. *kamchatica*	D	Mountains in Central Honshu, Japan	kamD16	Japan, Ishikawa, Mt. Hakusan, Ichinose, 1080 m
Arabidopsis kamchatica subsp. *kamchatica*	D	Mountains in Central Honshu, Japan	kamD17	Japan, Toyama, Mt. Shirouma, 2800 m
Arabidopsis kamchatica subsp. *kamchatica*	D	Mountains in Central Honshu, Japan	kamD18	Japan, Toyama, Tsurugigozen, 2740 m
Arabidopsis kamchatica subsp. *kamchatica*	D	Mountains in Central Honshu, Japan	kamD19	Japan, Toyama, Tateyama, Mikurigaike, 2400 m
Arabidopsis kamchatica subsp. *kamchatica*	D	Mountains in Central Honshu, Japan	kamD20	Japan, Toyama, Tateyama, Midorigaike, 2400 m
Arabidopsis kamchatica subsp. *kamchatica*	D	Mountains in Central Honshu, Japan	kamD21	Japan, Toyama, Kurobe-dam, 1500 m
Arabidopsis kamchatica subsp. *kamchatica*	D	Mountains in Central Honshu, Japan	kamD22	Japan, Nagano, Kamikochi, Myojin, 1520 m
Arabidopsis kamchatica subsp. *kamchatica*	D	Mountains in Central Honshu, Japan	kamD23	Japan, Nagano, Kamikochi, Shimomatashirodani-deai, 1570 m
Arabidopsis kamchatica subsp. *kamchatica*	D	Mountains in Central Honshu, Japan	kamD24	Japan, Yamanashi, Mt. Kitadake, 3090 m
Arabidopsis kamchatica subsp. *kamchatica*	D	Mountains in Central Honshu, Japan	kamD25	Japan, Shizuoka, Mt. Fuji, Subashiri, 1300 m
Arabidopsis kamchatica subsp. *kamchatica*	E	Lowland in Northern Honshu, Japan	kamE26	Japan, Niigata, Tsugawa
Arabidopsis kamchatica subsp. *kamchatica*	F	Hokkaido, Japan	kamF27	Japan, Hokkaido, Takinoue
Arabidopsis kamchatica subsp. *kamchatica*	F	Hokkaido, Japan	kamF28	Japan, Hokkaido, Asahikawa, Sounkyo, 640 m
Arabidopsis kamchatica subsp. *kamchatica*	F	Hokkaido, Japan	kamF29	Japan, Hokkaido, Kushiro, Obirashike, 20 m
Arabidopsis kamchatica subsp. *kamchatica*	F	Hokkaido, Japan	kamF30	Japan, Hokkaido, Kushiro, Kombumori, 5 m
Arabidopsis kamchatica subsp. *kamchatica*	G	Far East Russia	kamG31	Russia, Kamchatskii krai, Nachiki, basin of the Nachikinskoe ozero lake, 400 m
Arabidopsis kamchatica subsp. *kamchatica*	G	Far East Russia	kamG32	Russia, Kamchatskii krai, near the road from Petropavlovsk Kamchatskii to Esso, 260 m
Arabidopsis kamchatica subsp. *kamchatica*	G	Far East Russia	kamG33	Russia, Kamchatskii krai, Ganaly, close to the bridge over the river Vaktan Malkinskii, 300 r
Arabidopsis kamchatica subsp. *kamchatica*	G	Far East Russia	kamG34	Russia, Kamchatskii krai, Pushchino, close to the bridge over the river Denokhonok, 265 m
Arabidopsis kamchatica subsp. *kamchatica*	G	Far East Russia	kamG35	Russia, Kamchatskii krai, Petropavlovsk Kamchatskii, Mishenaya gora, 10 m
Arabidopsis kamchatica subsp. *kamchatica*	G	Far East Russia	kamG36	Russia, Kamchatskii krai, Elizovo, 70 m
Arabidopsis kamchatica subsp. *kamchatica*	G	Far East Russia	kamG37	Russia, Kamchatskii krai, Nachiki, Mt. Nachikinskoe zerkaltse, 730 m
Arabidopsis kamchatica subsp. *kamchatica*	G	Far East Russia	kamG38	Russia, Kamchatskii krai, Petropavlovsk Kamchatskii, Avachinskaya sopka, 640 m
Arabidopsis kamchatica subsp. *kamchatica*	G	Far East Russia	kamG39	Russia, Kamchatskii krai, Ganaly, close to the bridge over the river Vaktan Ganal'skii, 320 n
Arabidopsis kamchatica subsp. *kamchatica*	G	Far East Russia	kamG40	Russia, Kamchatskii krai, Pushchino, close to the bridge over the river Pravaya Kamchatka,
Arabidopsis kamchatica subsp. *kamchatica*	G	Far East Russia	kamG41	Russia, Kamchatskii krai, Esso, Srednii kamchatskii khrebet, 910 m
Arabidopsis kamchatica subsp. *kamchatica*	H	Alaska	kamH42	USA, Alaska, Kenai, 300 m
Arabidopsis kamchatica subsp. *kamchatica*	H	Alaska	kamH43	USA, Alaska, Chugach State Park, Potter, 5 m
Arabidopsis kamchatica subsp. *kamchatica*	H	Alaska	kamH44	USA, Alaska, Healy
Arabidopsis kamchatica subsp. *kamchatica*	H	Alaska	kamH45	USA, Alaska, Chena River, Chena Hot Springs Rd.
Arabidopsis kamchatica subsp. *kamchatica*	H	Alaska	kamH46	USA, Alaska, Richardson Highway, South of Darling Creek bridge
Arabidopsis kamchatica subsp. *kamchatica*	H	Alaska	kamH47	USA, Alaska, Portage Bay Rd
Arabidopsis kamchatica subsp. *kamchatica*	I	Canada and Washington	kamI48	Canada, Yukon, Mush Lake, 670 m
Arabidopsis kamchatica subsp. *kamchatica*	I	Canada and Washington	kamI49	USA, Washington, Mt. Baker
Arabidopsis kamchatica subsp. *kamchatica*	I	Canada and Washington	kamI50	USA, Washington, Mt. Baker
Arabidopsis lyrata subsp. *lyrata*			lyrlyr1	USA, North Carolina, Pores Knob, ca. 780 m
Arabidopsis lyrata subsp. *lyrata*			lyrlyr2	USA, North Carolina, Pores Knob, ca. 780 m
Arabidopsis lyrata subsp. *petraea*			lyrpet1	Russia, Karhumaki
Arabidopsis lyrata subsp. *petraea*			lyrpet2	Russia, Karhumaki
Arabidopsis lyrata subsp. *petraea*			lyrpet3	Germany, Stolberg, 300 m
Arabidopsis lyrata subsp. *petraea*			lyrpet4	Russia, Yakutiya (Sakha Republic), alluvium of Kolyma, banks of Suharnaya river
Arabidopsis lyrata subsp. *petraea*			lyrpet5	Russia, Yakutiya (Sakha Republic), alluvium of Kolyma, banks of Suharnaya river
Arabidopsis halleri subsp. *gemmifera*			halgem1	Japan, Hyogo, Taka, Omoide River, 200 m
Arabidopsis halleri subsp. *gemmifera*			halgem2	Japan, Osaka, Inagawa, Tadaginzan, 140 m
Arabidopsis halleri subsp. *gemmifera*			halgem3	Russia, Kamchatskii krai, Esso, Srednii kamchatskii khrebet, 660 m
Arabidopsis halleri subsp. *gemmifera*			halgem4	Russia, Kamchatskii krai, Esso, Srednii kamchatskii khrebet, 520 m
Arabidopsis halleri subsp. *gemmifera*			halgem5	Russia, Kamchatskii krai, Esso, valley of the river Ulavkavchan, 470 m
Arabidopsis halleri subsp. *gemmifera*			halgem6	Russia, Kamchatskii krai, Esso, valley of the river Ulavkavchan, 470 m
Arabidopsis halleri subsp. *gemmifera*			halgem7	Russia, Kamchatskii krai, Esso, 580 m
Arabidopsis halleri subsp. *gemmifera*			halgem8	Japan, Nagano, Kamikochi, Shimomatashirodani-deai, 1570 m
Arabidopsis halleri subsp. *dacica*			haldac1	Romania, Fagaras, Mts., Saua Caprei glacial lake, 2270 m
Arabidopsis halleri subsp. *tatrica*			haltat1	Slovakia, Belianske Tatry, 1200 m
Arabidopsis halleri subsp. *tatrica*			haltat2	Slovakia, Vysoke Tatry, 1800 m
Arabidopsis halleri subsp. *tatrica*			haltat3	Slovakia, Slovensky Raj, 1000 m
Arabidopsis halleri subsp. *ovirensis*			halovi1	Austria, Carinthia, Ebriach, 1557 m
Arabidopsis halleri subsp. *halleri*			halhal1	Switzerland, Ticino, Giubiasco, 400 m
Arabidopsis halleri subsp. *halleri*			halhal2	Switzerland, Ticino, Giubiasco, 400 m
Arabidopsis pedemontana			ped	Italy, north of Valle Po, north of Crissolo, Colle delle Porte, 2260 m

*two ITS haplotypes are obtained from this sample because one-base pair indel was identified.

** At trnL intron: a difference in a simple repeat, that was not considred in the analyses.

† Mountains in Central Honshu include Japan Alps, Mt. Fuji and Mt. Hakusan.

Selfing: n.a. not assayed, y: successful

Voucher specimens are deposited in the herbaria KYO, SAV, Z, and SHO (herbaria of Shoei Junior College, Kobe, Japan) when available.

Collector	selfing	*WER*-lyr	*WER*-hal	*CHS*-lyr	*CHS*-hal	ITS	cpDNA haplotype
H. Tsukaya	n.a.	kamA1L	kamA1H	kamA1L	kamA1H	kamA1	cpHap6
H. Tsukaya	n.a.	kamA2L	kamA2H	kamA2L	kamA2H	kamA2	cpHap6
H. Tsukaya	n.a.	kamA3L	kamA3Ha1, a2	kamA3L	kamA3H	kamA3a1, a2*	cpHap7
HK	n.a.	kamkwsB4L	kamkwsB4H	kamkwsB4L	kamkwsB4H	kamkwsB4	cpHap2
HK, KKS	n.a.	kamkwsB5L	kamkwsB5H	kamkwsB5L	kamkwsB5H	kamkwsB5	cpHap2
S. Fujii, KKS	y	kamkwsB6L	kamkwsB6H	kamkwsB6L	kamkwsB6H	kamkwsB6	cpHap4
S. Fujii, KKS	y	kamkwsB7L	kamkwsB7H	kamkwsB7L	kamkwsB7H	kamkwsB7	cpHap4
S. Fujii, KKS	y	kamkwsB8L	kamkwsB8H	kamkwsB8L	kamkwsB8H	kamkwsB8	cpHap2
HK	n.a.	kamkwsB9L	kamkwsB9H	kamkwsB9L	kamkwsB9H	kamkwsB9	cpHap4
HK	y	kamkwsB10L	kamkwsB10H	kamkwsB10L	kamkwsB10H	kamkwsB10	cpHap1
M. Kanaoka, KKS	y	kamC11L	kamC11H	kamC11L	deletion	kamC11	cpHap1
M. Kanaoka, KKS	y	kamC12L	kamC12H	kamC12La1, a2	deletion	kamC12	cpHap1
KKS	y	kamC13L	kamC13H	kamC13L	kamC13H	kamC13	cpHap3
KKS	n.a.	kamC14L	kamC14H	kamC14L	kamC14H	kamC14	cpHap3
HK, J. Sugisaka	y	kamD15L	kamD15H	kamD15L	kamD15H	kamD15	cpHap3
KKS	n.a.	kamD16L	kamD16H	kamD16L	kamD16H	kamD16	cpHap3
M. Kanaoka, KKS	n.a.	kamD17L	kamD17H	kamD17La1, a2	kamD17H	kamD17	cpHap1
KKS	n.a.	kamD18L	kamD18H	kamD18L	kamD18H	kamD18	cpHap5
KKS	n.a.	kamD19L	kamD19H	kamD19L	kamD19H	kamD19	cpHap3
KKS	y	kamD20L	kamD20H	kamD20L	kamD20H	kamD20	cpHap3
KKS	n.a.	kamD21L	kamD21H	kamD21L	kamD21H	kamD21	cpHap3
KKS	y	kamD22L	kamD22H	kamD22L	kamD22H	kamD22	cpHap3
HK	n.a.	kamD23L	kamD23H	kamD23L	deletion	kamD23	cpHap2
KKS	n.a.	kamD24L	kamD24H	kamD24L	kamD24Ha1, a2	kamD24	cpHap3
HK	y	kamD25L	kamD25H	kamD25L	kamD25H	kamD25	cpHap3
A. Kawabe	y	kamE26L	kamE26H	kamE26L	kamE26H	kamE26	cpHap1
KKS	n.a.	kamF27L	kamF27H	kamF27L	kamF27H	kamF27	cpHap3
H. Nakai, KKS	y	kamF28L	kamF28H	kamF28L	kamF28H	kamF28	cpHap3
KKS	n.a.	kamF29L	kamF29H	kamF29L	kamF29H	kamF29	cpHap3
KKS	y	kamF30L	kamF30H	kamF30L	kamF30H	kamF30	cpHap3
KM, VVY	n.a.	kamG31L	kamG31H	kamG31L	kamG31H	kamG31	cpHap1
KM, VVY, HK, RSI, KKS	n.a.	kamG32L	kamG32H	kamG32L	kamG32H	kamG32	cpHap1
KM, VVY, HK, RSI, KKS	n.a.	kamG33L	kamG33H	kamG33L	kamG33H	kamG33	cpHap1
KM, VVY, HK, RSI, KKS	n.a.	kamG34L	kamG34H	kamG34L	kamG34H	kamG34	cpHap1
KM, VVY, HK, RSI, KKS	n.a.	kamG35L	kamG35H	kamG35L	kamG35H	kamG35	cpHap1
KM, VVY	n.a.	kamG36L	kamG36H	kamG36L	kamG36H	kamG36	cpHap1**
KM, VVY	n.a.	kamG37L	kamG37H	kamG37L	kamG37H	kamG37	cpHap1
KM, VVY	n.a.	kamG38L	kamG38H	kamG38L	kamG38H	kamG38	cpHap1
KM, VVY, HK, RSI, KKS	n.a.	kamG39L	kamG39H	kamG39L	kamG39H	kamG39	cpHap1
KM, VVY, HK, RSI, KKS	n.a.	kamG40L	kamG40H	kamG40L	kamG40H	kamG40	cpHap1
KM, VVY, HK, RSI, KKS	n.a.	kamG41L	kamG41H	kamG41L	kamG41H	kamG41	cpHap1
A. Caicedo	n.a.	kamH42L	kamH42H	kamH42L	kamH42H	kamH42	cpHap1
KKS	y	kamH43L	kamH43H	kamH43L	kamH43H	kamH43	cpHap1
KKS	n.a.	kamH44L	kamH44H	kamH44L	kamH44H	kamH44	cpHap1
OS	y	kamH45L	kamH45H	kamH45L	kamH45H	kamH45	cpHap1
OS	y	kamH46L	kamH46H	kamH46L	kamH46H	kamH46	cpHap1
OS	y	kamH47L	kamH47H	kamH47L	kamH47H	kamH47	cpHap1
H. Schöb	n.a.	kamI48L	kamI48H	kamI48L	kamI48H	kamI48	cpHap1
OS	n.a.	kamI49L	kamI49H	kamI49L	kamI49H	kamI49	cpHap1
OS	n.a.	kamI50L	kamI50H	kamI50L	kamI50H	kamI50	cpHap1
KKS	n.a.	lyrlyr1		lyrlyr1a1, a2		lyrlyr1	cpHap8
KKS	y	lyrlyr2		lyrlyr2		lyrlyr2	cpHap8
OS	n.a.	lyrpet1				lyrpet1	cpHap9
OS	n.a.			lyrpet2		lyrpet2	cpHap9
M. Clauss, KKS	n.a.	lyrpet3		lyrpet3		lyrpet3	cpHap10
VVY	n.a.	lyrpet4		lyrpet4		lyrpet4	cpHap11
VVY	n.a.	lyrpet5		lyrpet5		lyrpet5	cpHap11
T. Kawagoe, KKS	y		halgem1		halgem1	halgem1	cpHap1
KKS	y		halgem2		halgem2	halgem2	cpHap1
KM, VVY, HK, RSI, KKS	n.a.		halgem3		halgem3	halgem3	cpHap2
KM, VVY, HK, RSI, KKS	n.a.		halgem4		halgem4	halgem4	cpHap2
KM, VVY, HK, RSI, KKS	n.a.		halgem5		halgem5	halgem5	cpHap2
KM, VVY, HK, RSI, KKS	n.a.		halgem6		halgem6	halgem6	cpHap2
KM, VVY, HK, RSI, KKS	n.a.		halgem7		halgem7	halgem7	cpHap5
KM, HK	n.a.		halgem8		halgem8	halgem8	cpHap3
M. Kolnik	n.a.		haldac1			haldac1	cpHap12
KM, JL, HK, KKS	n.a.		haltat1		haltat1a1, a2	haltat1	cpHap13
KM, JL, HK, KKS	n.a.		haltat2		haltat2	haltat2	cpHap14
KM, JL, HK, KKS	n.a.		haltat3			haltat3	cpHap15
M. Kolnik	n.a.		halovi1		halovi1	halovi1	cpHap16
RSI, KKS, T. Tsuchimatsu, M. Helling	n.a.		halhal1a1,a2			halhal1	cpHap17
RSI, KKS, T. Tsuchimatsu, M. Helling	n.a.		halhal2			halhal2	cpHap18
KKS, KM, JL	n.a.		ped		ped	ped	

