

RESEARCH ARTICLE

Open Access

Interspecific and interploidal gene flow in Central European *Arabidopsis* (Brassicaceae)

Marte H Jørgensen^{1*}, Dorothee Ehrich², Roswitha Schmickl³, Marcus A Koch³ and Anne K Brysting¹

Abstract

Background: Effects of polyploidisation on gene flow between natural populations are little known. Central European diploid and tetraploid populations of *Arabidopsis arenosa* and *A. lyrata* are here used to study interspecific and interploidal gene flow, using a combination of nuclear and plastid markers.

Results: Ploidal levels were confirmed by flow cytometry. Network analyses clearly separated diploids according to species. Tetraploids and diploids were highly intermingled within species, and some tetraploids intermingled with the other species, as well. Isolation with migration analyses suggested interspecific introgression from tetraploid *A. arenosa* to tetraploid *A. lyrata* and vice versa, and some interploidal gene flow, which was unidirectional from diploid to tetraploid in *A. arenosa* and bidirectional in *A. lyrata*.

Conclusions: Interspecific genetic isolation at diploid level combined with introgression at tetraploid level indicates that polyploidy may buffer against negative consequences of interspecific hybridisation. The role of introgression in polyploid systems may, however, differ between plant species, and even within the small genus *Arabidopsis*, we find very different evolutionary fates when it comes to introgression.

Background

Polyploidy, i.e. whole genome duplication, has long been considered a major evolutionary force in the Plant Kingdom [see e.g., [1-5]], and even though large advances in our understanding of polyploidy have been made during the last couple of decades, there are still many questions unanswered [reviewed by [6]]. We still don't have a general agreement on classification of polyploids, for instance. Some authors work with strict taxonomic definitions; autopolyploids are the result of polyploidisation events involving only a single species, and allopolyploids are the result of interspecific hybridisation [e.g., [3]]. Others base their definitions on inheritance patterns and the presence or absence of multivalents [e.g., [7]]. However, most would agree that auto- and allopolyploids are the extremes of a continuous range. There are also still controversies about how polyploids should be treated taxonomically. Soltis et al. [8] suggest that autopolyploids deserve species rank taxonomically, with ploidal level as part of the name. Others do not even

give allopolyploids species status due to lack of morphological distinctness (e.g., lack of diagnostic qualitative and discrete characters), and include them as subspecies of one of the parents [e.g., [9]]. Yet others separate morphological and biological species where the first may contain several of the latter [e.g., [10]].

Traditionally, polyploidisation events have been considered to result in total reproductive isolation of the new polyploid from the parent (s), and thus regarded as instant speciation [e.g., [11]]. More recent research has shown that recurrent formation of polyploids and triploid bridges contribute to interploidal gene flow [3,12,13]. To what extent, however, is still not known [6]. Multiple independent polyploidisation events have been shown to be common for both allopolyploids [e.g., [14,15]] and autopolyploids [e.g., [16-18]]. Population studies and modelling of sympatric *Chamerion angustifolium* (L.) Holub revealed that autotetraploids are not necessarily instantly isolated from their diploid progenitors, but that the isolation can become more prevalent through time [19,20]. Slotte et al. [21] showed that there is unidirectional gene flow from diploid *Capsella rubella* Reuter to its allotetraploid descendant *C. bursa-pastoris* (L.) Medicus. Furthermore, if polyploidisation events

* Correspondence: martej@ulrik.uio.no

¹Centre for Ecological and Evolutionary Synthesis (CEES), Department of Biology, University of Oslo, P.O. Box 1066 Blindern, NO-0316 Oslo, Norway
Full list of author information is available at the end of the article

result in immediate isolation from the progenitors, the result should be a major bottleneck. However, several studies have shown higher genetic diversity in polyploids compared to their progenitors [22-26], although this is not always the case [14,27-30]. The increased diversity may be the result of either recurrent formation of the polyploids [e.g., [1]], or past or ongoing interploidal gene flow through, for instance, triploid bridges [e.g., [19]]. These different models of polyploidisation can be seen as a gradient. Single event polyploidisation with subsequent reproductive isolation represents one end of this gradient whereas polyploidisation with ongoing gene flow or recurrent polyploidisation represents the other. Instances where polyploidisation is followed by historical gene flow which later stopped, or where polyploidisation is followed by reproductive isolation and subsequent gene flow in the form of secondary contact [31], could be considered as intermediate forms. Criteria and methods to distinguish between these different categories, though, have not yet been proposed [6].