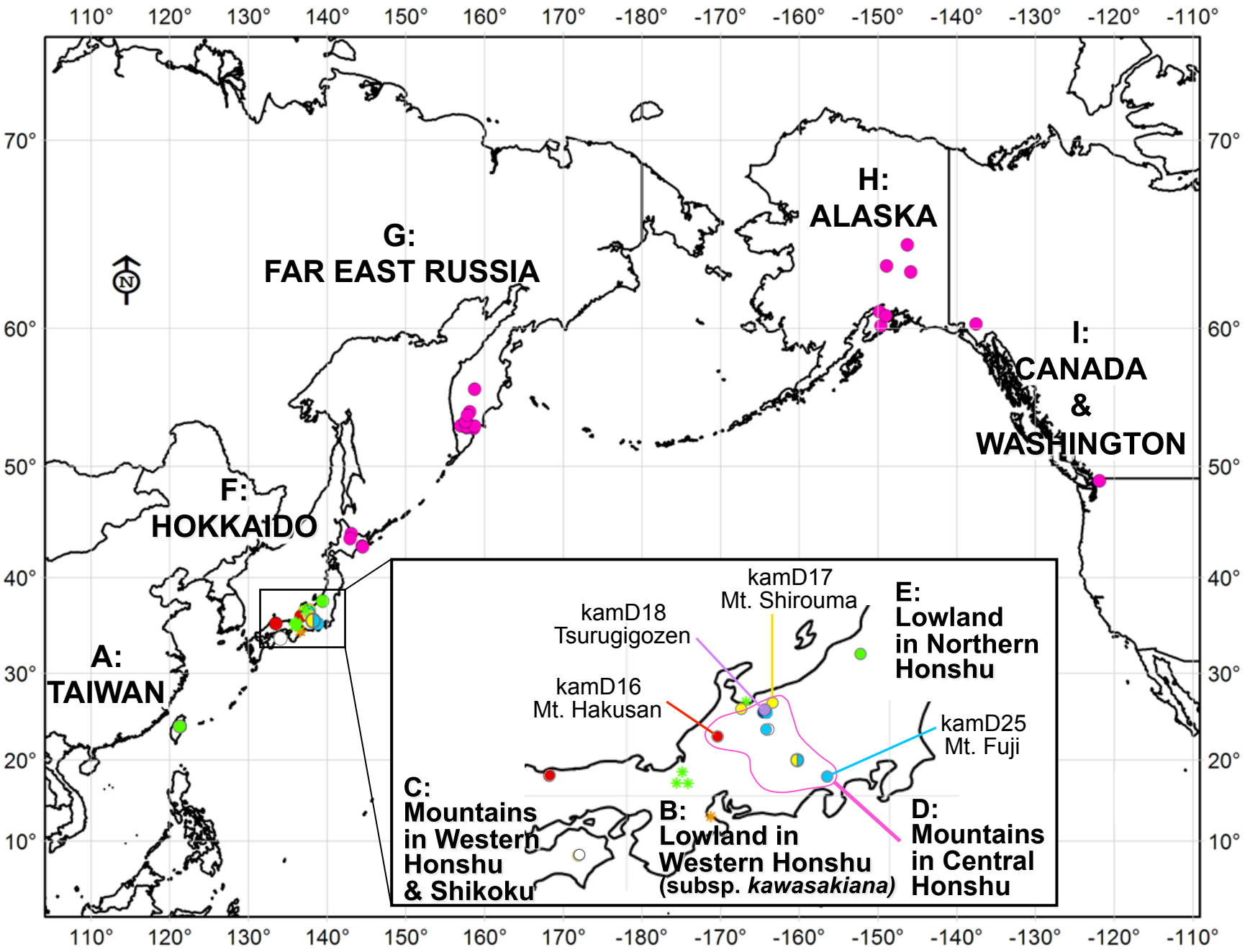


Fig.1 CHS-hal map

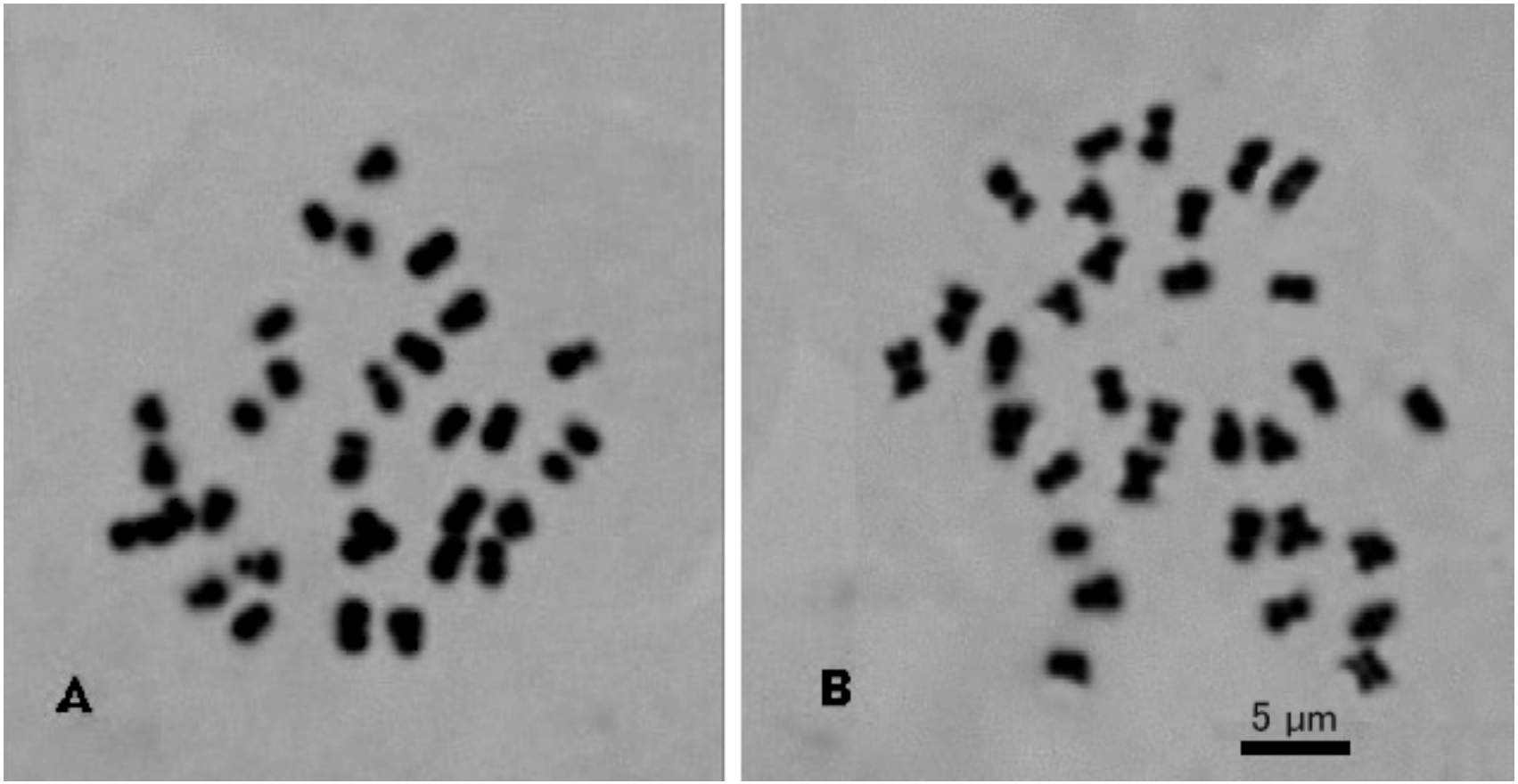
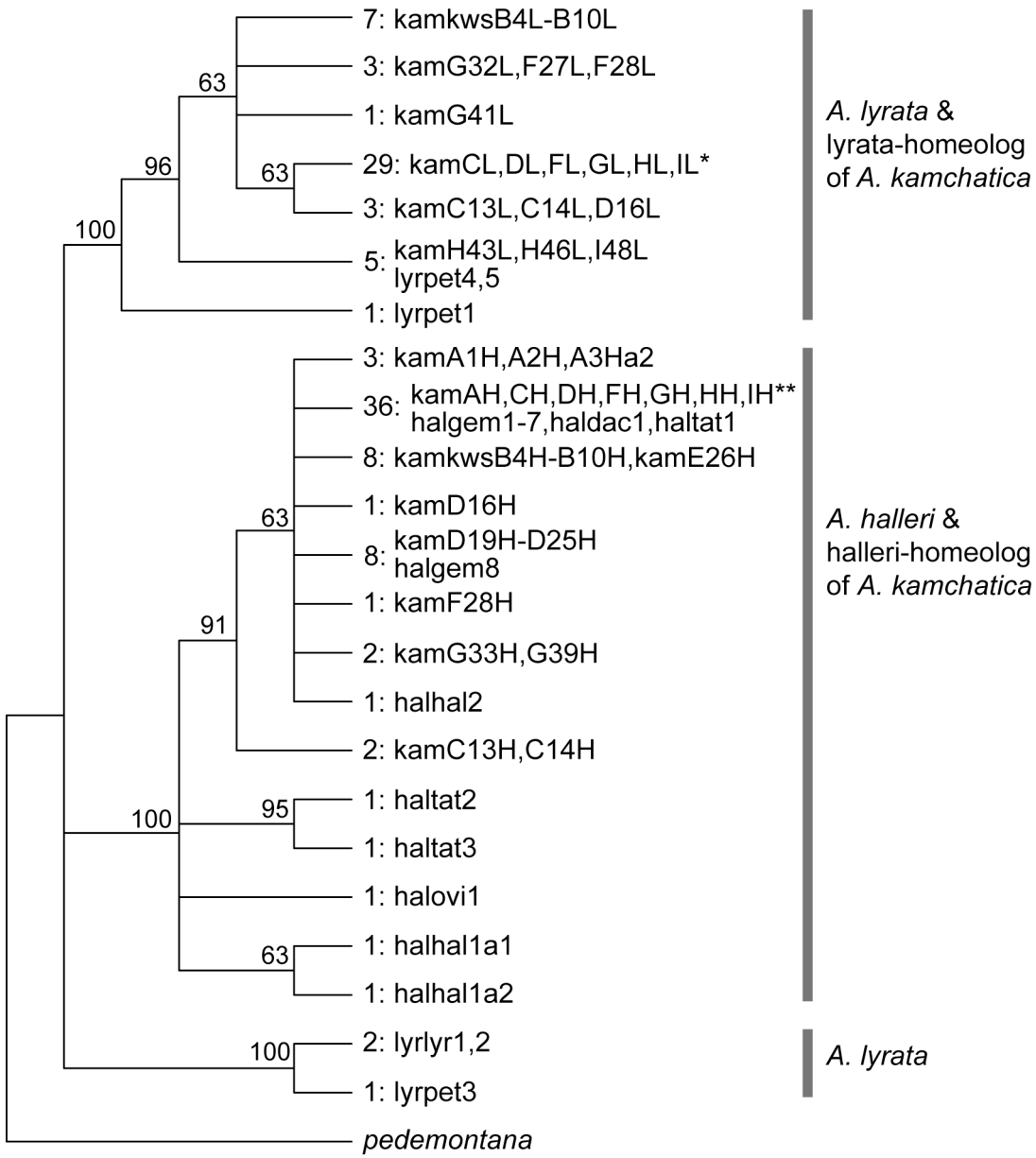


Fig. 2 chromosome



*kamC11L, C12L, D15L, D17L-D25L, F29L, F30L, G31L, G33L-G40L, H42L, H44L, H45L, H47L, I49L, I50L

**kamA3Ha1, C11H, C12H, D15H, D17H, D18H, F27H, F29H, F30H, G31H, G32H, G34H-G38H, G40H, G41H, H42H-47H, I48H-I50H

Fig.3 WER MP tree

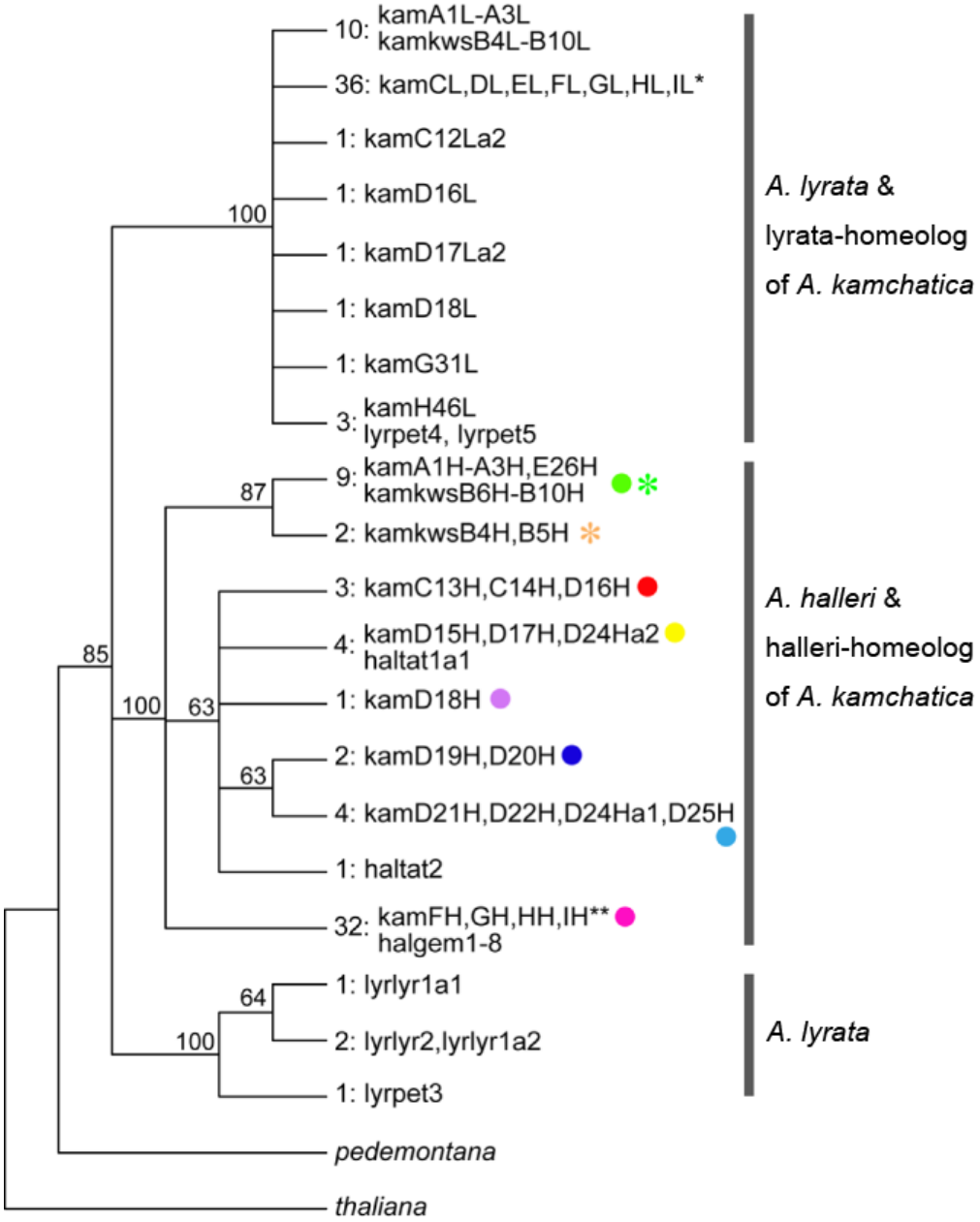
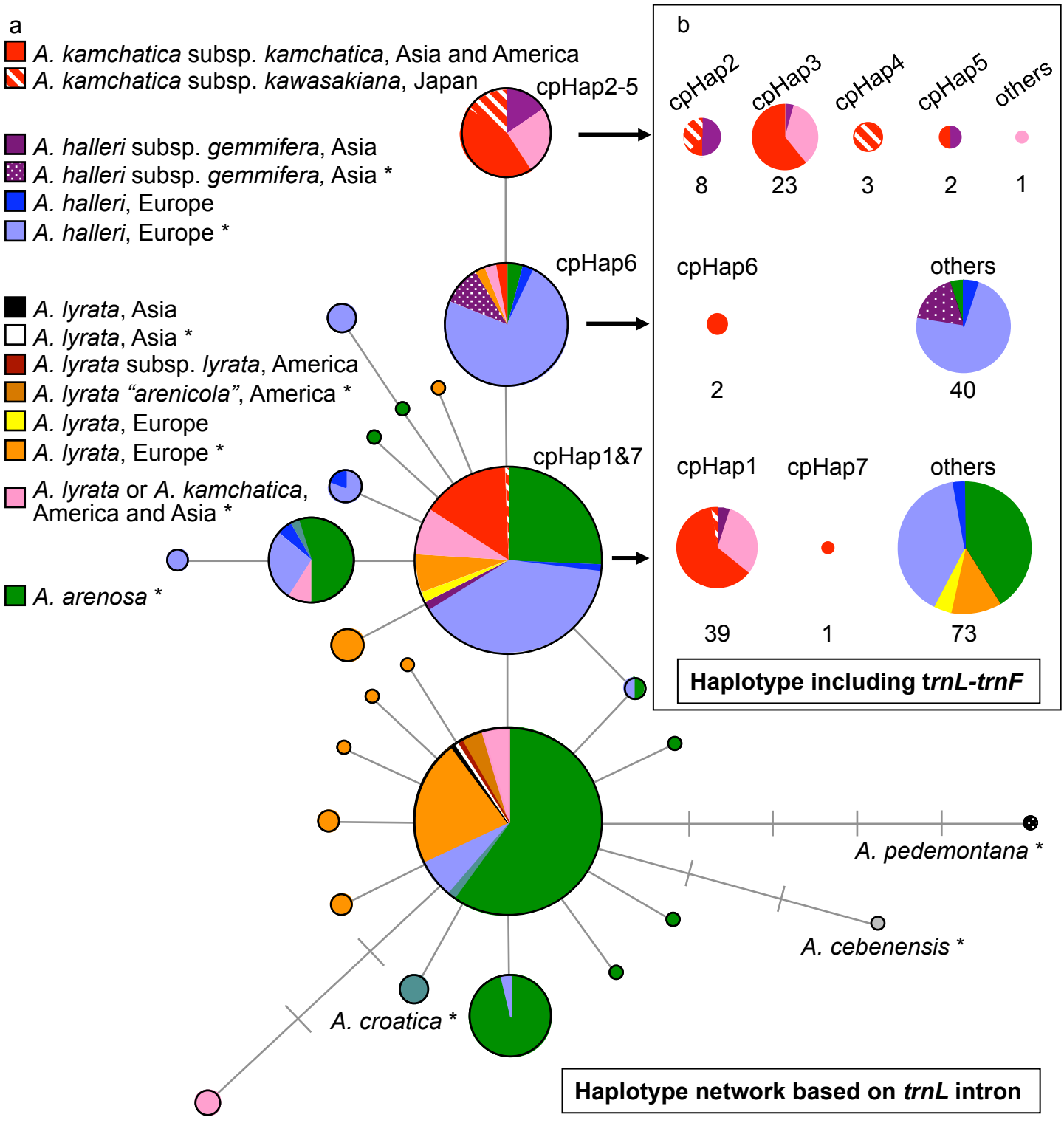


Fig.4 CHS MP tree


a
A. kamchatica subsp. kamchatica, Asia and America
A. kamchatica subsp. kawasakiana, Japan
A. halleri subsp. gemmifera, Asia
A. halleri subsp. gemmifera, Asia *
A. halleri, Europe
A. halleri, Europe *
A. lyrata, Asia
A. lyrata, Asia *
A. lyrata subsp. lyrata, America
A. lyrata "arenicola", America *
A. lyrata, Europe
A. lyrata, Europe *
A. lyrata or A. kamchatica, America and Asia *
A. arenosa *
cpHap2-5
cpHap6
cpHap1&7
A. croatica *
A. cebenensis *
A. pedemontana *
b
cpHap2
cpHap3
cpHap4
cpHap5
others
8
23
3
2
1
cpHap6
others
2
40
cpHap1
cpHap7
others
39
1
73
Haplotype including trnL-trnF
Haplotype network based on trnL intron

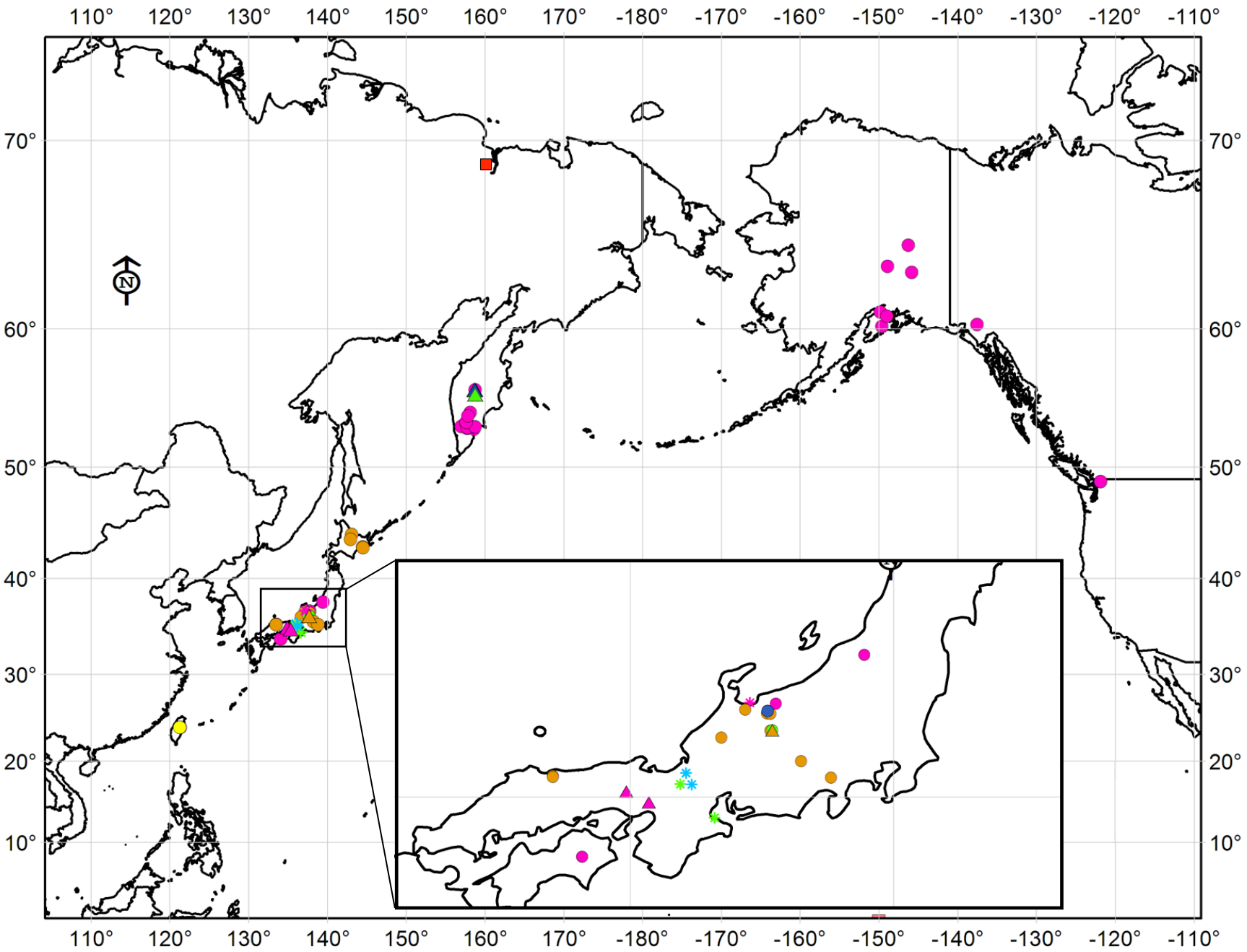
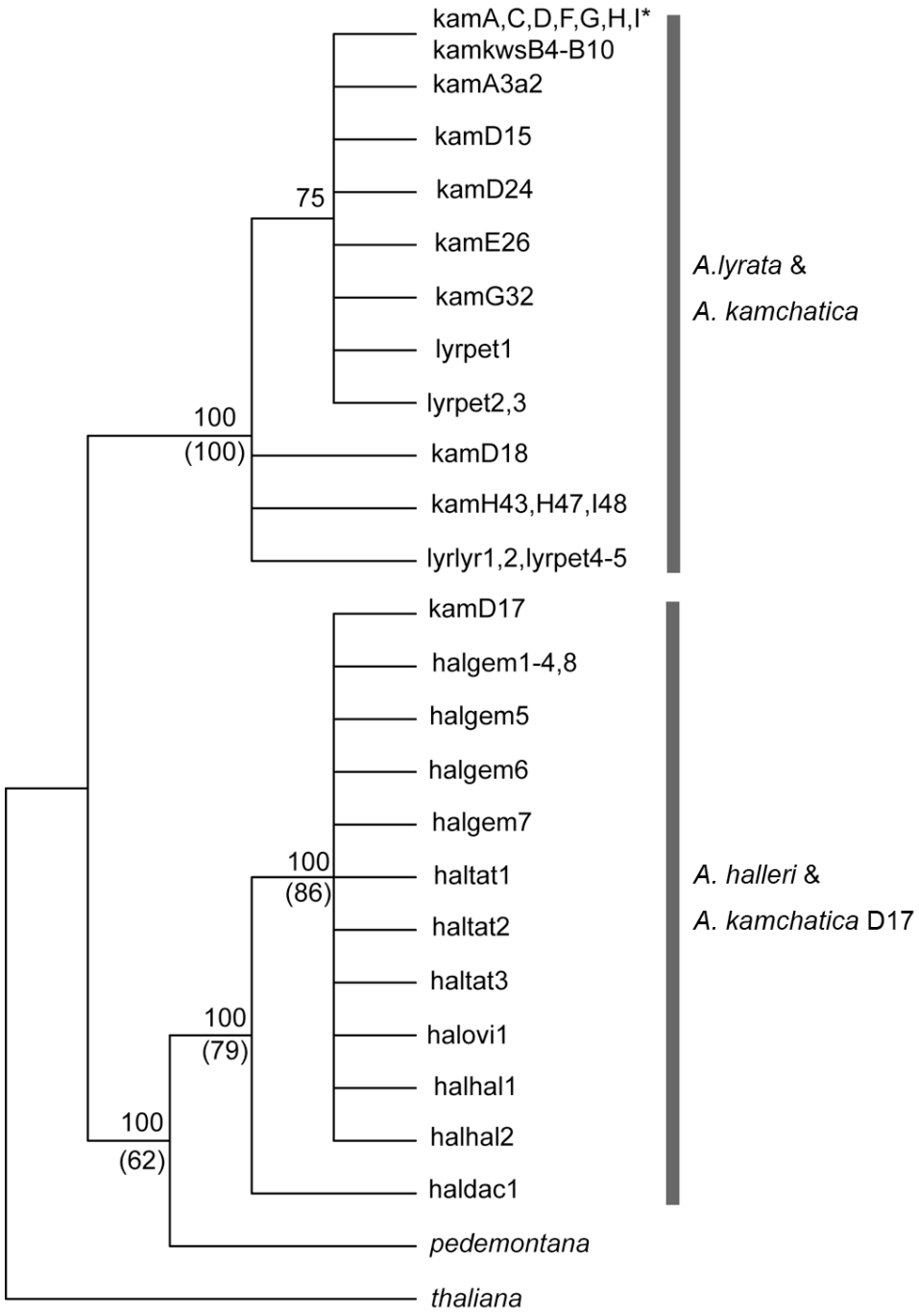


Fig.6 chloroplastHap map



*kamA1, A2, A3a1, C11-C14, D16, D19-D23, D25, F27-F30, G31, G33-G41, H42, H44-H46, I49, I50

Fig. 7 ITS MP

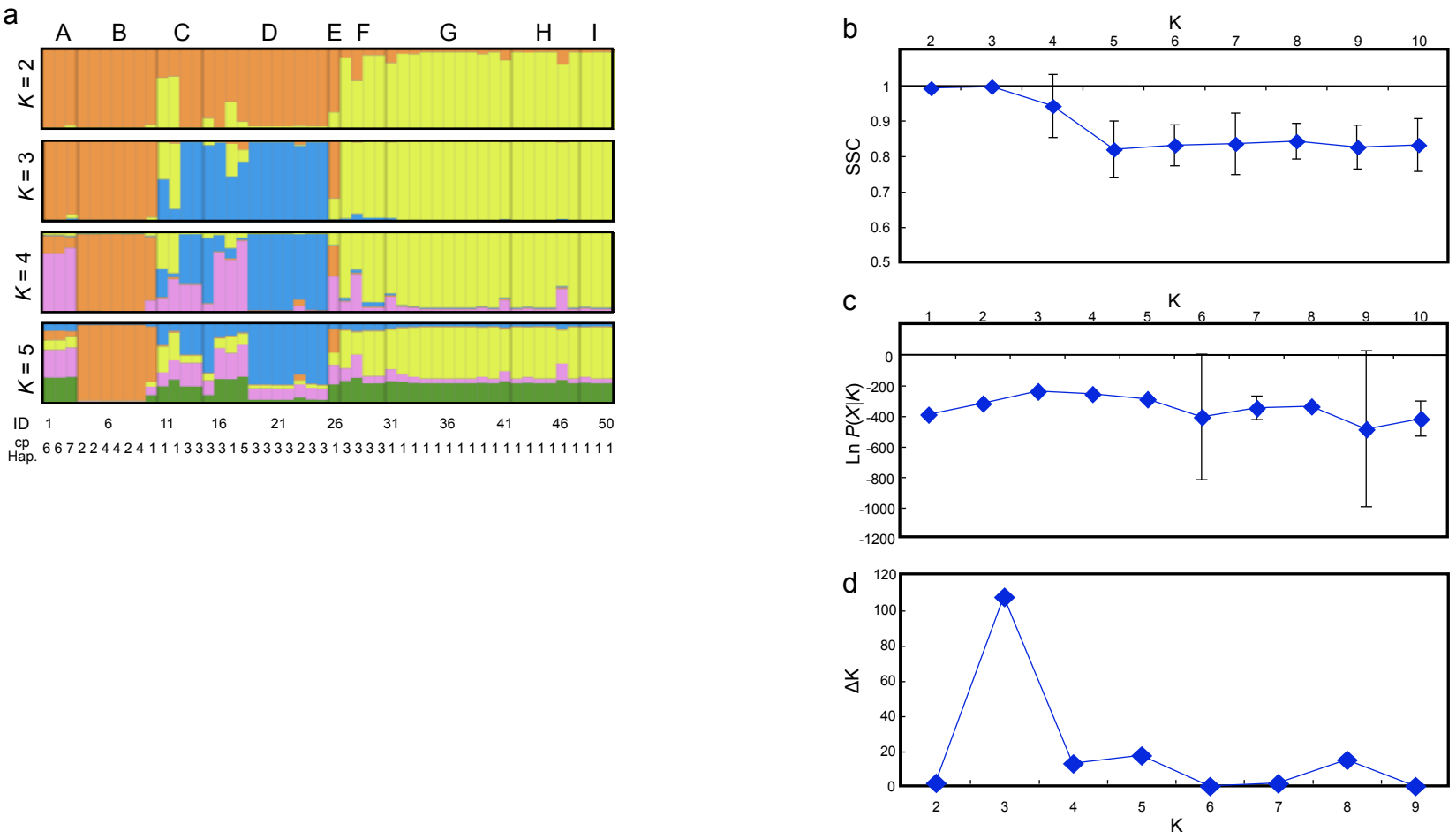


Fig.8 Structure

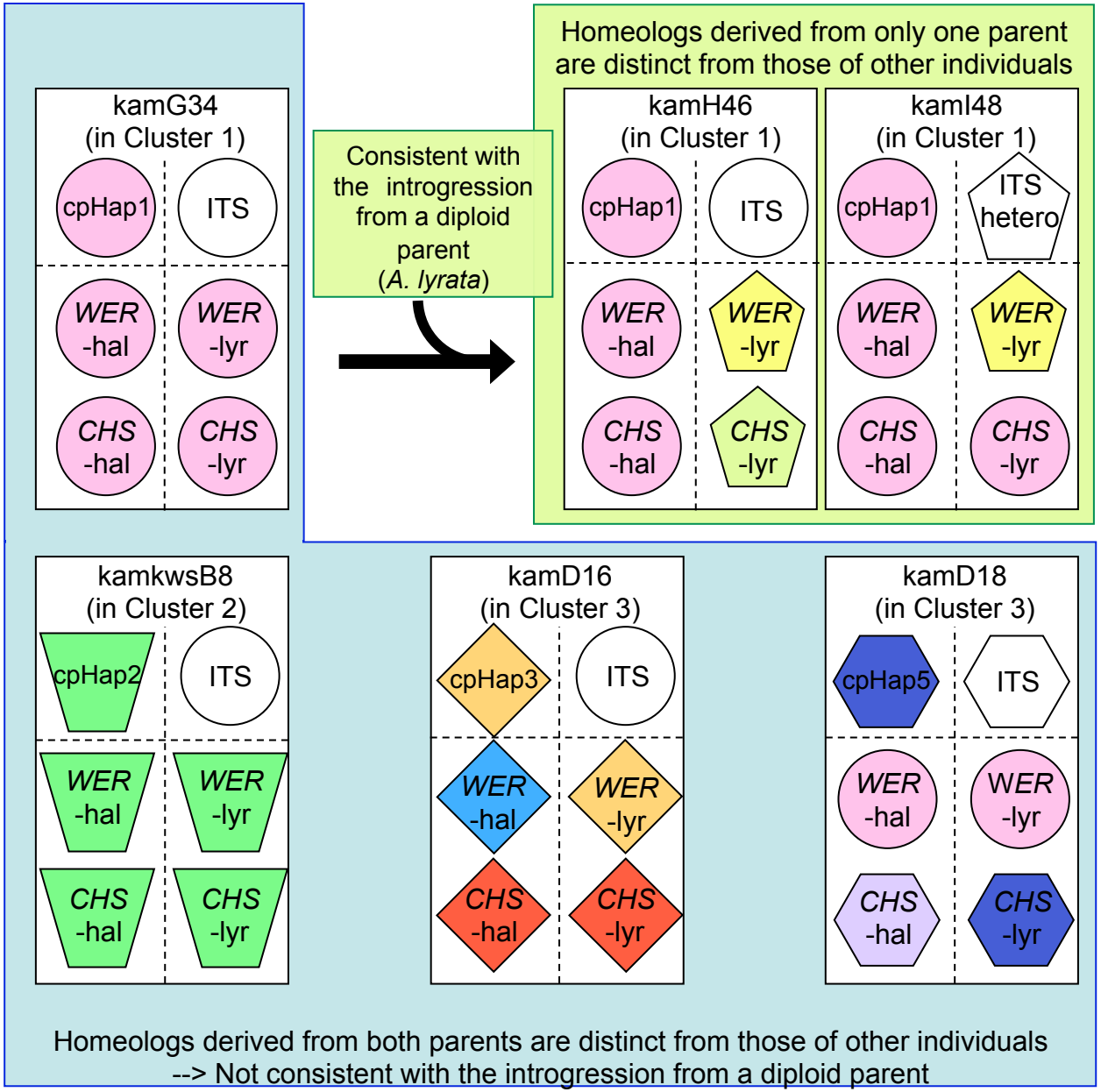

kamG34
(in Cluster 1)
cpHap1
ITS
WER -hal
WER -lyr
CHS -hal
CHS -lyr
Consistent with the introgression from a diploid parent (A. lyrata)
Homeologs derived from only one parent are distinct from those of other individuals
kamH46
(in Cluster 1)
kamI48
(in Cluster 1)
ITS hetero
kamkwsB8
(in Cluster 2)
cpHap2
kamD16
(in Cluster 3)
cpHap3
kamD18
(in Cluster 3)
cpHap5
Homeologs derived from both parents are distinct from those of other individuals
--> Not consistent with the introgression from a diploid parent

Fig. 9

Shimizu-Inatsugi *et al.*
Supporting Information (5 figures, 3 tables, 1 section of text)

Supporting Fig. S1. Primers used for amplification of *WER* and *CHS* genes, and the number of homeologs
a. Each arrow indicates the position of the primer and its direction. Gray boxes represent the coding region of each gene. See Supporting Table S1 for the details of the primers used.
b. Amplification of *WER*-hal and *WER*-lyr using primers WERF4-R3e1. The primers were designed in the conserved region so that they amplified both *WER*-hal and *WER*-lyr (e.g., E26, G31 and B4). Two homeologs of C13 and C14 had the same length, and see Text S1.
c. Digestion of PCR fragments to confirm the homeolog type. Each fragment was amplified with primers CHSkamF1-CHSR3 and digested by *Xba*I. The primers were designed in the conserved region so that they amplified both halleri- and lyrata-derived homeologs of *CHS*. *CHS*-hal but not *CHS*-lyr was digested, resulting in shorter bands.

Supporting Fig. S2. Segregating sites of the *CHS* gene and ITS region, and sequences of distinct cpDNA haplotypes
a, b. Summary of the segregating sites among the **a.** *CHS* gene and **b.** ITS region from *A. kamchatica*, *A. lyrata*, and *A. halleri*. Dots indicate the nucleotide identical to that of the upper row sequence, - indicates the deletion site and N indicates the site where the sequence was not obtained. The haplotype names correspond to those in Table 1. In ITS region, IUPAC ambiguity codes were used for coding polymorphic positions. In three individuals of North American *A. kamchatica* (kamH43, kamH47 and kamI48), three heterozygous sites were observed, which is consistent with the overlap of two haplotypes: a common haplotype in *A. kamchatica*, and a haplotype identical to the haplotype observed in four individuals of *A. lyrata*.
c. Sequences of the *trnL* intron and *trnL-trnF* intergenic spacer regions of distinct 18 cpDNA haplotypes (cpHap1–18) observed in *A. kamchatica*, *A. lyrata*, and *A. halleri*. The cpHap1* from the individual kamG36 has one additional nucleotide T at a polyT site of the *trnL* intron region, shown as * at the 300th site compared with cpHap1. This difference between cpHap1 and cpHap1* was ignored in all analyses. The alignment of *trnL-trnF* region was not clear due to tandem duplications. See Table 1 for the correspondence between individuals and haplotypes.

Supporting Fig. S3. Phylogenetic trees of *WER, CHS* and *ITS*
Accession abbreviations follow Table 1.
a. A single most-parsimonious tree obtained from the analysis based on nuclear *CHS* sequence data of the lyrata-clade. The data matrix includes 1589 aligned positions and additional coding of eight indels, i.e., it is longer than that of Fig. 4 because of inclusion of the longer promoter region, which was lacking or not amplified in the halleri clade. Bootstrap values above 50% are shown along the branches. The number to the right of each branch indicates how many sequences are included in the branch. The colored symbols correspond to that in Fig. S4c. The tree displays 14 unique sequences, representing 57 sequences obtained from 55 individuals.
b. Majority-rule consensus tree of the Bayesian inference based on nuclear *WER* sequence data (955 aligned positions). The posterior probability values of the nodes are indicated above the branches. The colored symbols correspond to that in either of Fig. S4a (for *WER*-lyr) or Fig. S4b (for *WER*-hal). *WER*-lyr of E26L (dark blue in Fig. S4a) and A1-3L (dark green in Fig. S4a) are not included in the tree because of the existence of long indels. The tree displays 24 unique sequences, representing 120 sequences obtained from 72 individuals.

c. Majority-rule consensus tree of the Bayesian inference based on nuclear *CHS* sequence data (1314 aligned positions). The posterior probability values of the nodes are indicated next to the branches. The tree displays 22 unique sequences, representing 118 sequences obtained from 68 individuals. Accession abbreviations follow Table 1.
d. Majority-rule consensus tree of the Bayesian inference based on nuclear ITS sequence data. The posterior probability values of the nodes are indicated above the branches. The tree displays 25 unique sequences, representing 75 individuals.

Supporting Fig. S4. Geographic distribution of *WER* and *CHS* haplotypes
Haplotype maps showing the geographic distribution of each haplotype of **a.** *WER*-lyr, **b.** *WER*-hal and **c.** *CHS*-lyr. Circles indicate *A. kamchatica* subsp. *kamchatica* and asterisks indicate subsp. *kawasakiana*. Heterozygotes are shown as half-circles. The upper map shows the Pacific Ocean rim, and the lower magnified map shows the Japanese archipelago. Different haplotypes of each homeolog are depicted in different colors. The color symbols in **a**, **b** and **c** correspond to those in the *WER*-lyr clade in Fig. S3b, *WER*-hal clade in Fig. S3b and *CHS*-lyr in Fig. S3a, respectively.

Supporting Fig. S5. Neighbor-joining tree of *CHS* including publicly available data
The evolutionary history was inferred using the Neighbor-Joining method. The optimal tree with the sum of branch length = 0.11702052 is shown. In addition to our data, published sequences, namely, seven sequences of *A. lyrata* subsp. *lyrata* from America (noted as lyrlyr AL), four sequences of *A. lyrata* subsp. *petraea* from Europe (as lyrpet AP), and eleven sequences of *A. halleri* from Europe (as hal CH) were included. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) is shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree.

Supporting Text 1. Sequencing, alignments, phylogenetic analyses, intrapopulation polymorphism, and Bayesian clustering

DNA extraction and sequencing
We conducted one round of selfing using 18 individuals to propagate, and repeated up to four rounds for subspecies *kawasakiana* with a short life-cycle. Genomic DNA was isolated from young leaves using the DNeasy Plant Mini kit (Qiagen). DNA sequencing was conducted at the Institute of Plant Biology, University of Zurich, with a Prism 3730 48-capillary automated sequencer (Applied Biosystems). The sequence alignments were done in Biolign version 4.0.6.2 (http://www2.maizegenetics.net/index.php?page=bioinformatics/index.html) and edited manually using the program BioEdit version 7.0.4.1 (Hall 1999). Ambiguous polymorphisms were rechecked with PCR reamplification and sequencing. The two cpDNA regions were concatenated and analyzed as a haplotype (called superhaplotype by Koch & Matschinger 2007).

Primer design
We designated the genes of polyploids, which are orthologs of "single-copy nuclear genes" in the parental species, as "low-copy nuclear genes" of polyploids. In contrast, the nuclear ribosomal ITS region is tandemly repeated and is not a low-copy nuclear gene. The PCR primers used in the study are listed in Table S1 and are shown schematically in Fig. S1. Design of the homeolog-specific primers and PCR amplification were conducted using the methods described by Lihova *et al.* (2006). Because of redundancy of genes in polyploid

species, homeologs often exhibit rearrangements or gene loss. In addition, the indels and SNPs used for the design of homeolog-specific primers are often polymorphic, even among individuals in the species. Thus, homeolog-specific primers designed based on a particular individual often yield nonspecific or unsuccessful amplification in other individuals, as we reported in the study of hexaploid *Cardamine asarifolia* (Lihova *et al.* 2006). In the present study, we designed multiple homeolog-specific primers to amplify the *WER* and *CHS* genes.

To amplify *WER* homeologs, we designed three primers in the 5′-upstream region and six primers in the 3′-downstream region based on the genome sequence of the closely related species *A. thaliana*. The PCR product by WERFU3 and WERRD6 turned out to be lyrata-homeolog specific. The PCR products by WERFU1 and WERRD1 included both homeologs. The product was sequenced directly using WERRD1 primer and yielded single peak sequences followed by double peaks. Using the methods described in Figure 2 of Lihova *et al.* (2006), we identified a 4-bp indel polymorphism and designed the halleri-homeolog specific primer WERgemRD1 at this position. Although sequencing of several individuals was successful using this primer, PCR amplification was weak in many individuals. Thus, based on the 3-bp indel polymorphism in the third exon, another specific primer, WERgemR3e2, was designed.

Three individuals from Taiwan did not yield a lyrata-homeolog of *WER* using primers WERFU3 and WERRD6. The forward primer WERlyrF1 in the second intron in combination with the reverse primer WERRD6 yielded specific amplification of a lyrata-homeolog, suggesting that rearrangements occurred in the former half of the gene.

To amplify *CHS* homeologs, Shimizu *et al.* (2005) designed the homeolog-specific primers CHSlyrFU1 and CHSgemFU1, which were designed at ~500 upstream of the start codon, and a common reverse primer CHSR1. In several individuals, these primers yielded low or no amplification, presumably because of rearrangements and mutations. To amplify the halleri-homeolog, the primer CHSgemFU15 was designed at ~150 upstream of the start codon. These primers amplified the halleri-homeolog in 47 of the 50 individuals; the other three individuals had presumable deletions or rearrangements (see below). The primer CHSlyrFU4 was designed to encompass two SNPs near the primer CHSlyrFU1, based on a comparison of the sequence of *A. lyrata* and *A. halleri* ssp. *gemmifera*.

Copy numbers

To verify the deletion and the copy numbers, primers were designed in the conserved sequences to amplify both homeologs together. As for the *WER* homeologs, WERF4 in exon 2 and WERR3e1 in exon 3 were designed to encompass the 54-bp insertion found in most of the halleri-homeologs. PCR yielded a longer band of the halleri-homeolog and a shorter band of lyrata-homeolog in all 50 individuals of *A. kamchatica* (Fig. S1b; e.g., G31 and B4), except for individuals with different indels (kamE26 from Tsugawa and kamC13 and kamC14 from Daisen). In kamE26, the lyrata-homeolog was longer than the halleri-homeolog (Fig. S1b). The PCR products of kamC13 and kamC14 were digested by *Pst*I and were confirmed to have two homeologs (data not shown). As for the *CHS* homeologs, CHSkamF1 in exon 1 and CHSR3 in the exon 2 were designed to encompass an SNP at the *Xba*I restriction site. By digesting the PCR products with *Xba*I, 47 of 50 individuals yielded bands of both lyrata and halleri-homeologs (Fig. S1c; e.g., G31 and B4). Three individuals (kamC11 and kamC12 from Shikokutsurugi and kamD23 from Shimomatashirodani-deai) yielded only one band that corresponded to the lyrata-homeolog. Direct sequencing (as described in Lihova *et al.* 2006) confirmed that only the lyrata-homeolog was amplified. In conjunction with the PCR failure of the halleri-homeolog described above, these data suggest strongly that the halleri-homeolog of *CHS* is deleted or rearranged in the three individuals. In contrast, diploid samples showed only one band (Fig. S1b, c and data not shown). In short, allopolyploidy was supported in all

50 individuals either by *WER* or *CHS*, and rearrangement or deletion occurred in several individuals.

Alignments and phylogenetic analyses

We aimed to achieve as much sample overlap between the individual data sets (*WER*, *CHS* and ITS) as possible, although this was not always possible for several reasons (sequence recombination, PCR failure, deletion, etc.).

MP analyses were conducted with PAUP* version 4.0b10 (Swofford 2001). Heuristic searches were made with the following settings: gaps treated as missing data, single-site polymorphisms as uncertainties, tree construction with stepwise addition, 1,000 replicates with random taxon addition, TBR branch swapping, no MAXTREES limits, and MULTREES option in effect. For character-state optimization, the ACCTRAN (accelerated character transformation) option was used. The most-parsimonious trees generated were summarized in the strict consensus and 50% majority-rule consensus trees. Bootstrap analyses (Felsenstein 1985) were performed using 100,000 resamplings with the fast-heuristic search as implemented in PAUP*.

The Bayesian inference was run using MrBayes version 3.0 beta4 (Huelsenbeck & Ronquist 2001). Four Markov chains were run for 20 million generations while adjusting the temperature difference between the cold and heated chains to achieve efficient swapping between the chains. Six substitution rates (nst = 6) and a gamma distribution (rates = gamma) were assumed. The trees were sampled every 100 generations and, finally, majority-rule consensus trees were computed that excluded the trees found in the burn-in phase (i.e., those generated before the likelihood values reached a plateau and fluctuated within a more or less stable range). The percentage of trees recovering an individual node is indicated on the consensus trees by the node's posterior probability.

WER homeologs

The alignment of *WER* homeologs (spanning from the middle of exon 1 to the middle of exon 3) comprised 120 sequences obtained from 72 individuals (summarized in Table S2). Only *A. pedemontana* was used as the outgroup species because inclusion of *A. thaliana* would introduce a high number of additional indels and complicate the sequence alignment and data analyses. The final alignment of *WER* comprised 24 unique haplotypes and included 955 aligned positions. Fourteen indels longer than 1 bp were introduced in the alignment, coded as additional binary to four-state characters, and included in the MP analyses. In total, 83 sites were variable, and 60 of them were parsimony informative.

MP analyses and Bayesian inferences resulted in very similar tree topologies, and in the following text, we report on the MP results only. The MP analysis of the *WER* data set resulted in a strict consensus tree (78 most-parsimonious trees, $L = 96$ steps, CI = 0.96, RI = 0.99), which displayed three main and well-supported (100% bootstrap) clades (Fig. 3). The *WER* sequences from the diploids *A. lyrata* and *A. halleri* were clearly differentiated from each other (placed in distinct clades). Two apparently different homeologs were retrieved from the tetraploids (*A. kamchatica*, including both subspecies) and placed in the respective clades of the diploids. Among the three clades resolved, one clade comprised all individuals of *A. halleri* and the corresponding homeolog from tetraploids (*WER*-hal), and the other clade comprised the Russian (both western and easternmost Russia) accessions of *A. lyrata* and the other homeolog from tetraploids (*WER*-lyr). Two accessions of *A. lyrata* from USA (North Carolina) and one from Germany of *A. lyrata* formed an additional clade.

Despite our limited sampling of the diploids, both *A. lyrata* and *A. halleri* were found to be more diverse than the tetraploids. We also identified haplotypes identical to those found in some tetraploids. Two accessions of *A. lyrata* from Far East Russia (Yakut, lyrpet4,

5) shared the same haplotype with *A. kamchatica* from Alaska and Yukon. Similarly, one accession of *A. halleri* subsp. *gemmifera* from Shimomatashirodani-deai (halgem8) shared the haplotype with *A. kamchatica* from the central mountains of Japan; and seven accessions of *A. halleri* subsp. *gemmifera* from Japan and Kamchatka (halgem1-7, along with two individuals of *A. halleri* from Europe) shared another, apparently widespread, haplotype with many individuals of *A. kamchatica* from nearly the entire area sampled.

Some geographic structure seems to be present among the *A. kamchatica* haplotypes. Within both the lyrata- and halleri-homeologs we found a widespread and common haplotype, which was observed in many accessions from the broad area sampled, and several haplotypes geographically restricted to one of the areas delimited (see Table 1 and area denoted as A–I). A specific haplotype was also found in the accessions of *A. kamchatica* subsp. *kawasakiana* (lowland in western Honshu). In the halleri-homeolog this haplotype was shared only with a single sample from the lowland of northern Honshu (kamE26H).

CHS homeologs

Two alignments of *CHS* homeologs were assembled (summarized in Table S2). Alignment 1 spanned from the promoter sequence to near the end of exon 2 and comprised 118 sequences obtained from 68 individuals. Alignment 2 included a longer region of the promoter sequence obtained only for the *lyrata*-clade (*A. lyrata* and lyrata-homeolog of tetraploids, see Results), i.e., 57 sequences obtained from 55 individuals. Alignment 2 was analyzed to determine whether better resolution can be seen within that clade. *Arabidopsis thaliana* and *A. pedemontana* were used as outgroups. Because of recombination, we excluded the *CHS* haplotypes of the three European diploid accessions from phylogenetic analyses. The recombinant sequences are shown in Fig. S2a. The all-accession-*CHS* alignment (Alignment 1) displayed 22 unique sequences. It had 1314 aligned positions and involved six indels longer than 1 bp, which were coded as additional binary to four-state characters in the MP analyses. One hundred and ten sites were variable, and 44 of them were parsimony informative.

The lyrata-*CHS* alignment (Alignment 2) comprised 14 unique sequences. It had 1589 aligned positions, and eight indels longer than 1 bp, which were coded as binary to four-state characters in the MP analyses. One hundred and seventeen sites were variable, and 25 of them were parsimony informative.

The MP analysis of the all-accession-CHS data set resulted in a strict consensus tree (four most-parsimonious trees, L = 127 steps, CI = 0.92, RI = 0.95), which displayed three main clades comprising: 1) all accessions of the diploid *A. halleri* and the respective homeolog from the tetraploids; 2) two accessions of *A. lyrata* (lyrpet4, 5) and the respective homeolog from the tetraploids; and 3) three accessions of *A. lyrata* (Fig. 4). Thus, as in the case of the *WER* data, the haplotypes from the diploids are clearly differentiated from each other, and two homeologs are proved to be present in the tetraploids.

We found considerable variation in *A. lyrata.* Two accessions of this species from Far East Russia (Yakut, lyrpet4, 5) shared their haplotype with *A. kamchatica* from Alaska (kamH46). All eight accessions of *A. halleri* subsp. *gemmifera* from Japan and Kamchatka (halgem1-8) had a haplotype identical to that found in *A. kamchatica* from Hokkaido, Far East Russia, Alaska, Yukon, and Washington state. Three accessions of *A. kamchatica* had a haplotype otherwise found in *A. halleri* in Europe (Slovakia, haltat1a1). In addition, *CHS* of lyrpet2 from western Russia was identical to the widespread haplotype of *A. kamchatica* along > 1 kb at the left side of the recombination breakpoint (Fig. S2a), whereas the 3′-side was identical to a few individuals of *A. lyrata*.