Arabidopsis (DC.) Heynh. is a small genus consisting mostly of diploids, but includes both allopolyploids [14,32-35] and taxonomic autopolyploids [36-41]. As the genus includes the geneticists' pet plant *A. thaliana* L., plenty of molecular tools are available also for its relatives [e.g., the recent release of the *A. lyrata* genome, [42]], making the genus ideal for studying polyploid evolution. In Central Europe two species have been recorded with two ploidal levels each: *A. arenosa* (L.) Lawalrée (hereafter *arenosa*) represents a complex species aggregate [38] with diploid taxa occurring mainly in the Carpathians and possibly in a few regions further south in Hungary and Croatia, whereas tetraploid *arenosa* is found in most of Central Europe. In contrast *A. lyrata* (L.) O'Kane & Al-Shehbaz (hereafter *lyrata*) is mostly diploid throughout its European distribution range, but several tetraploid populations are found in the Eastern Austrian Forealps and neighbouring regions [39,43,44]. Schmickl and Koch [45] detected significant levels of introgression from tetraploid *arenosa* into the gene pool of diploid and tetraploid *lyrata*, resulting in introgressed tetraploid populations in the Eastern Austrian Forealps and the northerly adjacent Danube Valley, using microsatellite markers and morphometrics. These results, based on genetic and morphological admixture and differentiation, suggest past and ongoing gene flow between the two taxa. To distinguish between ancient and recent gene flow, especially in polyploid complexes, has long been controversial, and no clear criteria have yet been commonly acknowledged [6]. Here we use low-copy nuclear and plastid DNA sequences from the *Arabidopsis* model system and different analytical methods to study interspecific and interploidal gene flow in *lyrata* and *arenosa*, specifically asking, 1) whether

interploidal gene flow takes place in one or both directions, 2) how polyploidisation affects interspecific introgression and 3) whether it is possible to distinguish between recurrent formation and introgression from progenitors.

Methods

Material

Three to five Central European populations of each ploidal level ($2x$ and $4x$) from both *arenosa* and *lyrata* were included in this study (Table 1; Figure 1) with a total of 16 populations. *Lyrata* $2x$, *lyrata* $4x$ and *arenosa* $4x$ were mostly sampled in close proximity in the Eastern Austrian Forealps. *Arenosa* $2x$ was sampled in Slovakian Carpathians. The material was collected in the wild as living plants or seeds from defined single mother plants and grown in the Botanical Garden, University of Heidelberg, and the Phytotron, University of Oslo. Fresh leaves of 1-43 plants from most populations (274 individuals in total; Table 1) were later collected for flow cytometry analyses. Leaves from two specimens per population were dried using silica gel to preserve DNA before extraction.

Flow cytometry

Relative nuclear DNA content of 274 specimens from 14 populations (Table 1) were estimated by flow cytometry analyses performed by G. Geenen (Plant Cytometry Services, Schijndel), using DAPI staining, the *Arabidopsis* buffer described in Doležal and Suda [46] and *Ilex crenata* Thunb. 'Fastigiata' as internal standard, otherwise following the protocol described in Jørgensen et al. [30]. The populations a4_GER and a4_AUT3 were not included in the analyses as we did not have living material from these at the time of the analysis. However, microsatellite data for these populations indicate that they are tetraploid (Schmickl and Koch, in preparation). T-tests were done in SPSS 16.0 (SPSS Inc., Chicago) to test for differences in means of nuclear DNA content between the taxa.

DNA extraction, cloning, and sequencing

Whole genomic DNA was extracted from leaf tissue using the DNeasy Plant Mini protocol (Qiagen, Hilden). Polymerase chain reaction (PCR) of the low-copy nuclear regions chalcone synthase (CHS) and short chain alcohol dehydrogenase (*scADH*) was carried out in 25 μ l volumes with 1 \times DyNAzyme EXT buffer (Finnzymes, Espoo), 0.2 mM of each dNTP, 0.6 μ M of each primer (Additional file 1, Table S1), 0.2 U DyNAzyme EXT DNA polymerase (Finnzymes), and 2 μ l 10 times diluted DNA template. Thermocycling conditions consisted of 3 min at 94°C, and 35 cycles of 30 s at 93°C, 30 s at 55°C, 2.5-3 min at 70°C, and a final extension

Table 1 Sampling of *Arabidopsis arenosa* and *A. lyrata* included in this study.