In both the lyrata- and halleri-homeologs, one widespread and common haplotype along with several restricted ones were found in *A. kamchatica*. Especially within the halleri-

homeolog, the haplotypes showed a geographic structure. Three subclades or groups can be recognized here: 1) accessions from lowland in western Honshu (= *A. kamchatica* subsp. *kawasakiana*), northern Honshu, and from Taiwan formed one subclade (areas A, B, and E); 2) those from the mountains of western and central Honshu (areas C and D) formed a second subclade; and 3) all the other accessions (i.e., Hokkaido, Far East Russia, Alaska, Yukon, Washington state; areas F–I) were characterized by another haplotype. The consensus tree based on the lyrata-*CHS* data set (Fig. S3a) showed a similar structure among the *A. kamchatica* haplotypes: those found in accessions from lowland in western Honshu, northern Honshu, from Taiwan (areas A, B, and E), and the common haplotype (areas F–I) formed a distinct clade, and were differentiated from the haplotypes from western and central Honshu (areas C and D).

Published *CHS* sequence data (Ramos-Onsins *et al.* 2004) were included in a Neighbor-Joining analysis using MEGA4 (Tamura *et al.* 2007) (Fig. S5). The Genbank accession numbers are AJ619886, AJ619888-619906, AJ619938, and AJ619939. They represent seven sequences of *A. lyrata* subsp. *lyrata* from America, four sequences of *A. lyrata* subsp. *petraea* from Europe, and eleven sequences of *A. halleri* from Europe (representing subsp. *halleri* as defined by Kolnik and Marhold 2006, judged from localities). The sequence AJ619887 was not used because the sequence was short. The evolutionary distances were computed using the Maximum Composite Likelihood method and are in the units of the number of base substitutions per site. All positions containing gaps and missing data were eliminated from the dataset (Complete deletion option). There were a total of 1275 positions in the final dataset.

ITS region

The ITS alignment comprised sequences from 75 individuals, was 619 positions long, and included only a single 1bp-long indel, which was not coded separately. Those of *Arabidopsis thaliana* and *A. pedemontana* were obtained from GenBank (AC006837 and DQ914842) and used as the outgroups. The alignment comprised 25 unique sequences; 59 sites were variable, and 14 of them were parsimony informative.

Intraindividual polymorphic sites were present only scarcely, and IUPAC ambiguity codes were used for coding such polymorphic positions. These sites were found more frequently in the diploids than in the tetraploids, which might suggest that the ITS sequences in the polyploids have been largely homogenized towards one of the parental types possibly due to higher rate of selfing by self-compatibility (see Alvarez & Wendel 2003).

The MP analysis resulted in a 50% majority-rule consensus tree (74,944 most-parsimonious trees, $L = 64$ steps, CI = 0.98, RI = 0.99) with two main and relatively well-supported (100% and 79% bootstrap) clades (Fig. 7). The two clades corresponded to *A. lyrata* and *A. halleri*, respectively. All *A. kamchatica* accessions (with one exception, kamD17) were placed in the clade of *A. lyrata.* A single individual, kamD17, originating from the central mountains in Japan was found in the clade of *A. halleri*. This indicates clearly that concerted evolution of the ITS region was highly effective here and homogenized its sequences towards a repeat type of one parental species, *A. lyrata*. Very little resolution was found within the two main clades, precluding any further inferences (Figs. 7 and S3d).

cpDNA haplotype network and intrapopulation polymorphism

Sequences of two regions of cpDNA (*trnL* intron and *trnL-trnF* intergenic spacer region) were obtained from 50 individuals of *A. kamchatica*, seven individuals of *A. lyrata*, and 15 individuals of *A. halleri* (Table 1). The final alignment of the *trnL* intron was 494 bp long. The *trnL-trnF* sequences varied considerably in length (271–772 bp) because of multiple *trnF* gene duplications and subsequent pseudogene formation (Koch *et al.* 2005). The alignment

was extremely difficult because of the tandem duplications, although it was based on the analysis by Koch *et al.* (2005). Multiple overlapping gaps at the 3′-end of the spacer were identified after alignment. The final alignment comprised 850 aligned positions. Sequences of the two regions from each individual were combined, and a single cpDNA haplotype was produced. Distinct 18-cpDNA haplotypes (cpHap1–18) were observed from *A. kamchatica*, *A. lyrata*, and *A. halleri* (Table 1). The sequences of these cpDNA haplotypes are shown in Fig. S2c.

Seven haplotypes (cpHap1-7) were observed from *A. kamchatica* (including subsp. *kawasakiana*). Four of them (cpHap1, 2, 3, 5) were observed also from *A. halleri*. subsp. *gemmifera,* and their sequences were distinct with different duplicated structure of *trnF* pseudogenes (Fig. S2c). Three other haplotypes were found only in *A. kamchatica*. The cpHap4 was short and found in *A. kamchatica* subsp. *kawasakiana,* and may represent a deletion derivative. The cpHap6 was found in Taiwan. The cpHap7 was also found in Taiwan, and is close to the cpHap1 with a SNP.

The minimum spanning network was drawn based on the *trnL* intron region of *Arabidopsis* species. In total 27 *trnL* intron region haplotypes obtained in this study and in Koch and Matschinger 2007 and Schmickl *et al.* 2008 (GenBank accession numbers DQ313494-313502, DQ313504-313508, DQ313510-313520, DQ914841, DQ529016) were included in this analysis (Fig. 5). *Arabidopsis thaliana* and *A. suecica* were not included because their cpDNA haplotypes are highly divergent. Total 582 individuals from Koch & Matschinger 2007 (SI Table 1) and Schmickl *et al.* 2008 (Supplementary material Table 1) are categorized according to the rules described below:

Arabis umbrosa from East Russia is categorized into *A. lyrata*, Asia. Because Schmicle *et al.* (2008) noted "the difficulties in assigning herbarium vouchers from Canada to ssp. *lyrata* or ssp. *kamchatica*", and because the taxonomic treatment of *A. kamchatica* has been a matter of debates, the individuals designated as *A. kamchatica* or *lyrata* from America or Asia in these references are categorized into one category, *A. lyrata* or *kamchatica*, America and Asia. *Arabidopsis arenosa* and *A. neglecta* are categorized into *A. arenosa*. The individuals designated as hybrid or without the information of haplotype were excluded. Several individuals most possibly overlapping in these two references (marked with the same herbarium number) were counted only once.

The network was constructed using median-joining method ($\varepsilon = 0$), implemented in the NETWORK program v. 4.5.1.0 (Bandelt *et al.*1999; freely available at www.fluxus-engineering.com). Every insertion and deletion was scored as a single mutational event regardless of its length. The *trnL-trnF* intergenic spacer region was not included in the construction of the network, because frequent and possibly parallel mutations in tandemly-duplicated copies made the alignment uncertain (Koch *et al.* 2005). Instead, we used the *trnL-trnF* intergenic spacer region to assess if cpDNA haplotype between those of *A. kamchatica* observed in this study and those of other *Arabidopsis* species are identical. In some of the individuals reported by Koch and Matschinger 2007 and Schmickl *et al.* 2008, only *trnL* intron sequence was reported and *trnL-trnF* regions was not available. Those individuals were included in *trnL* intron network (Fig. 5b), but were not used to count the haplotype including the *trnL-trnF* intergenic spacer region (Fig. 5a).

To survey the extent of polymorphism within populations, sequences of the *trnL-trnF* region were obtained from additional individuals from three selected populations: two populations of subsp. *kamchatica* (eight individuals from the population kamC14, and six individuals from the population kamC11, respectively) and one population of subsp. *kawasakiana* (five individuals from the population kamkwsB6) (Table S3). Despite the high levels of variation among populations in the *trnL-trnF* region, no polymorphism in *trnL-trnF*

was found in any local populations. This suggests that most of the polymorphisms are distributed among populations.

Bayesian clustering

To detect the population structure of *A. kamchatica* and assign individuals to populations, we used the Bayesian clustering algorithm implemented in program *structure* version 2.2 (http://pritch.bsd.uchicago.edu/structure.html) (Pritchard *et al.* 2000) (Fig. 8). The algorithm uses individual multilocus genotypic data and attempts to assign individuals to clusters under the predefined model with a certain number of clusters (*K*). In this study, two homeologous pairs of nuclear *WER* and *CHS* genes (*WER*-lyr, *WER*-hal, *CHS*-lyr and *CHS*-hal), and the cpDNA haplotype were included in the *structure* analysis. A high rate of selfing in *A. kamchatica* (Sugisaka & Kudoh 2008) violates the assumption of the Hardy–Weinberg equilibrium within a population in the *structure* analysis. Therefore, as recommended for complete-selfing species (Gao *et al.* 2007) and as commonly done in predominantly selfing species (e.g., Nordborg *et al.* 2005; Beck *et al.* 2008), the data were treated as haploid data. To prepare the haploid data set, the cpDNA haplotype and one haplotype from each nuclear locus were selected randomly and scored for each individual. Both substitutions and indels were used as information to distinguish haplotypes. As described in Results, the locus *WER*-lyr in three individuals from Taiwan appeared to have rearrangement or indels in the 5′-half of the gene and was treated as a distinct haplotype. The locus *CHS*-hal in three individuals was not amplified and was treated as missing data for these three individuals. The genetic distances between haplotypes were not considered in this program.

The model used in the *structure* analysis assumes no association between alleles from different loci arising from physical linkage by chromosomal proximity. To survey the association between different loci, the degree of the gametic disequilibrium D' (average of the absolute value of D_{ij}/D_{max} over all pairs of alleles from different loci weighted by the frequencies of the gametes) (Hedrick 1987) was calculated for four nuclear loci (six pairs of loci). Four of the six pairs (*WER*-lyr–*WER*-hal; *CHS*-lyr–*CHS*-hal; *WER*-lyr–*CHS*-hal and *CHS*-lyr–*WER*-hal) must reside on different chromosomes because they are the pairs from different parents. The D' value between these four pairs of loci was 0.788, 0.937, 0.806, and 0.818, respectively. These high gametic disequilibria observed between the loci from different parents were probably not the result of physical linkage but arose because of other factors such as the population structure. The other two pairs could potentially be in physical linkage (*WER*-hal–*CHS*-hal and *WER*-lyr–*CHS*-lyr), which would result in a higher D' value because *WER* and *CHS* reside on the same chromosome in the related species *A. thaliana*. However, the D' values (0.892 and 0.719 for *WER*-hal–*CHS*-hal and *WER*-lyr–*CHS*-lyr, respectively) are similar or lower than the pairs from different parents compared with the range of D' value of the pairs from different chromosomes described above (0.788-0.937). This indicates that the two pairs derived from the same parents do not have an elevated level of gametic (linkage) disequilibrium because of physical linkage and were thus treated as independent loci in the *structure* analysis.

Twenty independent runs with 100,000 iterations for the burn-in phase and 100,000 iterations for the data collection phase were conducted for different numbers of clusters ranging from $K = 1$ to 10. For all runs, admixture and correlated allele frequency models were used. Using the program *CLUMPP* (Jakobsson & Rosenberg 2007), the optimal alignments of 20 replicate clustering estimates were found for each number of clusters *K*. The *Greedy* algorithm (for $K = 2$ to 7) and the *LargeKGreedy* algorithm (for $K = 8$ to 10) of the program with 1,000 random input orders of 20 replicates were used. The averages of cluster membership coefficients were taken for all runs of each *K* with the optimal alignment, and the outputs were graphically displayed by the program *distruct*

(http://rosenberglab.bioinformatics.med.umich.edu/distruct.html) (Rosenberg 2004). To investigate the similarity of clustering estimates between different runs, the symmetric similarity coefficient (*SSC*) (Jakobsson & Rosenberg 2007) was computed for all pairs of runs with a given *K* using the program *CLUMPP*. The optimal number of clusters (*K*) was inferred based on evaluation of the ΔK statistic (Evanno *et al.* 2005).

Supporting Table S1. Primers used for PCR

Gene	homeolog type	Primer name	Primer sequence (5' to 3')	annealing (˚C)
WER	lyrata	WERFU3	TATACATAAATATTCCACTAGGTTCTG	57
		WERRD6	AATTGAAGAAACATTTAAAACATT	
	lyrata (Taiwan)	WERlyrF1	CTATTTCAAGAGAAGAAAAAACAGC	53
		WERRD6	AATTGAAGAAACATTTAAAACATT	
	lyrata & halleri	WERFU1	TCTCTCGTTTTATGATCTCTCTCG	57
		WERRD1	AGCCAATCATACACTACCACATCA	
	halleri	WERFU1	TCTCTCGTTTTATGATCTCTCTCG	53
		WERgemRD1	GTTTGATCAGCTTTGCATGCA	
	halleri	WERFU1	TCTCTCGTTTTATGATCTCTCTCG	53
		WERgemR3e2	TGTTTGGTTTTCTCATGATCT	
	lyrata & halleri	WERF4	TGTAGATTGAGGTGGATGAA	53
		WERR3e1	TGAACCCAAAGTGAACTCAAGTAG	
CHS	lyrata	CHSlyrFU1	TGGGAAGTGAAATCTCCTTATGGTG	57
		CHSR1	AGAGGAACGCTGTGCAAGAC	
	lyrata	CHSlyrFU4	GGTGGAGAAACTATACAACAAAT	57
		CHSR1	AGAGGAACGCTGTGCAAGAC	
	halleri	CHSgemFU1	GAAATCTCCGTAGTCCGTATGGTG	57
		CHSR1	AGAGGAACGCTGTGCAAGAC	
	halleri	CHSgemFU15	CTAACAACTAGCCACGTATATCTTC	* 67.5
		CHSR1	AGAGGAACGCTGTGCAAGAC	
	lyrata & halleri	CHSkamF1	CTAACCCTGAGAACCATGTG	53
		CHSR3	TATGGCACCATCAGAGTCTG	
ITS		ITSP1A	GGAAGGAGAAGTCGTAACAAGG	45
		ITSP4	TCCTCCGCTTATTGATATGC	
trnL_trnF		trnL/FIGSF	GGTTCAAGTCCCTCTATCCC	45
		trnL/FIGSR	GATTTTCAGTCCTCTGCTCTAC	
trnL intron		trnLintrF	CGAAATCGGTAGACGCTACG	38
		trnLintrR	GGGGATAGAGGGACTTGAAC	

*Temp. used for *kamchatica*, 55 for *halleri*

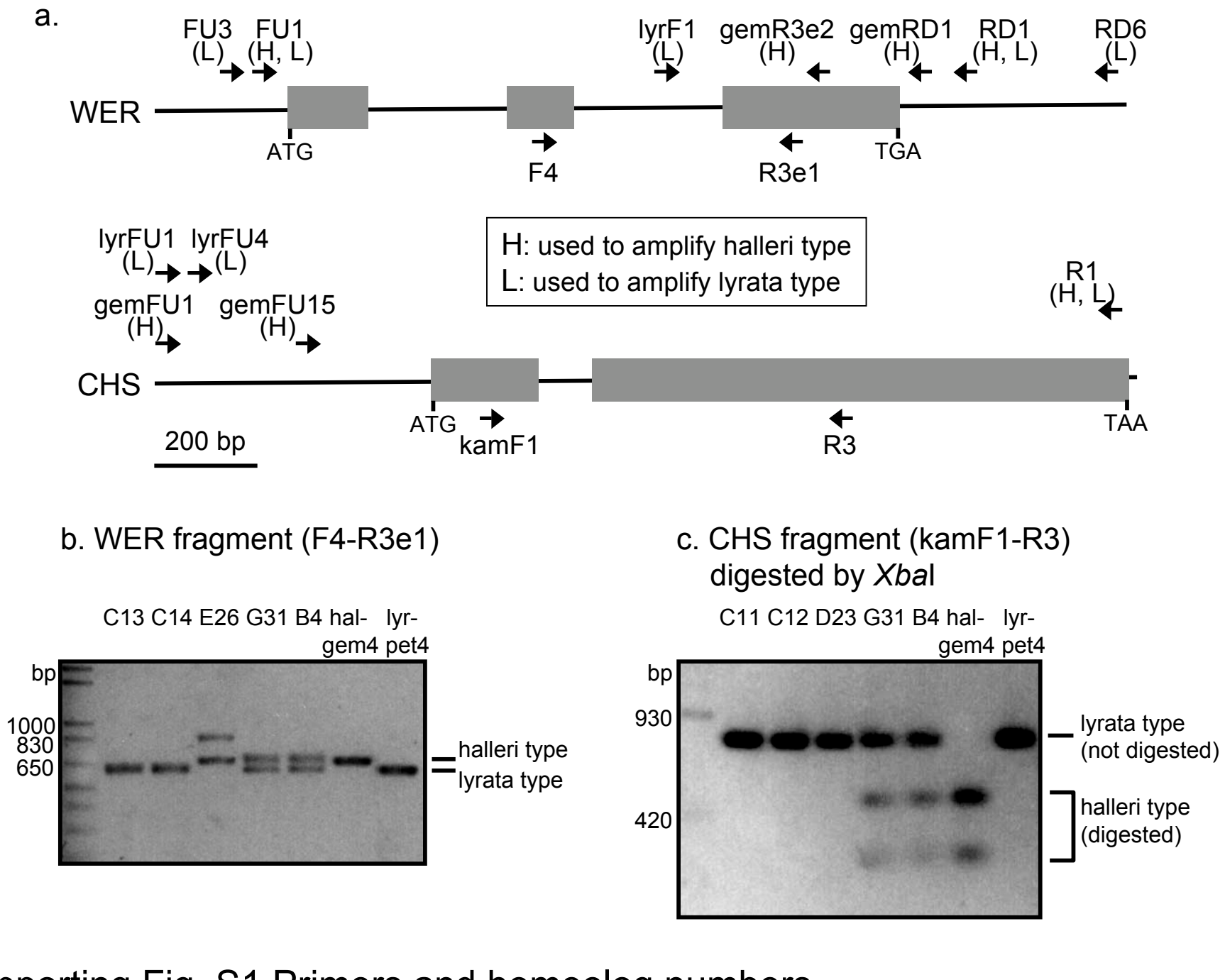
Supporting Table S2. Comparative information for nuclear DNA.

	nDNA			
	WER	**CHS-all-accessions**	**CHS-lyr-clade**	**ITS**
no. individuals	72	68	55	75
no. outgroup	1	2	2	2
no. sequences	120	118	57	75
sequence length (bp)	835-925	1286-1305	1551-1574	618-619
aligned length (bp)	955	1314	1589	619
no. coded indels	14	6	8	0
no. haplotypes	24	22	14	25
variable characters (% by aligned length, incl. indel	83 (8.6%)	110 (8.3%)	117 (7.4%)	59 (9.5%)
parsimony-informative characters (%,incl. indel coding)	60 (6.2%)	44 (3.3%)	25 (1.6%)	14 (2.3%)
no. MP trees	78	4	1	74,944
tree length (steps)	96	127	127	64
CI (consistency index)	0.96	0.92	0.96	0.98
RI (retention index)	0.99	0.95	0.85	0.99
no. haplotypes of halleri-homeolog of *A. kamchatica*	8	8		
nucleotide diversity π of halleri-homeolog of *A. kamchatica*	0.0008	0.0026		
no. haplotypes of lyrata-homeolog of *A. kamchatica*	8	11		
nucleotide diversity π of lyrata-homeolog of *A. kamchatica*	0.0006	0.0012		

One haplotype was used from each individual in the calculation of nucleotide diversity π (Tajima 1983)
WER-lyrata of kamA1-3 and kamE26 accessions were removed due to large indels

Table S3 Samples for intra-populaton variation analyses (only *trnL-trnF* of cpDNA was surveyed)

Name of taxon	Area		sample name	Population	Collector	*trnL-trnF* of cpDNA
Arabidopsis kamchatica subsp. *kamchatica*	C	Mountains in Western Honshu and Shikoku, Japan	kamC101	Japan, Tokushima, Mt. Shikokutsurugi	M. Kanaoka, KKS	cpHap1
Arabidopsis kamchatica subsp. *kamchatica*	C	Mountains in Western Honshu and Shikoku, Japan	kamC102	Japan, Tokushima, Mt. Shikokutsurugi	M. Kanaoka, KKS	cpHap1
Arabidopsis kamchatica subsp. *kamchatica*	C	Mountains in Western Honshu and Shikoku, Japan	kamC103	Japan, Tokushima, Mt. Shikokutsurugi	M. Kanaoka, KKS	cpHap1
Arabidopsis kamchatica subsp. *kamchatica*	C	Mountains in Western Honshu and Shikoku, Japan	kamC104	Japan, Tokushima, Mt. Shikokutsurugi	M. Kanaoka, KKS	cpHap1
Arabidopsis kamchatica subsp. *kamchatica*	C	Mountains in Western Honshu and Shikoku, Japan	kamC105	Japan, Tokushima, Mt. Shikokutsurugi	M. Kanaoka, KKS	cpHap1
Arabidopsis kamchatica subsp. *kamchatica*	C	Mountains in Western Honshu and Shikoku, Japan	kamC106	Japan, Tokushima, Mt. Shikokutsurugi	M. Kanaoka, KKS	cpHap1
Arabidopsis kamchatica subsp. *kamchatica*	C	Mountains in Western Honshu and Shikoku, Japan	kamC107	Japan, Tottori, Mt. Daisen	KKS	cpHap3
Arabidopsis kamchatica subsp. *kamchatica*	C	Mountains in Western Honshu and Shikoku, Japan	kamC108	Japan, Tottori, Mt. Daisen	KKS	cpHap3
Arabidopsis kamchatica subsp. *kamchatica*	C	Mountains in Western Honshu and Shikoku, Japan	kamC109	Japan, Tottori, Mt. Daisen	KKS	cpHap3
Arabidopsis kamchatica subsp. *kamchatica*	C	Mountains in Western Honshu and Shikoku, Japan	kamC110	Japan, Tottori, Mt. Daisen	KKS	cpHap3
Arabidopsis kamchatica subsp. *kamchatica*	C	Mountains in Western Honshu and Shikoku, Japan	kamC111	Japan, Tottori, Mt. Daisen	KKS	cpHap3
Arabidopsis kamchatica subsp. *kamchatica*	C	Mountains in Western Honshu and Shikoku, Japan	kamC112	Japan, Tottori, Mt. Daisen	KKS	cpHap3
Arabidopsis kamchatica subsp. *kamchatica*	C	Mountains in Western Honshu and Shikoku, Japan	kamC113	Japan, Tottori, Mt. Daisen	KKS	cpHap3
Arabidopsis kamchatica subsp. *kamchatica*	C	Mountains in Western Honshu and Shikoku, Japan	kamC114	Japan, Tottori, Mt. Daisen	KKS	cpHap3
Arabidopsis kamchatica subsp. *kawasakiana*	B	Lowland in Western Honshu, Japan	kamkwsB115	Japan, Shiga, Takashima	KKS	cpHap4
Arabidopsis kamchatica subsp. *kawasakiana*	B	Lowland in Western Honshu, Japan	kamkwsB116	Japan, Shiga, Takashima	KKS	cpHap4
Arabidopsis kamchatica subsp. *kawasakiana*	B	Lowland in Western Honshu, Japan	kamkwsB117	Japan, Shiga, Takashima	KKS	cpHap4
Arabidopsis kamchatica subsp. *kawasakiana*	B	Lowland in Western Honshu, Japan	kamkwsB118	Japan, Shiga, Takashima	KKS	cpHap4
Arabidopsis kamchatica subsp. *kawasakiana*	B	Lowland in Western Honshu, Japan	kamkwsB119	Japan, Shiga, Takashima	KKS	cpHap4



Supporting Fig. S1 Primers and homeolog numbers

Haplotype name	Frequency	Segregating sites
kamA1L-A3L,kamkwsB10L	4	C T G C C A C T A T G C C C G - A C C G C T T C T T C T T C C C C A G A C C G C T C C T C A G C G T G
kamkwsB4L-B9L	6	. C . . . . . - . . . . .
kamCL,DL *	12	. . A . . . . . T . . . . .
kamC12La2	1	. . A . . . . . C . . . . .
kamD16L	1	. . A . . . . . T . . . . . G . . . . .
kamD17La1	1	. . A T . . . . . T . . . . .
kamD17La2	1	. . A T . . . . . T . . . . . - . . . . .
kamD18L	1	T . A . . . . . T . T . . . . .
kamEL,FL,GL,HL,IL **	23	. . . . . T . . . . .
kamG31L	1	. . . . . T . . . . . G . . . . .
kamH46L,lyrpet4,lyrpet5	3	. . A . . . . C . . . . . T . . . . .
lyrpet2	1	. . A . . . . . T . . . . . T . T A . T . T T A T . C A
lyrpet3	1	. . A . . C . . . . . A T . T A T A C C T C . . C G T . . T T . . . T . T A . T . T T A T . C A
lyrlyr1a1	1	N N N N T . . . . . A T . T A T A C C T C . . C G T . . T T . . . T . T A . T . T T A T . C A
lyrlyr1a2,lyrlyr2	2	N N N N T . . . . . A T . T A T A C C T C . . C G T . A T T . . . T . T A . T . T T A T . C A
kamA1H-A3H,E26H,kamkwsB6H-B10H	9	N N N N . . T . . G . A . . A T T T A T A C C T C . . . . . T C C G . T . . . . A . . . . T . .
kamkwsB4H,B5H	2	N N N N . . T . . G . A . . A T T T A T A C C T C . A . . . . . T C C G . T . . . . A . . . . T . .
kamC13H,C14H,D16H	3	N N N N . . . . . G . A G . A T . T A T A C C T C . . . . . T C C G . T . . T . A . . . . T . .
kamD15H,D17H,D24Ha2,haltat1a1	4	N N N N . . . . . G . A . . A T . T A T A C C T C . . . . . T C C G . T . . T . A . . . . T . .
kamD18H	1	N N N N . . . . . G . A . . A T . T A T A C C T C . . . . . T C C G . T . . T . A . . . . T . .
kamD19H,D20H	2	N N N N . . . . G G C A . . A T . T A T A C C T C . . . . . T C C G . T . . T . A . . . . T . .
kamD21H,D22H,D24Ha1,D25H	4	N N N N . . . . . G C A . . A T . T A T A C C T C . . . . . T C C G . T . . T . A . . . . T . .
kamFH,GH,HH,IH*,halgem1-8	32	N N N N . . . . . G . A . . A T . T A T A C C T C . . . . . T C C G . T . . . . A . . . . T . .
haltat1a2	1	N N N N . . . . . G . A . . A T T T A T A C C T C . . . . . T C C G . T . . T . A . . . . T . .
haltat2	1	N N N N . . . . . G . A . . A T . T A T A C C T C . . . . . T C C G . A . . T . A . . . . T . .
halovi1	1	N N N N . . . . . G . A . . A T T T A T A C C T C . . . . . T C C G . A . . T . A . . . . T . .

* kamC11L, C12La1, C13L, C14L, D15L, D19L-D25L

** kamE26L, F27L-F30L, G32L-G41L, H42L-H45L, H47L, I48L-I50L

Supporting Fig. S2a Segregating sites of *CHS*

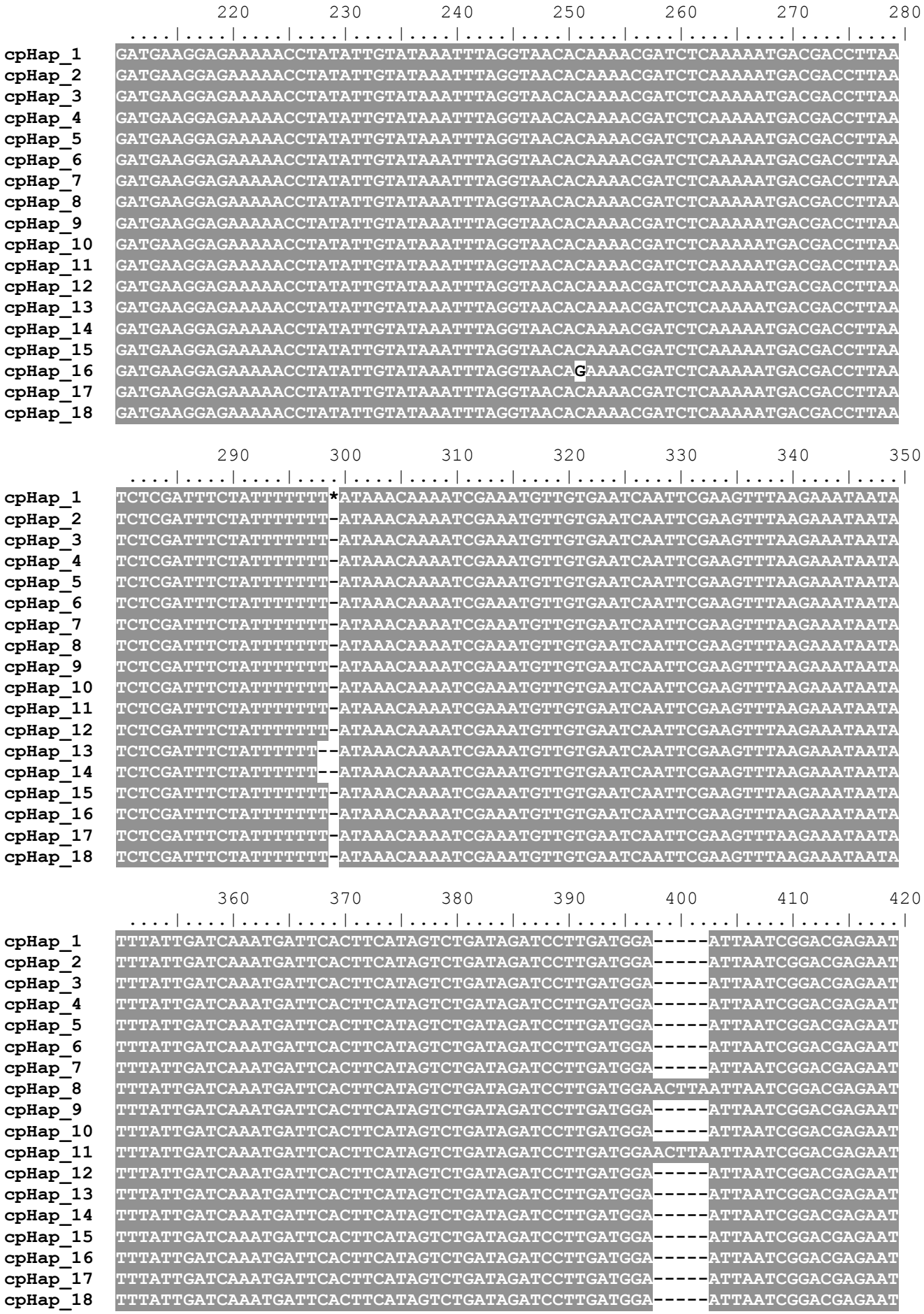
Haplotype name	Frequency	Segregating sites
kamA,C,D,F,G,H,I*,kamkwsB4–B10	41	C C G C C T T C G C G C A G G A G G T G T A A C C C T T G G T G C T A C C G
kamA3a2	1	. . . . . – . . . . .
kamD15	1	G . . . . . K . . . . . R
kamD17	1	. . . T T C . . A T A . T T A G . A . A . T . . T T C C . . . . . C G . . .
kamD18	1	. . . . . G . . .
kamD24	1	. . . . . K . . . . .
kamE26	1	. . . . . W . . . . . Y . . . . .
kamG32	1	. . . . . R . . . . . R . . . . .
kamH43,H47,I48	3	. . . . . Y . R S . .
lyrpet1	1	. . . . . R . . . . . W K . . . . .
lyrpet2,3	2	. . . . . A T . . . . .
lyrlyr1,2,lyrpet4,5	4	. . . . . T . G G . .
haldac1	1	. . . . . C C . . . A . T . A G . A . . . . . C C . . . . . C G . . .
halgem1–4, 8	5	. . . . . C . . . T A . T . A G . A . A . T . . . . C C A . . . . C G . . .
halgem5	1	. . . . . C . . . T A . T . A G . A . A . T . . . . Y C A . . . . C G . . .
halgem6	1	. . . Y . C . . R T A . T . A G . A . A . T . . Y . Y C R K . . . C G . Y .
halgem7	1	. . . . . C . . . T A . T . A G . A . A . T . . Y . Y C R K . . . C G . Y .
halhal1	1	. Y . . . C . . . T A . T . A G . A . A . . . . . C C R . . . . C G . . .
halhal2	1	. Y R . . C . Y R T A Y T . A G . A . A . W . . . . C C R . . . . C G . . .
haltat1	1	. . . . . C . . . T A . T . A G . A . A . . . . . C C R . . . . C G . . .
haltat2	1	. . . . . C . . . T A . T . A G . A . A . . . . . C C A . W . . C G . . .
haltat3	1	. . . . . C . . . T A . T . A G . A . A . . . . . Y C A . . S . C G . . .
halovi1	1	. T . . . C . . . T A . T . A G . A . A . . . . . C C A . . . . C G . . .

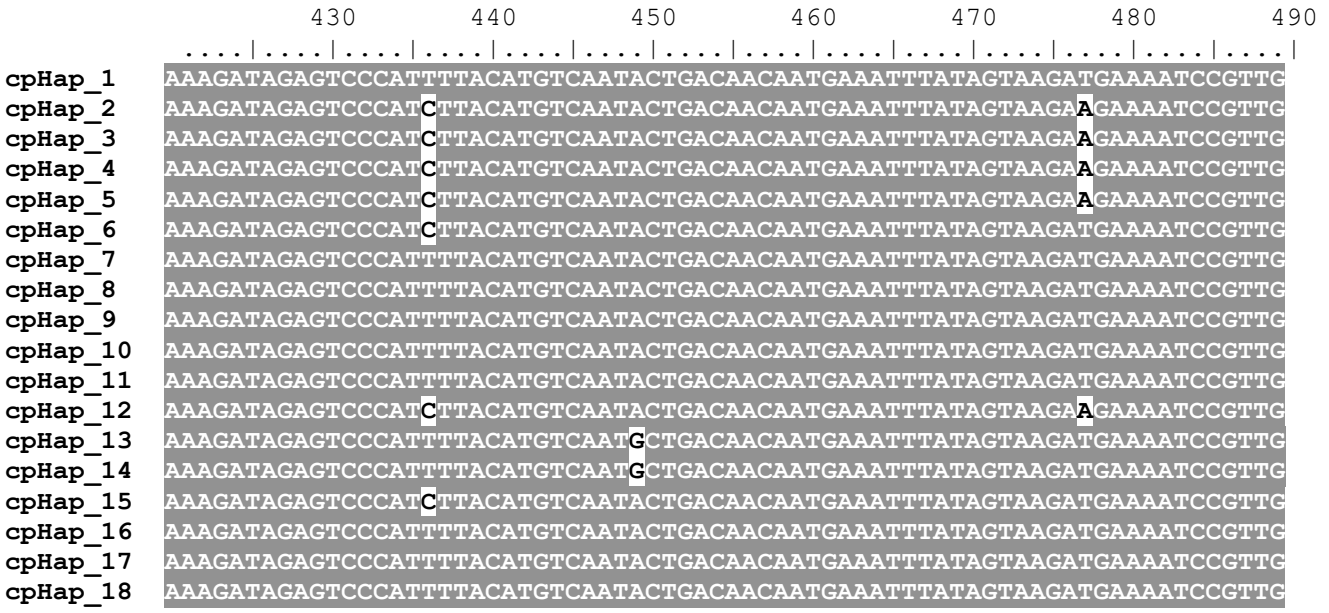
* kamA1, A2, A3a1, C11–C14, D16, D19–D23, D25, F27–F30, G31, G33–G41, H42, H44–H46, I49, I50, kamkwsB4–B10

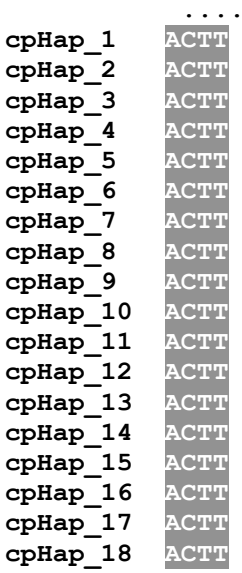
Supporting Fig. S2b Segregating sites of ITS

***trnL* intron**

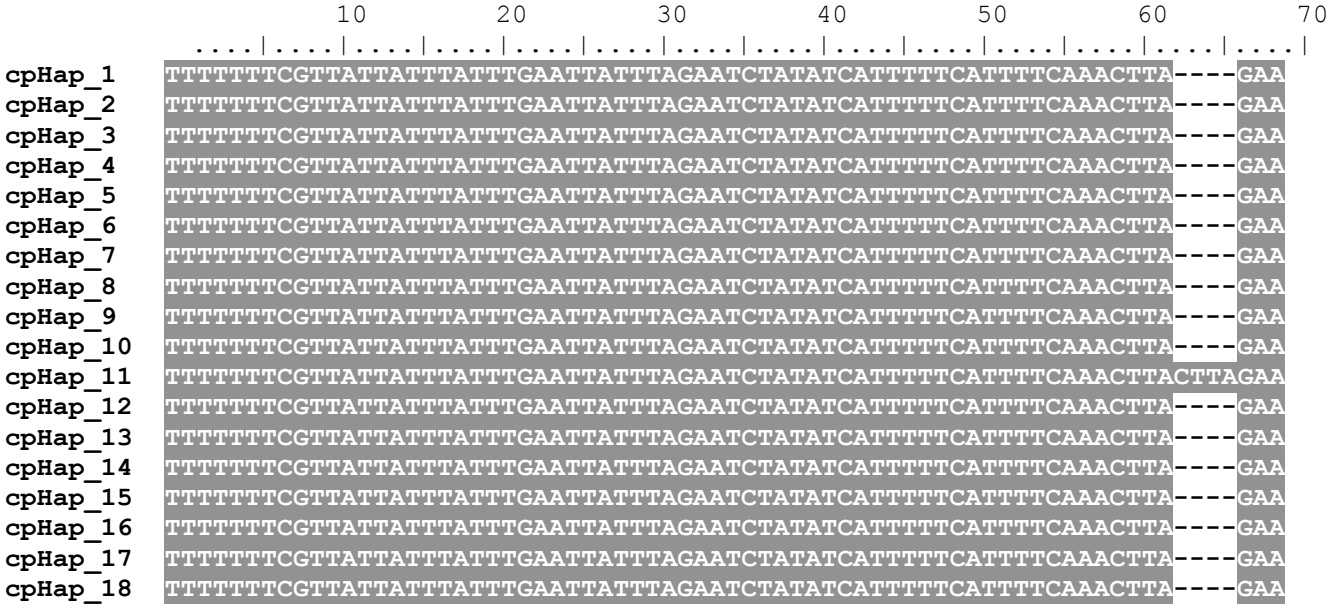
```
                10        20        30        40        50        60        70
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
cpHap_1   TACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAATTAACAATGGGCAATCCTGAGCCAAATCCTG
cpHap_2   TACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAATTAACAATGGGCAATCCTGAGCCAAATCCTG
cpHap_3   TACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAATTAACAATGGGCAATCCTGAGCCAAATCCTG
cpHap_4   TACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAATTAACAATGGGCAATCCTGAGCCAAATCCTG
cpHap_5   TACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAATTAACAATGGGCAATCCTGAGCCAAATCCTG
cpHap_6   TACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAATTAACAATGGGCAATCCTGAGCCAAATCCTG
cpHap_7   TACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAATTAACAATGGGCAATCCTGAGCCAAATCCTG
cpHap_8   TACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAATTAACAATGGGCAATCCTGAGCCAAATCCTG
cpHap_9   TACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAATTAACAATGGGCAATCCTGAGCCAAATCCTG
cpHap_10  TACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAATTAACAATGGGCAATCCTGAGCCAAATCCTG
cpHap_11  TACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAATTAACAATGGGCAATCCTGAGCCAAATCCTG
cpHap_12  TACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAATTAACAATGGGCAATCCTGAGCCAAATCCTG
cpHap_13  TACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAATTAACAATGGGCAATCCTGAGCCAAATCCTG
cpHap_14  TACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAATTAACAATGGGCAATCCTGAGCCAAATCCTG
cpHap_15  TACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAATTAACAATGGGCAATCCTGAGCCAAATCCTG
cpHap_16  TACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAATTAACAATGGGCAATCCTGAGCCAAATCCTG
cpHap_17  TACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAATTAACAATGGGCAATCCTGAGCCAAATCCTG
cpHap_18  TACTAAGTGATAACTTTCAAATTCAGAGAAACCCTGGAATTAACAATGGGCAATCCTGAGCCAAATCCTG

                80        90       100       110       120       130       140
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
cpHap_1   GTTTACGCGAACAAACCGGAGTTTACAAAGCGCGAAAAAAGGGATAGGTGCAGAGACTCAATGGAAGCTG
cpHap_2   GTTTACGCGAACAAACCGGAGTTTACAAAGCGCGAAAAAAGGGATAGGTGCAGAGACTCAATGGAAGCTG
cpHap_3   GTTTACGCGAACAAACCGGAGTTTACAAAGCGCGAAAAAAGGGATAGGTGCAGAGACTCAATGGAAGCTG
cpHap_4   GTTTACGCGAACAAACCGGAGTTTACAAAGCGCGAAAAAAGGGATAGGTGCAGAGACTCAATGGAAGCTG
cpHap_5   GTTTACGCGAACAAACCGGAGTTTACAAAGCGCGAAAAAAGGGATAGGTGCAGAGACTCAATGGAAGCTG
cpHap_6   GTTTACGCGAACAAACCGGAGTTTACAAAGCGCGAAAAAAGGGATAGGTGCAGAGACTCAATGGAAGCTG
cpHap_7   GTTTACGCGAACAAACCGGAGTTTACAAAGCGCGAAAAAAGGGATAGGTGCAGAGACTCAATGGAAGCTG
cpHap_8   GTTTACGCGAACAAACCGGAGTTTACAAAGCGCGAAAAAAGGGATAGGTGCAGAGACTCAATGGAAGCTG
cpHap_9   GTTTACGCGAACAAACCGGAGTTTACAAAGCGCGAAAAAAGGGATAGGTGCAGAGACTCAATGGAAGCTG
cpHap_10  GTTTACGCGAACAAACCGGAGTTTACAAAGCGCGAAAAAAGGGATAGGTGCAGAGACTCAATGGAAGCTG
cpHap_11  GTTTACGCGAACAAACCGGAGTTTACAAAGCGCGAAAAAAGGGATAGGTGCAGAGACTCAATGGAAGCTG
cpHap_12  GTTTACGCGAACAAACCGGAGTTTACAAAGCGCGAAAAAAGGGATAGGTGCAGAGACTCAATGGAAGCTG
cpHap_13  GTTTACGCGAACAAACCGGAGTTTACAAAGCGCGAAAAAAGGGATAGGTGCAGAGACTCAATGGAAGCTG
cpHap_14  GTTTACGCGAACAAACCGGAGTTTACAAAGCGCGAAAAAAGGGATAGGTGCAGAGACTCAATGGAAGCTG
cpHap_15  GTTTACGCGAACAAACCGGAGTTTACAAAGCGCGAAAAAAGGGATAGGTGCAGAGACTCAATGGAAGCTG
cpHap_16  GTTTACGCGAACAAACCGGAGTTTACAAAGCGCGAAAAAAGGGATAGGTGCAGAGACTCAATGGAAGCTG
cpHap_17  GTTTACGCGAACAAACCGGAGTTTACAAAGCGCGAAAAAAGGGATAGGTGCAGAGACTCAATGGAAGCTG
cpHap_18  GTTTACGCGAACAAACCGGAGTTTACAAAGCGCGAAAAAAGGGATAGGTGCAGAGACTCAATGGAAGCTG

               150       160       170       180       190       200       210
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
cpHap_1   TTCTAACAAATGGAGTTCACTACCTTGTGTTGATAAAGGAATCCTTCGATCGAAACTTCAAATCAAAAAG
cpHap_2   TTCTAACAAATGGAGTTCACTACCTTGTGTTGATAAAGGAATCCTTCGATCGAAACTTCAAATCAAAAAG
cpHap_3   TTCTAACAAATGGAGTTCACTACCTTGTGTTGATAAAGGAATCCTTCGATCGAAACTTCAAATCAAAAAG
cpHap_4   TTCTAACAAATGGAGTTCACTACCTTGTGTTGATAAAGGAATCCTTCGATCGAAACTTCAAATCAAAAAG
cpHap_5   TTCTAACAAATGGAGTTCACTACCTTGTGTTGATAAAGGAATCCTTCGATCGAAACTTCAAATCAAAAAG
cpHap_6   TTCTAACAAATGGAGTTCACTACCTTGTGTTGATAAAGGAATCCTTCGATCGAAACTTCAAATCAAAAAG
cpHap_7   TTCTAACAAATGGAGTTCACTACCTTGTGTTGATAAAGGAATCCTTCGATCGAAACTTCAAATCAAAAAG
cpHap_8   TTCTAACAAATGGAGTTCACTACCTTGTGTTGATAAAGGAATCCTTCGATCGAAACTTCAAATCAAAAAG
cpHap_9   TTCTAACAAATGGAGTTCACTACCTTGTGTTGATAAAGGAATCCTTCGATCGAAACTTCAAATCAAAAAG
cpHap_10  TTCTAACAAATGGAGTTCACTACCTTGTGTTGATAAAGGAATCCTTCGATCGAAACTTCAAATCAAAAAG
cpHap_11  TTCTAACAAATGGAGTTCACTACCTTGTGTTGATAAAGGAATCCTTCGATCGAAACTTCAAATCAAAAAG
cpHap_12  TTCTAACAAATGGAGTTCACTACCTTGTGTTGATAAAGGAATCCTTCGATCGAAACTTCAAATCAAAAAG
cpHap_13  TTCTAACAAATGGAGTTCACTACCTTGTGTTGATAAAGGAATCCTTCGATCGAAACTTCAAATCAAAAAG
cpHap_14  TTCTAACAAATGGAGTTCACTACCTTGTGTTGATAAAGGAATCCTTCGATCGAAACTTCAAATCAAAAAG
cpHap_15  TTCTAACAAATGGAGTTCACTACCTTGTGTTGATAAAGGAATCCTTCGATCGAAACTTCAAATCAAAAAG
cpHap_16  TTCTAACAAATGGAGTTCACTACCTTGTGTTGATAAAGGAATCCTTCGATCGAAACTTCAAATCAAAAAG
cpHap_17  TTCTAACAAATGGAGTTCACTACCTTGTGTTGATAAAGGAATCCTTCGATCGAAACTTCAAATCAAAAAG
cpHap_18  TTCTAACAAATGGAGTTCACTACCTTGTGTTGATAAAGGAATCCTTCGATCGAAACTTCAAATCAAAAAG
```