Taxon	Ploidal level	Population ¹	Locality ²	# FC ³
<i>arenosa</i>	2x	a2_SVK1	SVK: Vysoké Tatry; Prešovský kraj; Belianske Tatry; Zadné Med'odoly Valley; Kopské Sedlo (131)	5
		a2_SVK2	SVK: Nízke Tatry Mts.; Pusté Pole (915140)	30
		a2_SVK3	SVK: Veľká Fatra Mts.; Harmanec; Malý Šturec Sedlo (915141)	33
	4x ⁴	a4_GER	GER: Southern Germany; Swabian Alps; Wental; Felsenmeer (123)	-
		a4_AUT1	AUT: Lower Austria; Eastern Alps; SSW St. Aegydt am Neuwalde; Kernhof; rocky batter next to street opposite depot (81-915142)	26
		a4_AUT2	AUT: Lower Austria; Waldviertel; Wachau; NNE Weißenkirchen; Achleiten (3)	12
		a4_AUT3	AUT: Lower Austria; Kamp Valley; S Stiefern; parking site with view on railway bridge (89)	-
	a4_AUT4	AUT: Lower Austria; Waldviertel; Wachau; forest road from Scheibenbach towards Pfaffental (20)	4	
<i>lyrata</i>	2x ⁵	I2_GER	GER: Bavaria; Veldenstein Forest; street from Velden to Pfaffenhofen (112)	17
		I2_CZE	CZE: SW Brno; NW Ivanice; between Nova Ves and Oslavany; slope above Oslava River (96)	9
		I2_AUT1	AUT: Lower Austria; street from Pernitz to Pottenstein (88-915143)	43
		I2_AUT2	AUT: Lower Austria; S Vienna; Bad Vöslau; rocks near Vöslauer Hütte (74-915145)	28
	4x	I4_AUT1	AUT: Lower Austria; Waldviertel; Wachau; E Dürnstein; small hill N Franzosendenkmal (13)	24
		I4_AUT2	AUT: Lower Austria; S Vienna; Mödling; Castle ruin Mödling (66-915144)	21
		I4_AUT3	AUT: Lower Austria; Dunkelstein Forest; Wachau; N Bacharnsdorf (50)	21
		I4_AUT4	AUT: Lower Austria; Schrambach between Freiland and Lilienfeld (116)	1

¹Two individuals from each population were sequenced.

²Country names are abbreviated as follows: AUT - Austria, CZE - Czech Republic, GER - Germany, and SVK - Slovakia. Brackets following localities give original collection number given in Schmickl and Koch [45].

³# FC gives the number of individuals analysed with flow cytometry. The populations a4_GER and a4_AUT3 were not included in the flow cytometry analysis, but multi-allelic microsatellite loci suggest they are tetraploid (Schmickl and Koch, unpublished).

⁴a4_AUT1 contained a single diploid individual, the others were tetraploid.

⁵I2_CZE contained a single triploid individual, the others were diploid.