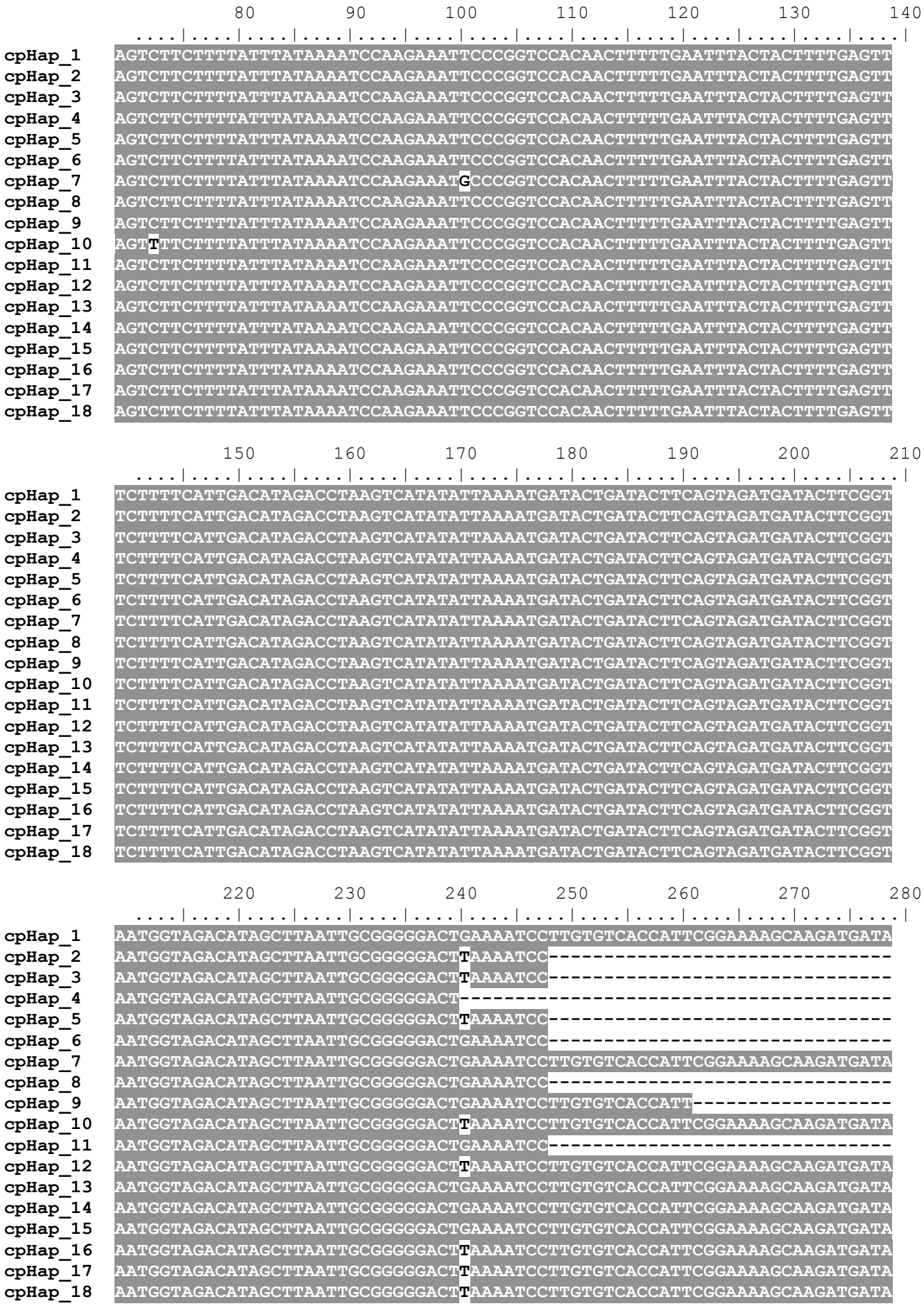

220 230 240 250 260 270 280
cpHap_1 GATGAAGGAGAAAAACCTATATTGTATAAATTTAGGTAACACAAAACGATCTCAAAAATGACGACCTTAA
cpHap_2 GATGAAGGAGAAAAACCTATATTGTATAAATTTAGGTAACACAAAACGATCTCAAAAATGACGACCTTAA
cpHap_3 GATGAAGGAGAAAAACCTATATTGTATAAATTTAGGTAACACAAAACGATCTCAAAAATGACGACCTTAA
cpHap_4 GATGAAGGAGAAAAACCTATATTGTATAAATTTAGGTAACACAAAACGATCTCAAAAATGACGACCTTAA
cpHap_5 GATGAAGGAGAAAAACCTATATTGTATAAATTTAGGTAACACAAAACGATCTCAAAAATGACGACCTTAA
cpHap_6 GATGAAGGAGAAAAACCTATATTGTATAAATTTAGGTAACACAAAACGATCTCAAAAATGACGACCTTAA
cpHap_7 GATGAAGGAGAAAAACCTATATTGTATAAATTTAGGTAACACAAAACGATCTCAAAAATGACGACCTTAA
cpHap_8 GATGAAGGAGAAAAACCTATATTGTATAAATTTAGGTAACACAAAACGATCTCAAAAATGACGACCTTAA
cpHap_9 GATGAAGGAGAAAAACCTATATTGTATAAATTTAGGTAACACAAAACGATCTCAAAAATGACGACCTTAA
cpHap_10 GATGAAGGAGAAAAACCTATATTGTATAAATTTAGGTAACACAAAACGATCTCAAAAATGACGACCTTAA
cpHap_11 GATGAAGGAGAAAAACCTATATTGTATAAATTTAGGTAACACAAAACGATCTCAAAAATGACGACCTTAA
cpHap_12 GATGAAGGAGAAAAACCTATATTGTATAAATTTAGGTAACACAAAACGATCTCAAAAATGACGACCTTAA
cpHap_13 GATGAAGGAGAAAAACCTATATTGTATAAATTTAGGTAACACAAAACGATCTCAAAAATGACGACCTTAA
cpHap_14 GATGAAGGAGAAAAACCTATATTGTATAAATTTAGGTAACACAAAACGATCTCAAAAATGACGACCTTAA
cpHap_15 GATGAAGGAGAAAAACCTATATTGTATAAATTTAGGTAACACAAAACGATCTCAAAAATGACGACCTTAA
cpHap_16 GATGAAGGAGAAAAACCTATATTGTATAAATTTAGGTAACAGAAAACGATCTCAAAAATGACGACCTTAA
cpHap_17 GATGAAGGAGAAAAACCTATATTGTATAAATTTAGGTAACACAAAACGATCTCAAAAATGACGACCTTAA
cpHap_18 GATGAAGGAGAAAAACCTATATTGTATAAATTTAGGTAACACAAAACGATCTCAAAAATGACGACCTTAA
290 300 310 320 330 340 350
cpHap_1 TCTCGATTTCTATTTTTTT*ATAAACAAAATCGAAATGTTGTGAATCAATTCGAAGTTTAAGAAATAATA
cpHap_2 TCTCGATTTCTATTTTTTT-ATAAACAAAATCGAAATGTTGTGAATCAATTCGAAGTTTAAGAAATAATA
cpHap_3 TCTCGATTTCTATTTTTTT-ATAAACAAAATCGAAATGTTGTGAATCAATTCGAAGTTTAAGAAATAATA
cpHap_4 TCTCGATTTCTATTTTTTT-ATAAACAAAATCGAAATGTTGTGAATCAATTCGAAGTTTAAGAAATAATA
cpHap_5 TCTCGATTTCTATTTTTTT-ATAAACAAAATCGAAATGTTGTGAATCAATTCGAAGTTTAAGAAATAATA
cpHap_6 TCTCGATTTCTATTTTTTT-ATAAACAAAATCGAAATGTTGTGAATCAATTCGAAGTTTAAGAAATAATA
cpHap_7 TCTCGATTTCTATTTTTTT-ATAAACAAAATCGAAATGTTGTGAATCAATTCGAAGTTTAAGAAATAATA
cpHap_8 TCTCGATTTCTATTTTTTT-ATAAACAAAATCGAAATGTTGTGAATCAATTCGAAGTTTAAGAAATAATA
cpHap_9 TCTCGATTTCTATTTTTTT-ATAAACAAAATCGAAATGTTGTGAATCAATTCGAAGTTTAAGAAATAATA
cpHap_10 TCTCGATTTCTATTTTTTT-ATAAACAAAATCGAAATGTTGTGAATCAATTCGAAGTTTAAGAAATAATA
cpHap_11 TCTCGATTTCTATTTTTTT-ATAAACAAAATCGAAATGTTGTGAATCAATTCGAAGTTTAAGAAATAATA
cpHap_12 TCTCGATTTCTATTTTTTT-ATAAACAAAATCGAAATGTTGTGAATCAATTCGAAGTTTAAGAAATAATA
cpHap_13 TCTCGATTTCTATTTTTT--ATAAACAAAATCGAAATGTTGTGAATCAATTCGAAGTTTAAGAAATAATA
cpHap_14 TCTCGATTTCTATTTTTT--ATAAACAAAATCGAAATGTTGTGAATCAATTCGAAGTTTAAGAAATAATA
cpHap_15 TCTCGATTTCTATTTTTTT-ATAAACAAAATCGAAATGTTGTGAATCAATTCGAAGTTTAAGAAATAATA
cpHap_16 TCTCGATTTCTATTTTTTT-ATAAACAAAATCGAAATGTTGTGAATCAATTCGAAGTTTAAGAAATAATA
cpHap_17 TCTCGATTTCTATTTTTTT-ATAAACAAAATCGAAATGTTGTGAATCAATTCGAAGTTTAAGAAATAATA
cpHap_18 TCTCGATTTCTATTTTTTT-ATAAACAAAATCGAAATGTTGTGAATCAATTCGAAGTTTAAGAAATAATA
360 370 380 390 400 410 420
cpHap_1 TTTATTGATCAAATGATTCACTTCATAGTCTGATAGATCCTTGATGGA-----ATTAATCGGACGAGAAT
cpHap_2 TTTATTGATCAAATGATTCACTTCATAGTCTGATAGATCCTTGATGGA-----ATTAATCGGACGAGAAT
cpHap_3 TTTATTGATCAAATGATTCACTTCATAGTCTGATAGATCCTTGATGGA-----ATTAATCGGACGAGAAT
cpHap_4 TTTATTGATCAAATGATTCACTTCATAGTCTGATAGATCCTTGATGGA-----ATTAATCGGACGAGAAT
cpHap_5 TTTATTGATCAAATGATTCACTTCATAGTCTGATAGATCCTTGATGGA-----ATTAATCGGACGAGAAT
cpHap_6 TTTATTGATCAAATGATTCACTTCATAGTCTGATAGATCCTTGATGGA-----ATTAATCGGACGAGAAT
cpHap_7 TTTATTGATCAAATGATTCACTTCATAGTCTGATAGATCCTTGATGGA-----ATTAATCGGACGAGAAT
cpHap_8 TTTATTGATCAAATGATTCACTTCATAGTCTGATAGATCCTTGATGGAACTTAATTAATCGGACGAGAAT
cpHap_9 TTTATTGATCAAATGATTCACTTCATAGTCTGATAGATCCTTGATGGA-----ATTAATCGGACGAGAAT
cpHap_10 TTTATTGATCAAATGATTCACTTCATAGTCTGATAGATCCTTGATGGA-----ATTAATCGGACGAGAAT
cpHap_11 TTTATTGATCAAATGATTCACTTCATAGTCTGATAGATCCTTGATGGAACTTAATTAATCGGACGAGAAT
cpHap_12 TTTATTGATCAAATGATTCACTTCATAGTCTGATAGATCCTTGATGGA-----ATTAATCGGACGAGAAT
cpHap_13 TTTATTGATCAAATGATTCACTTCATAGTCTGATAGATCCTTGATGGA-----ATTAATCGGACGAGAAT
cpHap_14 TTTATTGATCAAATGATTCACTTCATAGTCTGATAGATCCTTGATGGA-----ATTAATCGGACGAGAAT
cpHap_15 TTTATTGATCAAATGATTCACTTCATAGTCTGATAGATCCTTGATGGA-----ATTAATCGGACGAGAAT
cpHap_16 TTTATTGATCAAATGATTCACTTCATAGTCTGATAGATCCTTGATGGA-----ATTAATCGGACGAGAAT
cpHap_17 TTTATTGATCAAATGATTCACTTCATAGTCTGATAGATCCTTGATGGA-----ATTAATCGGACGAGAAT
cpHap_18 TTTATTGATCAAATGATTCACTTCATAGTCTGATAGATCCTTGATGGA-----ATTAATCGGACGAGAAT

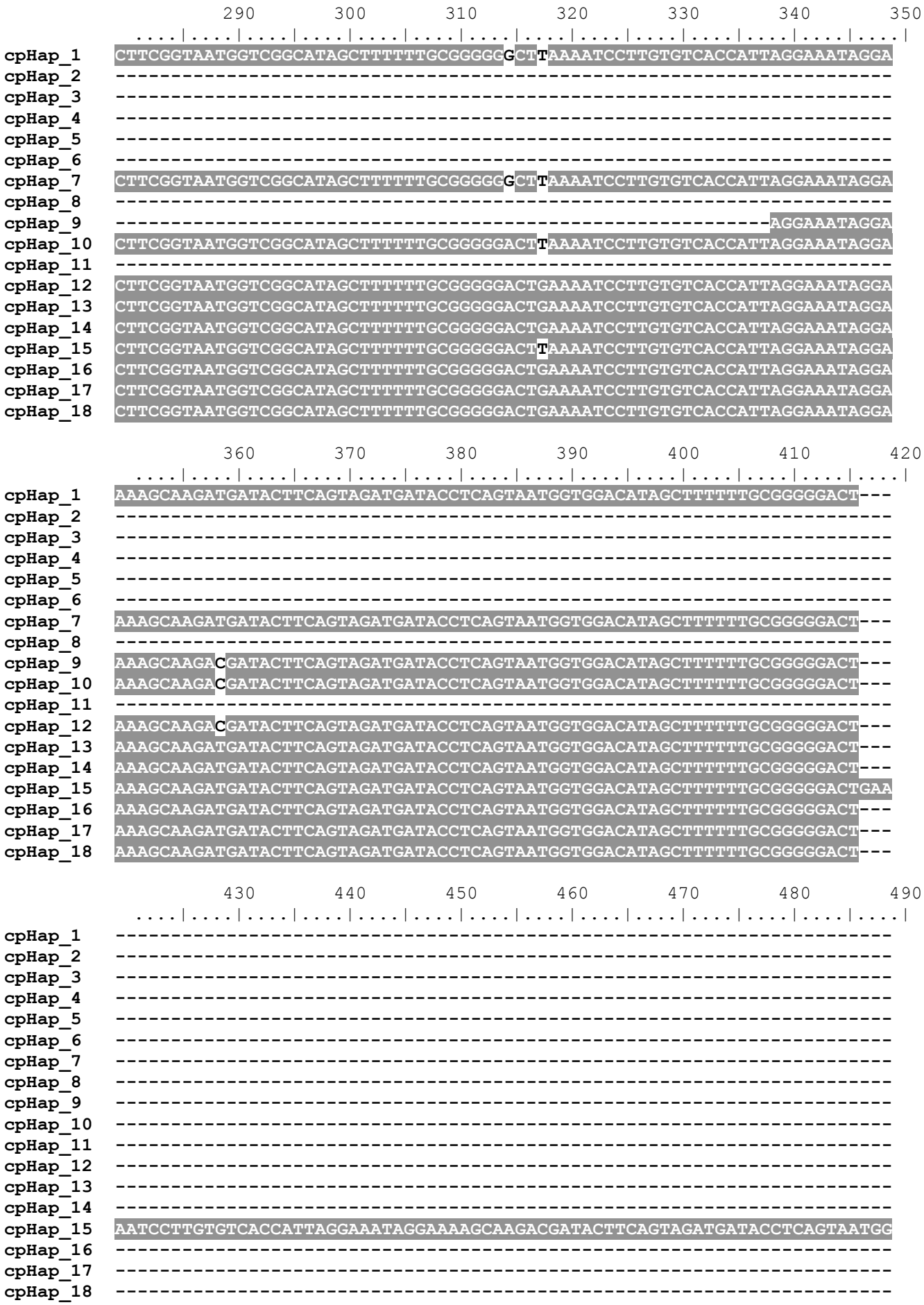


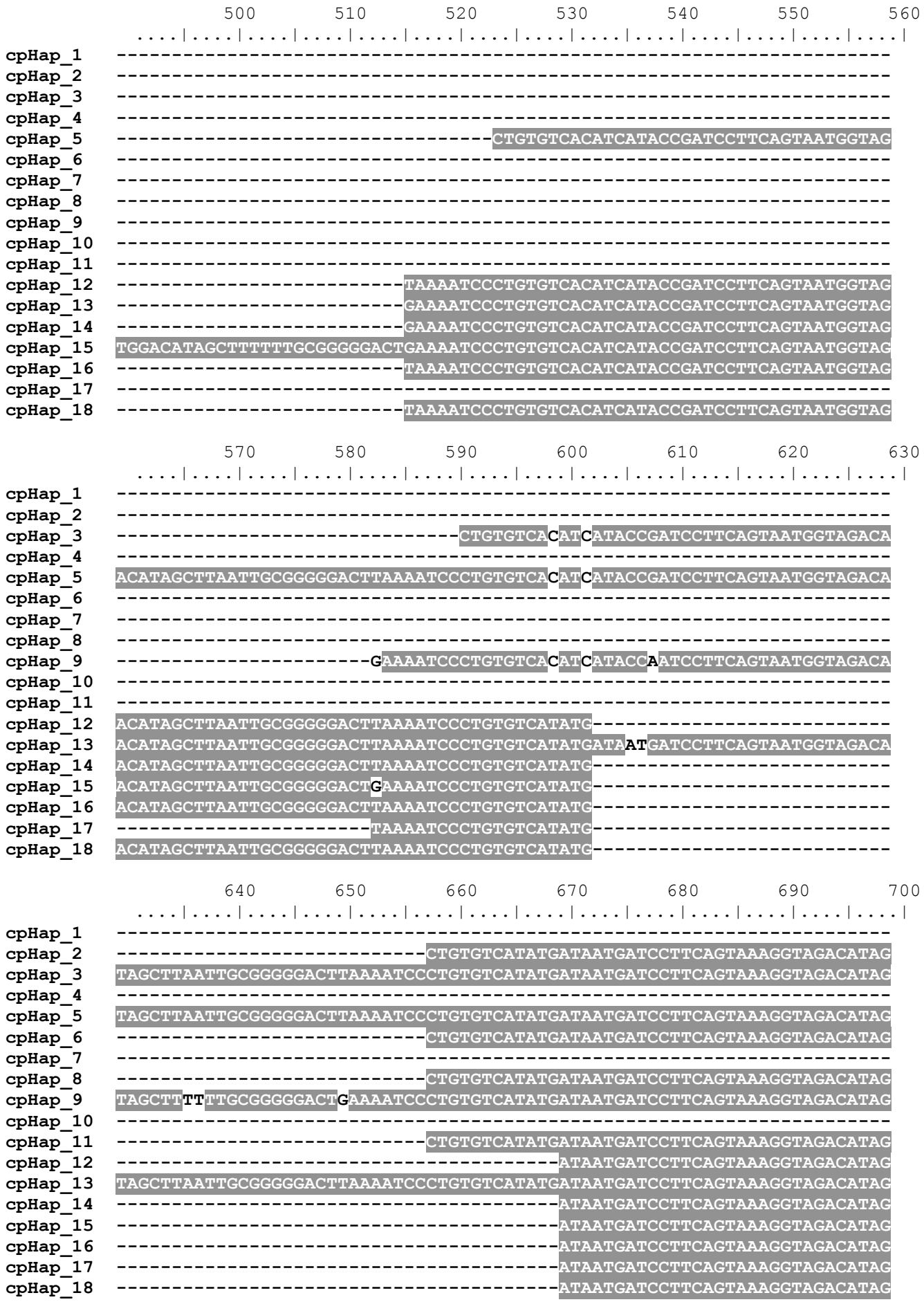


***trnL-trnF* intergenic region**




80 90 100 110 120 130 140
cpHap_1 AGTCTTCTTTTATTTATAAAATCCAAGAAATTCCCGGTCCACAACTTTTTGAATTTACTACTTTTGAGTT
cpHap_2 AGTCTTCTTTTATTTATAAAATCCAAGAAATTCCCGGTCCACAACTTTTTGAATTTACTACTTTTGAGTT
cpHap_3 AGTCTTCTTTTATTTATAAAATCCAAGAAATTCCCGGTCCACAACTTTTTGAATTTACTACTTTTGAGTT
cpHap_4 AGTCTTCTTTTATTTATAAAATCCAAGAAATTCCCGGTCCACAACTTTTTGAATTTACTACTTTTGAGTT
cpHap_5 AGTCTTCTTTTATTTATAAAATCCAAGAAATTCCCGGTCCACAACTTTTTGAATTTACTACTTTTGAGTT
cpHap_6 AGTCTTCTTTTATTTATAAAATCCAAGAAATTCCCGGTCCACAACTTTTTGAATTTACTACTTTTGAGTT
cpHap_7 AGTCTTCTTTTATTTATAAAATCCAAGAAATGCCCGGTCCACAACTTTTTGAATTTACTACTTTTGAGTT
cpHap_8 AGTCTTCTTTTATTTATAAAATCCAAGAAATTCCCGGTCCACAACTTTTTGAATTTACTACTTTTGAGTT
cpHap_9 AGTCTTCTTTTATTTATAAAATCCAAGAAATTCCCGGTCCACAACTTTTTGAATTTACTACTTTTGAGTT
cpHap_10 AGTTTTCTTTTATTTATAAAATCCAAGAAATTCCCGGTCCACAACTTTTTGAATTTACTACTTTTGAGTT
cpHap_11 AGTCTTCTTTTATTTATAAAATCCAAGAAATTCCCGGTCCACAACTTTTTGAATTTACTACTTTTGAGTT
cpHap_12 AGTCTTCTTTTATTTATAAAATCCAAGAAATTCCCGGTCCACAACTTTTTGAATTTACTACTTTTGAGTT
cpHap_13 AGTCTTCTTTTATTTATAAAATCCAAGAAATTCCCGGTCCACAACTTTTTGAATTTACTACTTTTGAGTT
cpHap_14 AGTCTTCTTTTATTTATAAAATCCAAGAAATTCCCGGTCCACAACTTTTTGAATTTACTACTTTTGAGTT
cpHap_15 AGTCTTCTTTTATTTATAAAATCCAAGAAATTCCCGGTCCACAACTTTTTGAATTTACTACTTTTGAGTT
cpHap_16 AGTCTTCTTTTATTTATAAAATCCAAGAAATTCCCGGTCCACAACTTTTTGAATTTACTACTTTTGAGTT
cpHap_17 AGTCTTCTTTTATTTATAAAATCCAAGAAATTCCCGGTCCACAACTTTTTGAATTTACTACTTTTGAGTT
cpHap_18 AGTCTTCTTTTATTTATAAAATCCAAGAAATTCCCGGTCCACAACTTTTTGAATTTACTACTTTTGAGTT
150 160 170 180 190 200 210
cpHap_1 TCTTTTCATTGACATAGACCTAAGTCATATATTAAAATGATACTGATACTTCAGTAGATGATACTTCGGT
cpHap_2 TCTTTTCATTGACATAGACCTAAGTCATATATTAAAATGATACTGATACTTCAGTAGATGATACTTCGGT
cpHap_3 TCTTTTCATTGACATAGACCTAAGTCATATATTAAAATGATACTGATACTTCAGTAGATGATACTTCGGT
cpHap_4 TCTTTTCATTGACATAGACCTAAGTCATATATTAAAATGATACTGATACTTCAGTAGATGATACTTCGGT
cpHap_5 TCTTTTCATTGACATAGACCTAAGTCATATATTAAAATGATACTGATACTTCAGTAGATGATACTTCGGT
cpHap_6 TCTTTTCATTGACATAGACCTAAGTCATATATTAAAATGATACTGATACTTCAGTAGATGATACTTCGGT
cpHap_7 TCTTTTCATTGACATAGACCTAAGTCATATATTAAAATGATACTGATACTTCAGTAGATGATACTTCGGT
cpHap_8 TCTTTTCATTGACATAGACCTAAGTCATATATTAAAATGATACTGATACTTCAGTAGATGATACTTCGGT
cpHap_9 TCTTTTCATTGACATAGACCTAAGTCATATATTAAAATGATACTGATACTTCAGTAGATGATACTTCGGT
cpHap_10 TCTTTTCATTGACATAGACCTAAGTCATATATTAAAATGATACTGATACTTCAGTAGATGATACTTCGGT
cpHap_11 TCTTTTCATTGACATAGACCTAAGTCATATATTAAAATGATACTGATACTTCAGTAGATGATACTTCGGT
cpHap_12 TCTTTTCATTGACATAGACCTAAGTCATATATTAAAATGATACTGATACTTCAGTAGATGATACTTCGGT
cpHap_13 TCTTTTCATTGACATAGACCTAAGTCATATATTAAAATGATACTGATACTTCAGTAGATGATACTTCGGT
cpHap_14 TCTTTTCATTGACATAGACCTAAGTCATATATTAAAATGATACTGATACTTCAGTAGATGATACTTCGGT
cpHap_15 TCTTTTCATTGACATAGACCTAAGTCATATATTAAAATGATACTGATACTTCAGTAGATGATACTTCGGT
cpHap_16 TCTTTTCATTGACATAGACCTAAGTCATATATTAAAATGATACTGATACTTCAGTAGATGATACTTCGGT
cpHap_17 TCTTTTCATTGACATAGACCTAAGTCATATATTAAAATGATACTGATACTTCAGTAGATGATACTTCGGT
cpHap_18 TCTTTTCATTGACATAGACCTAAGTCATATATTAAAATGATACTGATACTTCAGTAGATGATACTTCGGT
220 230 240 250 260 270 280
cpHap_1 AATGGTAGACATAGCTTAATTGCGGGGGACTGAAAATCCTTGTGTCACCATTCGGAAAAGCAAGATGATA
cpHap_2 AATGGTAGACATAGCTTAATTGCGGGGGACTTAAAATCC----------
cpHap_3 AATGGTAGACATAGCTTAATTGCGGGGGACTTAAAATCC----------
cpHap_4 AATGGTAGACATAGCTTAATTGCGGGGGACT----------
cpHap_5 AATGGTAGACATAGCTTAATTGCGGGGGACTTAAAATCC----------
cpHap_6 AATGGTAGACATAGCTTAATTGCGGGGGACTGAAAATCC----------
cpHap_7 AATGGTAGACATAGCTTAATTGCGGGGGACTGAAAATCCTTGTGTCACCATTCGGAAAAGCAAGATGATA
cpHap_8 AATGGTAGACATAGCTTAATTGCGGGGGACTGAAAATCC----------
cpHap_9 AATGGTAGACATAGCTTAATTGCGGGGGACTGAAAATCCTTGTGTCACCATT----------
cpHap_10 AATGGTAGACATAGCTTAATTGCGGGGGACTTAAAATCCTTGTGTCACCATTCGGAAAAGCAAGATGATA
cpHap_11 AATGGTAGACATAGCTTAATTGCGGGGGACTGAAAATCC----------
cpHap_12 AATGGTAGACATAGCTTAATTGCGGGGGACTTAAAATCCTTGTGTCACCATTCGGAAAAGCAAGATGATA
cpHap_13 AATGGTAGACATAGCTTAATTGCGGGGGACTGAAAATCCTTGTGTCACCATTCGGAAAAGCAAGATGATA
cpHap_14 AATGGTAGACATAGCTTAATTGCGGGGGACTGAAAATCCTTGTGTCACCATTCGGAAAAGCAAGATGATA
cpHap_15 AATGGTAGACATAGCTTAATTGCGGGGGACTGAAAATCCTTGTGTCACCATTCGGAAAAGCAAGATGATA
cpHap_16 AATGGTAGACATAGCTTAATTGCGGGGGACTTAAAATCCTTGTGTCACCATTCGGAAAAGCAAGATGATA
cpHap_17 AATGGTAGACATAGCTTAATTGCGGGGGACTTAAAATCCTTGTGTCACCATTCGGAAAAGCAAGATGATA
cpHap_18 AATGGTAGACATAGCTTAATTGCGGGGGACTTAAAATCCTTGTGTCACCATTCGGAAAAGCAAGATGATA