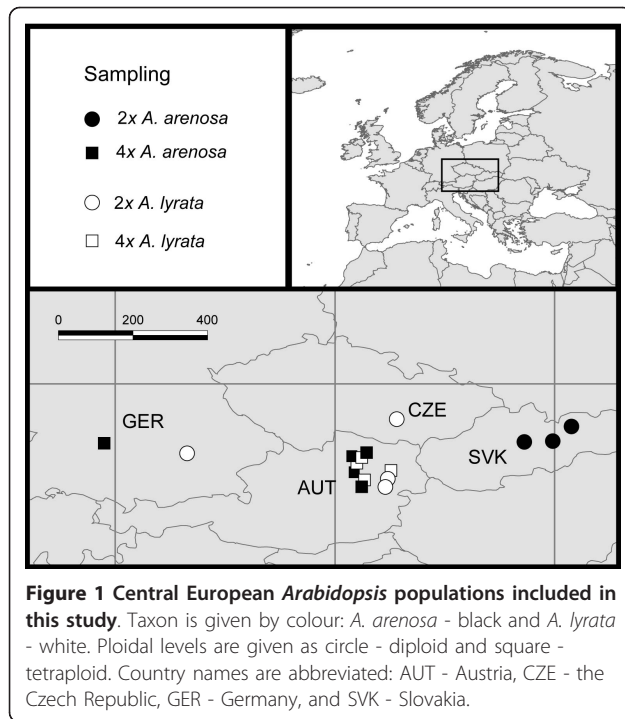
for 5 min at 70°C. PCR products were cloned using the TOPO-TA kit for sequencing with the pCR4-TOPO vector (Invitrogen, Carlsbad). Colonies were checked for inserts by running a PCR with M13 or T7 primers. At least six insert-containing clones from each PCR reaction were sequenced in both directions. The plastid region *trnL-F* was amplified using PuReTaq Ready-To-Go PCR beads (GE Healthcare, Waukesha) with 0.6 µM of each primer (Additional file 1, Table S1) and 2 µL 10 times diluted DNA template. For each DNA region, both strands were sequenced using BigDye v3.1 7 cycle sequencing kit (Applied Biosystems, Foster City) and M13F/M13R or T7 primers. Products of the cycle-sequencing reactions were separated on an ABI 3700 Genetic Analyzer (Applied Biosystems). The resulting sequences were assembled and edited using Vector NTI advance 10 (Invitrogen), and consensus sequences representing each allele and alignments were made using BioEdit version 7.0.5 [47]. For the plastid *trnL-F* region, only the first 700 base pairs were included in the analyses, to avoid unambiguous alignment due to the presence of pseudogenes [48,49].

PCR-mediated recombinants (chimeras) constitute a well-known problem in PCR-based cloning protocols [e.g. [50-55]], and to distinguish between PCR-mediated and real recombinants is not possible via PCR-based methods. However, the risk of obtaining

recombinants as PCR artefacts increases with the concentration of template [51], and the expected frequency of these should be lower than for real recombinants [50,52]. In this study we omitted clones that were recombinants of other cloned sequences from the same individual, and that were present at low frequencies, as PCR artefacts. Discrepant bases supported by only a single clone were assumed to be due to polymerase reading error and were corrected based on consensus sequences of other clones from the same individual.

Data analyses

Intragenic recombination events may be relatively common [56], and should be taken into consideration when choosing methods for phylogenetic analysis [57]. Minimum number of recombination events [RM; [58]] per region was calculated using DnaSP version 5.10 [59]. As we found substantial recombination for most regions (Additional file 1, Table S1), phylogenetic relationships were analysed for each region using neighbour networks [60], with Jukes-Cantor distances in the program SplitsTree4 [61]. Gaps were included following the simple coding strategy introduced by Simmons and Ochoterena [62] as implemented in the software SeqState version 1.4.1 [63]. For all marker systems the datasets were analysed by: 1) splitting the individuals into subsets



according to taxa, 2) including all individuals, and 3) including only diploids.

When two or three different alleles are found in a tetraploid individual, it is not possible to determine the true genotype although dosage may give an indication. Computational methods based among others on the EM algorithm have been developed to infer genotypes assuming random mating and populations at equilibrium [e.g. [64]]. As we have sampled two individuals from several distinct populations, we cannot assume equilibrium, and chose not to use any statistical method. To roughly assess the impact of assuming different numbers of allele copies, we reconstituted genotypes at random using the following approach: For each tetraploid individual with two or three alleles, a random number between 1 and 3 was generated. For individuals with two alleles, 1 corresponded to three copies of the first allele and one copy of the second (the order was arbitrary), 2 corresponded to two copies of each allele and 3 corresponded to one copy of the first allele and three copies of the second. For individuals with three alleles, 1 corresponded to duplicating the first allele, 2 to duplicating the second allele and 3 to duplicating the third allele. Three different datasets (D1, D2, D3) were generated using this approach. Assuming that the three allele proportions 1:3, 2:2, and 3:1 are equally probable for tetraploids with two distinct alleles at a locus leads to a deficit of 2:2 individuals compared to equilibrium expectations. Therefore we created a fourth dataset (D22) where all individuals with two distinct alleles were

considered to have a 2:2 genotype. The four datasets were used both in diversity calculations and isolation with migration (IM) analyses.

Diversity indices were calculated for each molecular region and each species/ploidal level, separately. For the tetraploids we calculated the indices for the duplicated datasets D1, D2, D3, and D22, and averaged the estimates. Arlequin version 3.11 [65] was used to calculate gene diversity (H_d), nucleotide diversity averaged over loci (π), and average number of nucleotide differences (k) with standard deviations. Differences in diversity between ploidy levels and between taxa were assessed by two sided T-tests computed using the standard deviations given by Arlequin.

The parameters of the IM model were estimated as implemented in the program IMA2 [66,67] to assess the importance and direction of gene flow between the ploidal levels and species. Samples of *arenosa* and *lyrata* were first analysed separately, to determine whether there is gene flow between ploidal levels and in which direction it occurs. Second, in order to assess gene flow between species, both taxa with both ploidal levels were analysed together in an analysis with four populations. We assumed one ancestral population for each species, and one ancestral population for the whole complex. In such an analysis with four populations a large number of parameters have to be estimated, requiring a large amount of data to obtain reliable estimates. As the three loci available here were somewhat limited in that respect, we also analysed only the two diploid taxa to assess evidence for interspecific gene flow. All analyses involving tetraploids were carried out for the four different datasets of tetraploid genotypes. One of the assumptions of the IM model is that there is no important gene exchange with populations not included in the analysis. As there was evidence for significant gene flow between ploidal levels in both species, we did not analyse the two tetraploids together. Another assumption of IMA2 is that there is no recombination. Tests for recombination [58,68] showed, however, that there was considerable recombination in the nuclear sequences used here. The program IMgc [69] was used to find the largest subsets of the data matrix without signs of recombination (non-recombining blocks) by removing either sequences or variable sites. The program can prioritise the number of sequences kept or the number of variable sites. We first used the default value 1 for the prioritising parameter. As some of the sequences got very short (datasets def; Additional file 2, Table S2), we produced additional subsets using a value of 0.5, retaining more variable sites and fewer sequences (datasets seq). This option was used for *arenosa* and *lyrata*, but it could not be used for the total and diploid datasets, because it reduced the number of sequences of diploid *arenosa* to three or less

(Table S2). Reducing the data to non-recombining blocks reduces the amount of data and leads to a loss of diversity, which may lead to a downward bias in estimates of effective population sizes obtained from IM. Divergence time and gene flow estimates have, however, been shown to be largely unaffected [70].

The parameter estimates provided by IMA2 are scaled by the mutation rate. In order to convert them to demographic estimates, a mutation rate needs to be assumed. We followed the procedure of Slotte et al. [21]: we assumed a substitution rate of $6.5 \cdot 10^{-9}$ [71,72] as a lower boundary for the mutation rate and the synonymous substitution rate of $1.5 \cdot 10^{-8}$ per site per generation [73] as an upper boundary, and used the mean of these two estimates to calculate per-fragment mutation rates (Additional file 3, Table S3).

IMA2 uses a Bayesian approach and Markov Chain Monte Carlo (MCMC) simulations to estimate parameters. Priors for effective population sizes ($q = 4N_e \cdot \mu$, where μ is the migration rate per fragment, not per bp), time since divergence ($t = \text{time} \cdot \mu$) and migration rates ($m = \text{migration rate} / \mu$) were initially chosen as recommended in the IM manual [74] and adjusted according to the results of preliminary analyses. For the final runs the following values were used (all values are scaled by $\mu = 1.21 \cdot 10^{-5}$): $q = 17$, $t = 10$ and $m = 5$ for *arenosa*, $q = 20$, $t = 7$ and $m = 5$ for *lyrata*, $q = 15$, $t = 10$ and $m = 5$ for the total dataset, and $q = 12$, $t = 10$ and $m = 2$ for the diploid dataset. The number of chains and the heating scheme were also tested in several preliminary runs. For the final runs we used 20 chains and heating parameters of $h_a = 0.96$ and $h_b = 0.9$ for analyses with two populations and 80 chains, and $h_a = 0.999$ and $h_b = 0.3$ for analyses with four populations. The length of the burnin was 1 000 000 MCMC iterations and estimates were based on between 10 and 27 million iterations. Mixing was assessed by trend plots for estimates over the runs and by effective sample size (ESS) values. Convergence was assessed by repeating runs several times with different random seeds. Some IMA2 runs were performed on BioHPC, Computational Biology Service Unit, Cornell University.