290 300 310 320 330 340 350
cpHap_1 CTTCGGTAATGGTCGGCATAGCTTTTTTGCGGGGGGCTTAAAATCCTTGTGTCACCATTAGGAAATAGGA
cpHap_2 ----------
cpHap_3 ----------
cpHap_4 ----------
cpHap_5 ----------
cpHap_6 ----------
cpHap_7 CTTCGGTAATGGTCGGCATAGCTTTTTTGCGGGGGGCTTAAAATCCTTGTGTCACCATTAGGAAATAGGA
cpHap_8 ----------
cpHap_9 ----------AGGAAATAGGA
cpHap_10 CTTCGGTAATGGTCGGCATAGCTTTTTTGCGGGGGACTTAAAATCCTTGTGTCACCATTAGGAAATAGGA
cpHap_11 ----------
cpHap_12 CTTCGGTAATGGTCGGCATAGCTTTTTTGCGGGGGACTGAAAATCCTTGTGTCACCATTAGGAAATAGGA
cpHap_13 CTTCGGTAATGGTCGGCATAGCTTTTTTGCGGGGGACTGAAAATCCTTGTGTCACCATTAGGAAATAGGA
cpHap_14 CTTCGGTAATGGTCGGCATAGCTTTTTTGCGGGGGACTGAAAATCCTTGTGTCACCATTAGGAAATAGGA
cpHap_15 CTTCGGTAATGGTCGGCATAGCTTTTTTGCGGGGGACTTAAAATCCTTGTGTCACCATTAGGAAATAGGA
cpHap_16 CTTCGGTAATGGTCGGCATAGCTTTTTTGCGGGGGACTGAAAATCCTTGTGTCACCATTAGGAAATAGGA
cpHap_17 CTTCGGTAATGGTCGGCATAGCTTTTTTGCGGGGGACTGAAAATCCTTGTGTCACCATTAGGAAATAGGA
cpHap_18 CTTCGGTAATGGTCGGCATAGCTTTTTTGCGGGGGACTGAAAATCCTTGTGTCACCATTAGGAAATAGGA
360 370 380 390 400 410 420
cpHap_1 AAAGCAAGATGATACTTCAGTAGATGATACCTCAGTAATGGTGGACATAGCTTTTTTGCGGGGGACT---
cpHap_2 ----------
cpHap_3 ----------
cpHap_4 ----------
cpHap_5 ----------
cpHap_6 ----------
cpHap_7 AAAGCAAGATGATACTTCAGTAGATGATACCTCAGTAATGGTGGACATAGCTTTTTTGCGGGGGACT---
cpHap_8 ----------
cpHap_9 AAAGCAAGACGATACTTCAGTAGATGATACCTCAGTAATGGTGGACATAGCTTTTTTGCGGGGGACT---
cpHap_10 AAAGCAAGACGATACTTCAGTAGATGATACCTCAGTAATGGTGGACATAGCTTTTTTGCGGGGGACT---
cpHap_11 ----------
cpHap_12 AAAGCAAGACGATACTTCAGTAGATGATACCTCAGTAATGGTGGACATAGCTTTTTTGCGGGGGACT---
cpHap_13 AAAGCAAGATGATACTTCAGTAGATGATACCTCAGTAATGGTGGACATAGCTTTTTTGCGGGGGACT---
cpHap_14 AAAGCAAGATGATACTTCAGTAGATGATACCTCAGTAATGGTGGACATAGCTTTTTTGCGGGGGACT---
cpHap_15 AAAGCAAGATGATACTTCAGTAGATGATACCTCAGTAATGGTGGACATAGCTTTTTTGCGGGGGACTGAA
cpHap_16 AAAGCAAGATGATACTTCAGTAGATGATACCTCAGTAATGGTGGACATAGCTTTTTTGCGGGGGACT---
cpHap_17 AAAGCAAGATGATACTTCAGTAGATGATACCTCAGTAATGGTGGACATAGCTTTTTTGCGGGGGACT---
cpHap_18 AAAGCAAGATGATACTTCAGTAGATGATACCTCAGTAATGGTGGACATAGCTTTTTTGCGGGGGACT---
430 440 450 460 470 480 490
cpHap_1 ----------
cpHap_2 ----------
cpHap_3 ----------
cpHap_4 ----------
cpHap_5 ----------
cpHap_6 ----------
cpHap_7 ----------
cpHap_8 ----------
cpHap_9 ----------
cpHap_10 ----------
cpHap_11 ----------
cpHap_12 ----------
cpHap_13 ----------
cpHap_14 ----------
cpHap_15 AATCCTTGTGTCACCATTAGGAAATAGGAAAAGCAAGACGATACTTCAGTAGATGATACCTCAGTAATGG
cpHap_16 ----------
cpHap_17 ----------
cpHap_18 ----------


500 510 520 530 540 550 560
cpHap_1 ----------
cpHap_2 ----------
cpHap_3 ----------
cpHap_4 ----------
cpHap_5 ----------CTGTGTCACATCATACCGATCCTTCAGTAATGGTAG
cpHap_6 ----------
cpHap_7 ----------
cpHap_8 ----------
cpHap_9 ----------
cpHap_10 ----------
cpHap_11 ----------
cpHap_12 ----------TAAAATCCCTGTGTCACATCATACCGATCCTTCAGTAATGGTAG
cpHap_13 ----------GAAAATCCCTGTGTCACATCATACCGATCCTTCAGTAATGGTAG
cpHap_14 ----------GAAAATCCCTGTGTCACATCATACCGATCCTTCAGTAATGGTAG
cpHap_15 TGGACATAGCTTTTTTGCGGGGGACTGAAAATCCCTGTGTCACATCATACCGATCCTTCAGTAATGGTAG
cpHap_16 ----------TAAAATCCCTGTGTCACATCATACCGATCCTTCAGTAATGGTAG
cpHap_17 ----------
cpHap_18 ----------TAAAATCCCTGTGTCACATCATACCGATCCTTCAGTAATGGTAG
570 580 590 600 610 620 630
cpHap_1 ----------
cpHap_2 ----------
cpHap_3 ----------CTGTGTCACATCATACCGATCCTTCAGTAATGGTAGACA
cpHap_4 ----------
cpHap_5 ACATAGCTTAATTGCGGGGGACTTAAAATCCCTGTGTCACATCATACCGATCCTTCAGTAATGGTAGACA
cpHap_6 ----------
cpHap_7 ----------
cpHap_8 ----------
cpHap_9 ----------GAAAATCCCTGTGTCACATCATACCATCCTTCAGTAATGGTAGACA
cpHap_10 ----------
cpHap_11 ----------
cpHap_12 ACATAGCTTAATTGCGGGGGACTTAAAATCCCTGTGTCATATG----------
cpHap_13 ACATAGCTTAATTGCGGGGGACTTAAAATCCCTGTGTCATATGATAATGATCCTTCAGTAATGGTAGACA
cpHap_14 ACATAGCTTAATTGCGGGGGACTTAAAATCCCTGTGTCATATG----------
cpHap_15 ACATAGCTTAATTGCGGGGGACTGAAAATCCCTGTGTCATATG----------
cpHap_16 ACATAGCTTAATTGCGGGGGACTTAAAATCCCTGTGTCATATG----------
cpHap_17 ----------TAAAATCCCTGTGTCATATG----------
cpHap_18 ACATAGCTTAATTGCGGGGGACTTAAAATCCCTGTGTCATATG----------
640 650 660 670 680 690 700
cpHap_1 ----------
cpHap_2 ----------CTGTGTCATATGATAATGATCCTTCAGTAAAGGTAGACATAG
cpHap_3 TAGCTTAATTGCGGGGGACTTAAAATCCCTGTGTCATATGATAATGATCCTTCAGTAAAGGTAGACATAG
cpHap_4 ----------
cpHap_5 TAGCTTAATTGCGGGGGACTTAAAATCCCTGTGTCATATGATAATGATCCTTCAGTAAAGGTAGACATAG
cpHap_6 ----------CTGTGTCATATGATAATGATCCTTCAGTAAAGGTAGACATAG
cpHap_7 ----------
cpHap_8 ----------CTGTGTCATATGATAATGATCCTTCAGTAAAGGTAGACATAG
cpHap_9 TAGCTTTTTTGCGGGGGACTGAAAATCCCTGTGTCATATGATAATGATCCTTCAGTAAAGGTAGACATAG
cpHap_10 ----------
cpHap_11 ----------CTGTGTCATATGATAATGATCCTTCAGTAAAGGTAGACATAG
cpHap_12 ----------ATAATGATCCTTCAGTAAAGGTAGACATAG
cpHap_13 TAGCTTAATTGCGGGGGACTTAAAATCCCTGTGTCATATGATAATGATCCTTCAGTAAAGGTAGACATAG
cpHap_14 ----------ATAATGATCCTTCAGTAAAGGTAGACATAG
cpHap_15 ----------ATAATGATCCTTCAGTAAAGGTAGACATAG
cpHap_16 ----------ATAATGATCCTTCAGTAAAGGTAGACATAG
cpHap_17 ----------ATAATGATCCTTCAGTAAAGGTAGACATAG
cpHap_18 ----------ATAATGATCCTTCAGTAAAGGTAGACATAG

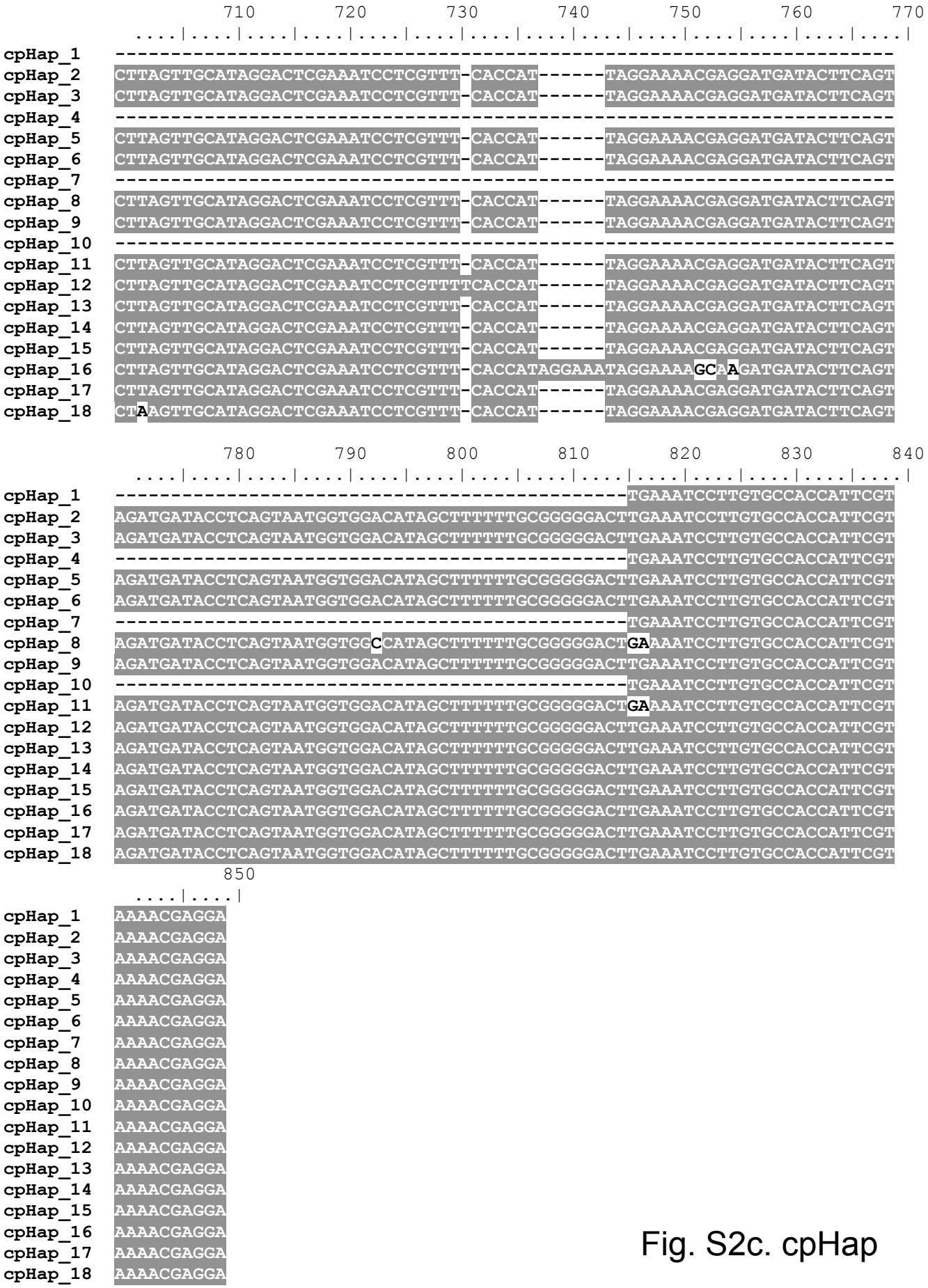
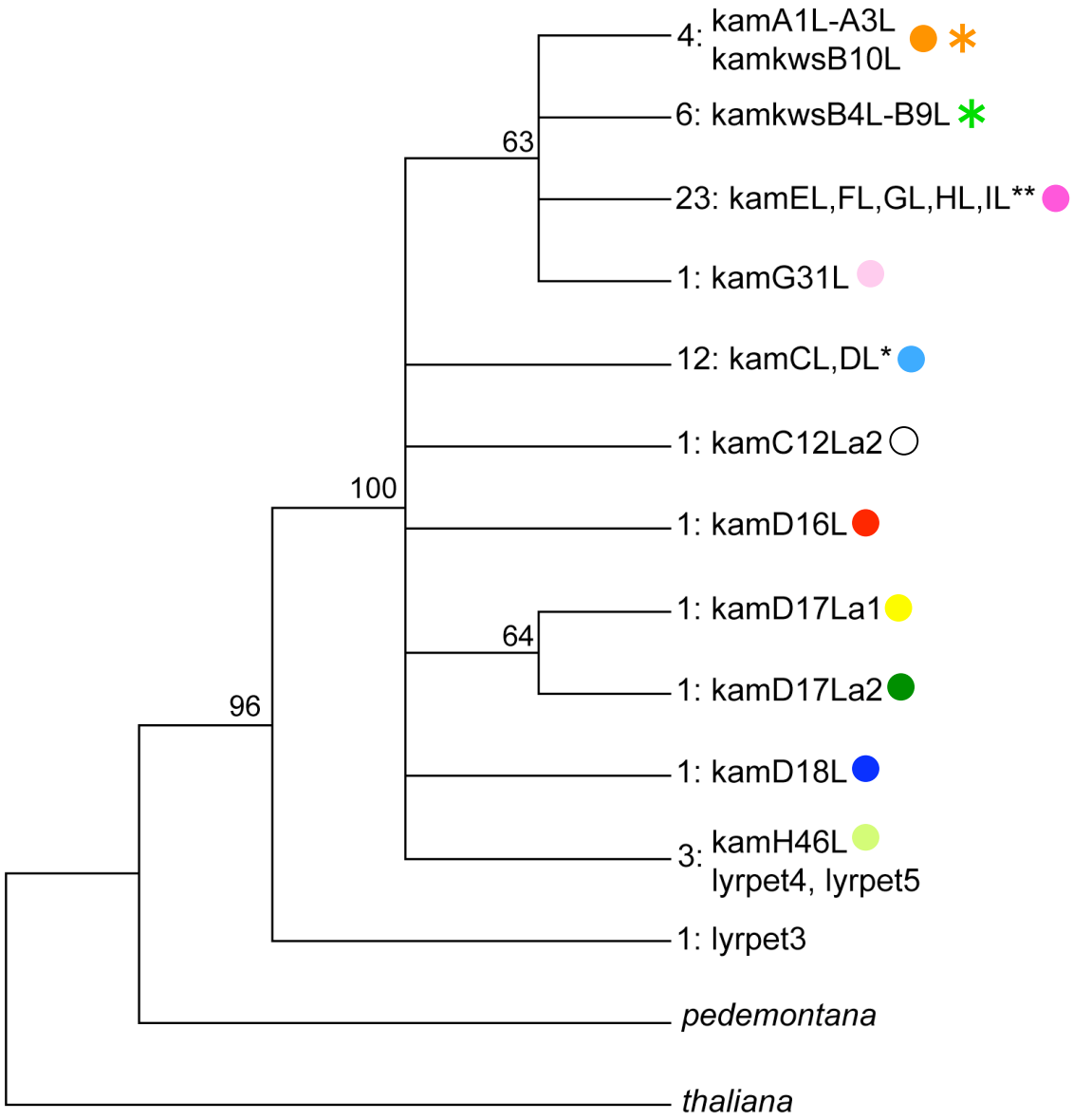
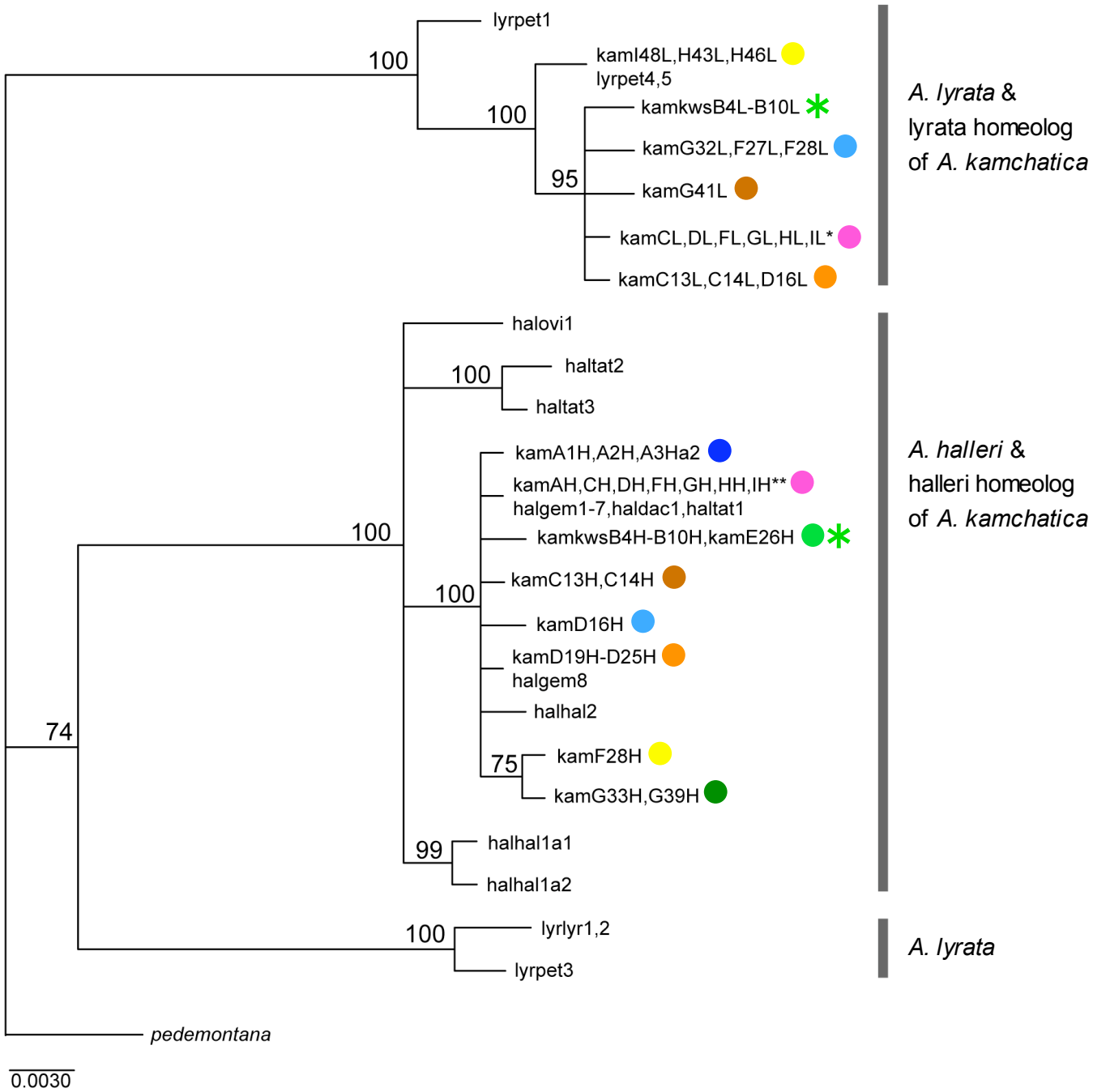