Results

DNA content

The ploidal levels for the 14 populations examined are given in Table 1. Only two populations showed signs of more than one ploidal level. The tetraploid *arenosa* population a4_AUT1 from Kernhof in Austria included one diploid individual, and the diploid *lyrata* population l2_CZE from NW Ivanice in the Czech Republic contained one triploid. The T-test showed that the two taxa had significantly different DNA content within ploidal levels for both diploids and tetraploids, *lyrata* having a

slightly larger nuclear DNA content than *arenosa* in both cases (0.23 vs. 0.20, and 0.44 vs. 0.43, $P < 0.001$).

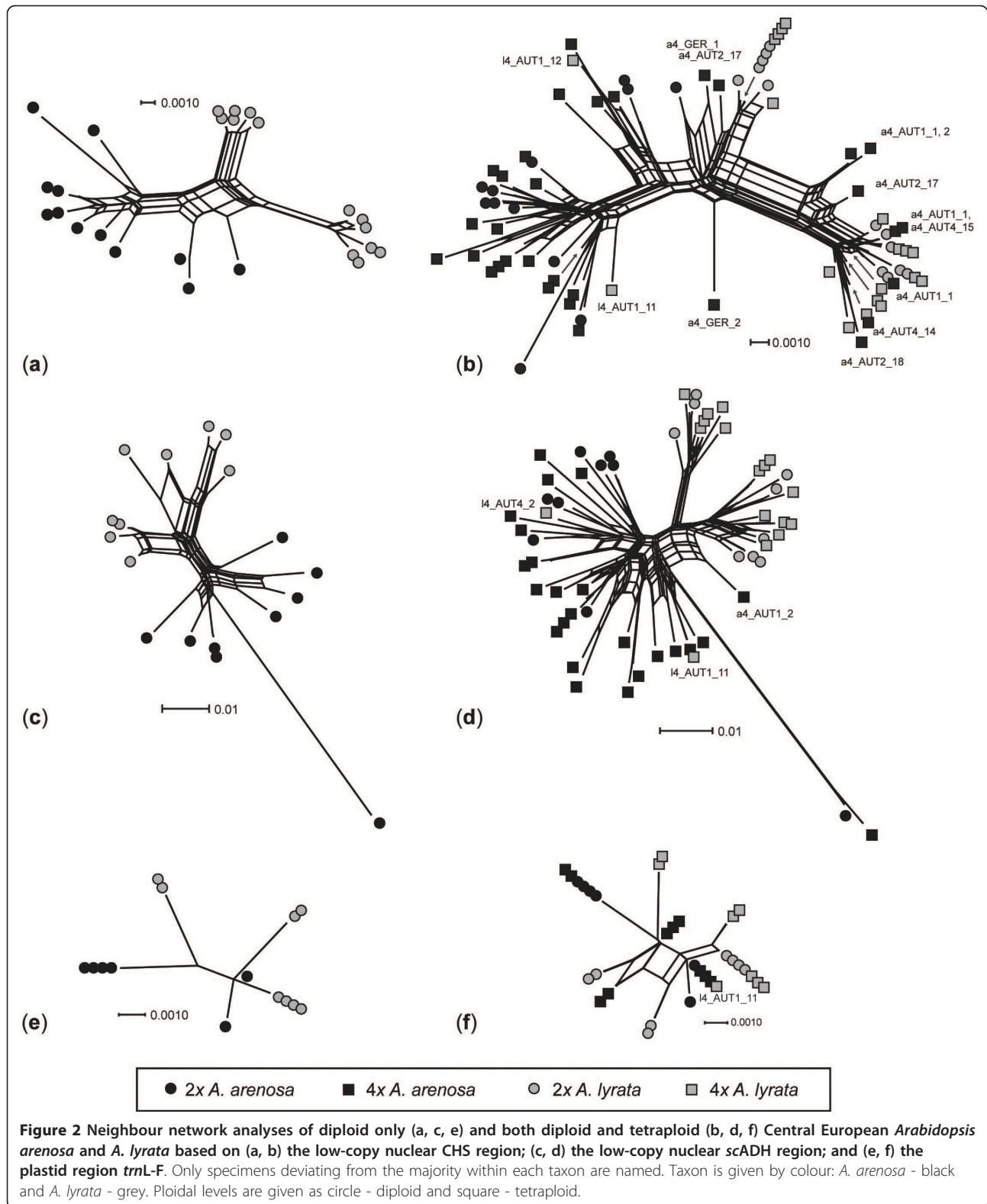
Sequence variation and diversity

The obtained sequences were deposited to GenBank with accession numbers GQ386471-GQ386654; 75 sequences of CHS, 59 sequences of *scADH*, and 32 sequences of *trnL-F* (Additional file 1, Table S1). Substantial recombination has taken place for both low-copy nuclear regions. Minimum number of recombination events was 16 for CHS and 24 for *scADH* (Additional file 1, Table S1). For the plastid *trnL-F* region, only a single recombination event was detected.

When analysed alone, the diploids were separated into two groups corresponding to named taxa in the neighbour networks based on the nuclear markers (Figure 2a, c), and partly also the plastid region (Figure 2e). CHS split the *lyrata* diploids into two distinct groups with absolutely no geographical structure; both clusters included specimens from Germany, the Czech Republic, and Austria (Figure 2a). There was no apparent geographical structure among *arenosa* specimens either. The analysis of *scADH* gave no additional information (Figure 2d). The two taxa didn't share *trnL-F* haplotypes, but all three *lyrata* haplotypes clustered closer to *arenosa* than to each other (Figure 2e).

Adding the tetraploids to the neighbour networks complicated the picture (Figure 2b, d, f). The majority of the tetraploids clustered according to taxa; tetraploid *lyrata* clustered with diploid *lyrata*, and tetraploid *arenosa* with diploid *arenosa*. There were, however, exceptions for all the marker systems. The CHS network grouped five tetraploid *lyrata* sequences with tetraploid *arenosa* (Figure 2b; Additional file 4, Table S4). These represent three specimens (with a mixture