Fig. S2c. cpHap



*kamC11L, C12La1, C13L, C14L, D15L, D19L-D25L
**kamE26L, F27L-F30L, G32L-G41L, H42L-H45L, H47L, I48L-I50L

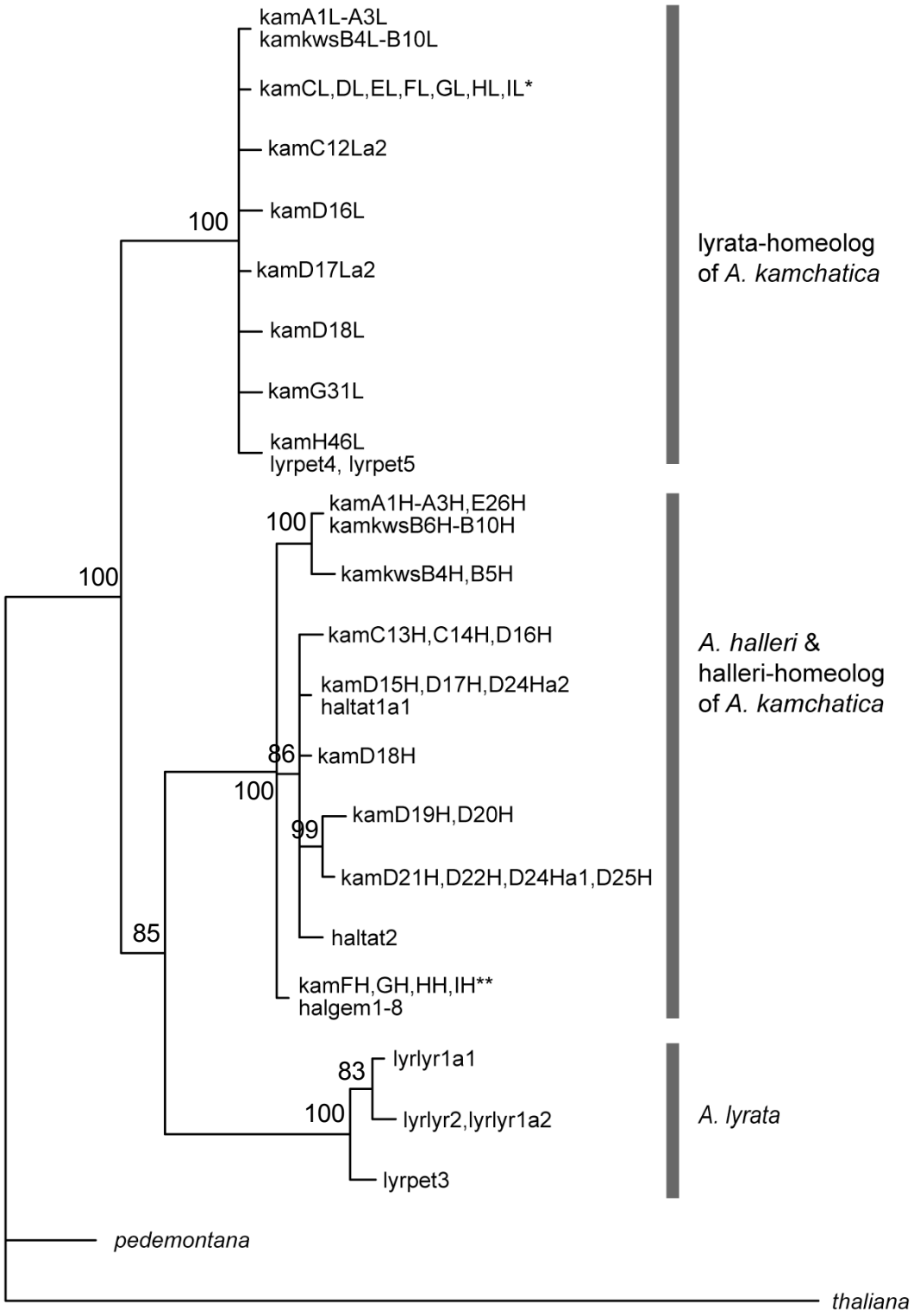
Supporting Fig. S3a *CHS* most-parsimonious tree, lyrata clade



*kamC11L, C12L, D15L, D17L-D25L, F29L, F30L, G31L, G33L-G40L, H42L, H44L, H45L, H47L, I49L, I50L

**kamA3Ha1, C11H, C12H, D15H, D17H, D18H, F27H, F29H, F30H, G31H, G32H, G34H-G38H, G40H, G41H, H42H-47H, I48H-I50H

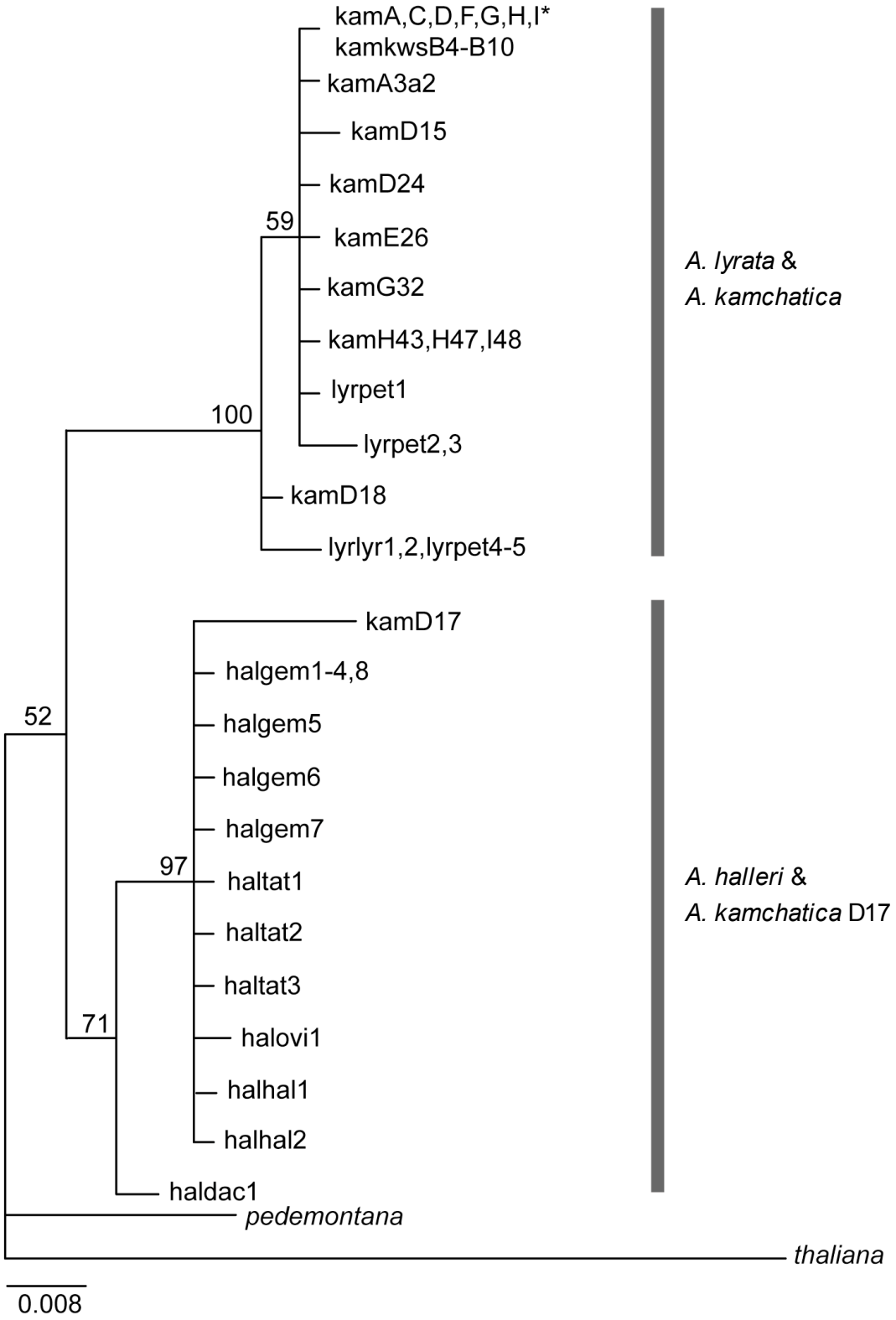
Supporting Fig. S3b *WER* Bayesian tree



*kamC11L, C12La1, C13L, C14L, D15L, D17La1, D19L-D25L, E26L, F27L-F30L, G32L-G41L, H42L-H45L, H47L, I48L-I50L

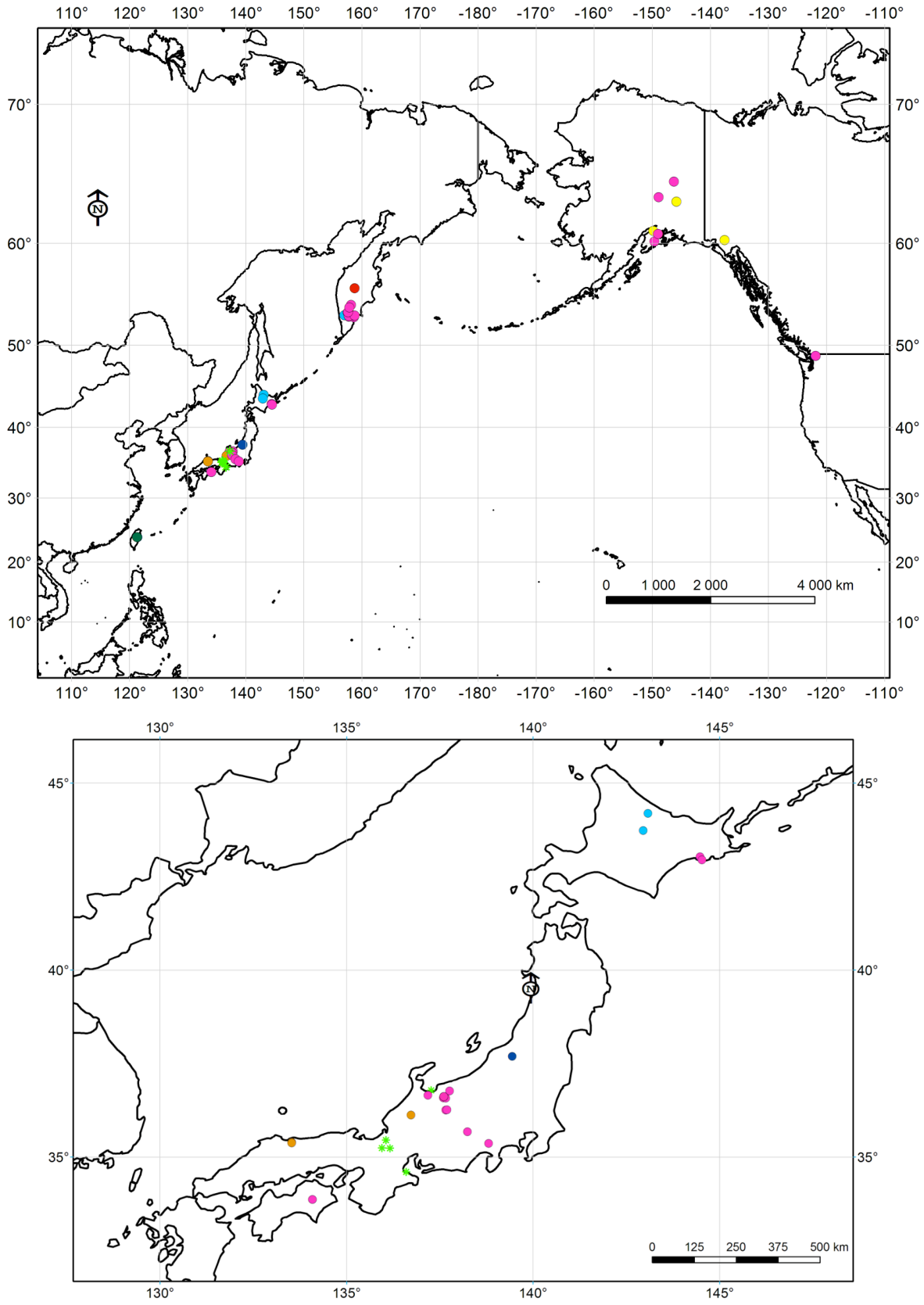
**kamF27H-F30H, G31H-G41, H42H-H47H, I48H-I50H

Supporting Fig. S3c *CHS* Bayesian tree

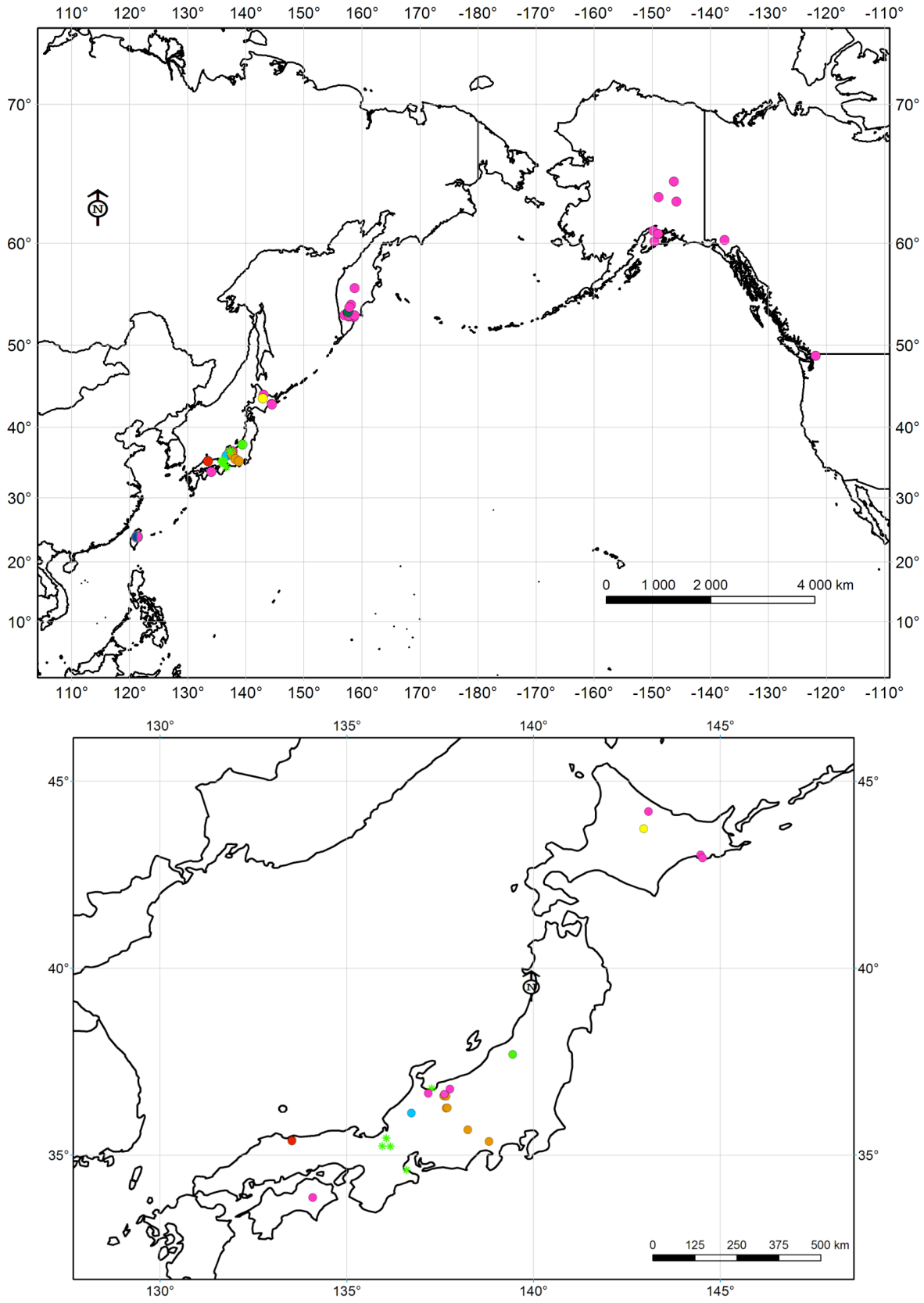


*kamA1, A2, A3a1, C11-C14, D16, D19-D23, D25, F27-F30, G31, G33-G41, H42, H44-H46, I49, I50

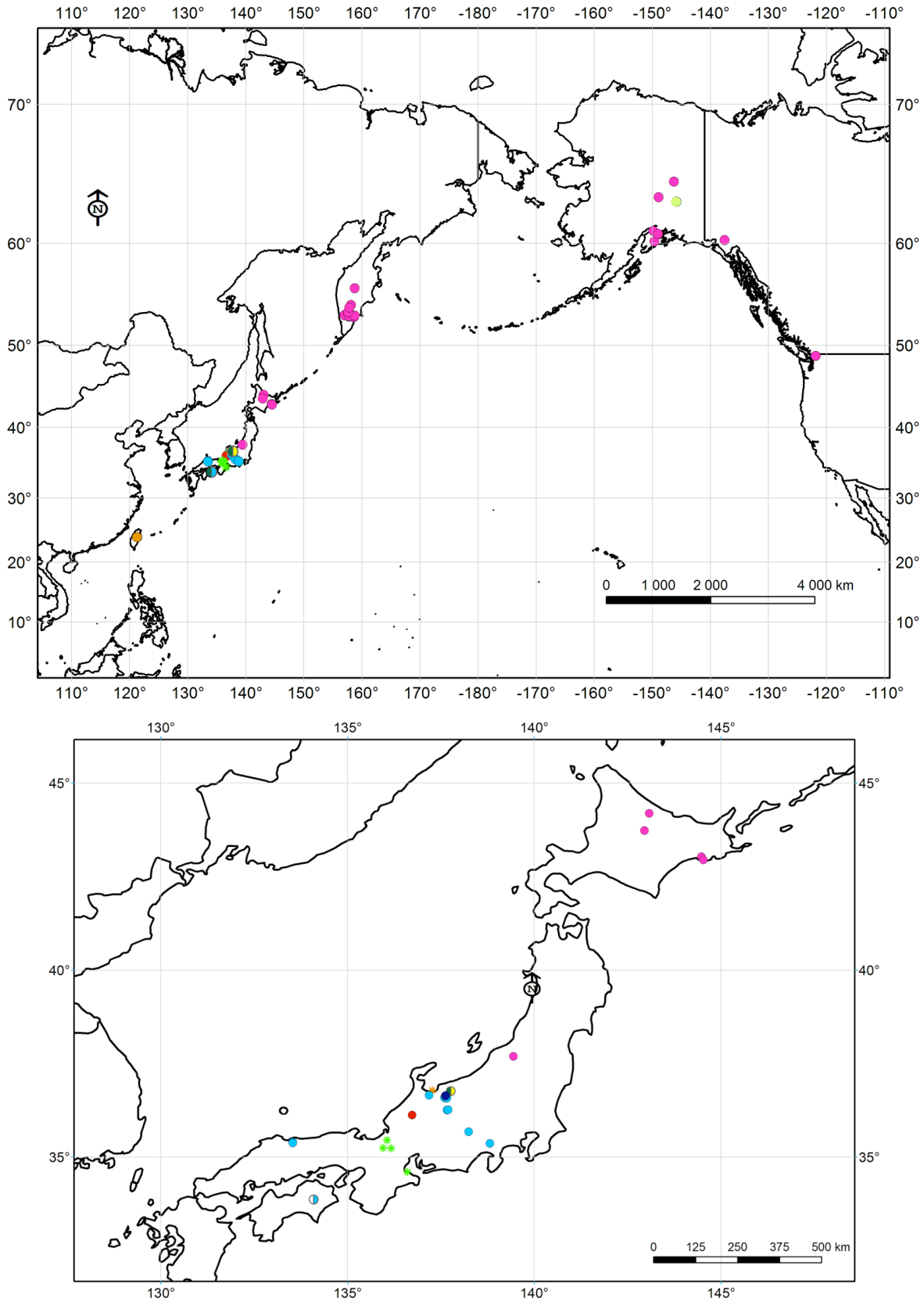
Supporting Fig. S3d ITS Bayesian tree



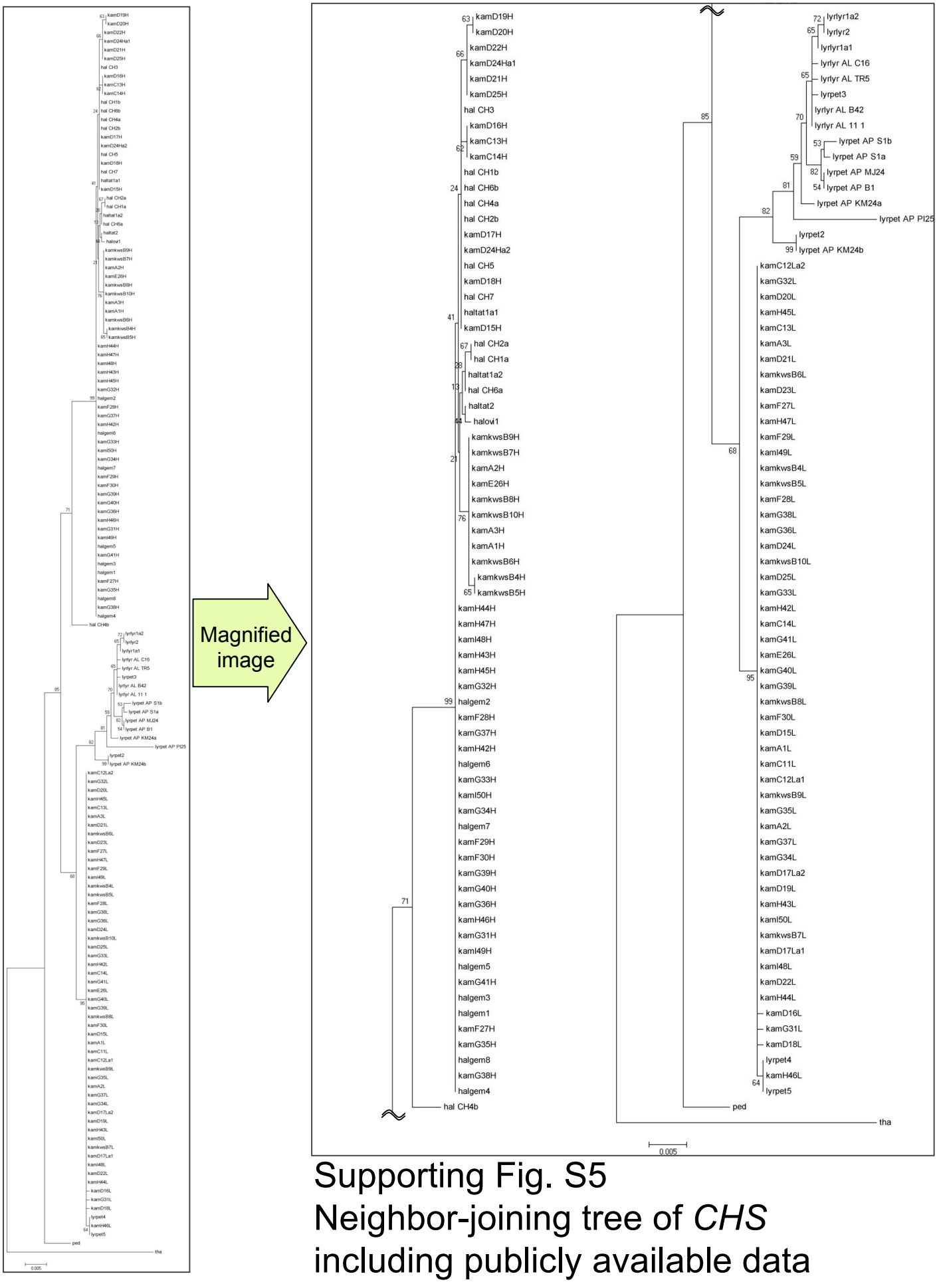
Supporting Fig. S4a *WER lyrata* type



Supporting Fig. S4b *WER halleri* type



Supporting Fig. S4c *CHS lyrata* type



Supporting Fig. S5
Neighbor-joining tree of *CHS* including publicly available data