of *lyrata*- and *arenosa*-like alleles) of which two are from the same population (l4_AUT1), collected in Wachau, and the last one from Schrambach, also in Lower Austria (population l4_AUT4). In the *scADH* network, one of the specimens from the Wachau population (l4_AUT1_11) shared an allele with a tetraploid *arenosa* collected just a few kilometres away (a4_AUT2_18), whereas the specimen from the Schrambach population (l4_AUT4_2) clustered with a tetraploid *arenosa* from Wachau (a4_AUT4_15; Figure 2d; Additional file 4, Table S4).

In the CHS network ten tetraploid *arenosa* sequences clustered with the *lyrata* groups (Figure 2b). These represent seven specimens (with a mixture of *lyrata*- and *arenosa*-like alleles; Additional file 4, Table S4), most of them from Lower Austria (populations a4_AUT1, a4_AUT2, and a4_AUT4), but a single one from Germany (a4_GER). Only one of these tetraploid *arenosa* specimens (a4_AUT1_2) contained a *lyrata*-like



scADH allele and clustered with *lyrata* in the network (Figure 2d; Additional file 4, Table S4).

The plastid *trnL-F* network separated specimens according to taxa with one exception: the same tetraploid *lyrata* specimen from Wachau (l4_AUT1_11), which clustered with *arenosa* also in the CHS and *scADH* networks, shared a haplotype with diploid and tetraploid *arenosa* (a2_SVK2, a4_AUT2, and a4_GER, Figure 2f).

To summarise, these networks basically told the same story with major splits between *lyrata* and *arenosa*, and with ploidal levels to a high degree intermingled within each taxon. Deviations from this pattern were found more or less in the same populations for the different markers; tetraploid *arenosa*: a4_AUT1, a4_AUT2, a4_AUT4 and tetraploid *lyrata*: l4_AUT1 and l4_AUT4, all populations from Lower Austria where the two taxa are sympatric (Table 1; Additional file 4, Table S4).

Analysing the taxa separately for all marker systems showed that the specimens did not cluster according to ploidal level (Figure 3). For both species, the CHS and *scADH* networks separated groups of a few tetraploids from the remaining specimens, corresponding to the deviations mentioned above (Figure 3a, b). Otherwise specimens of different ploidal levels are completely intermingled.

Both nuclear regions showed high levels of gene diversity ($H_d = 0.93-0.99$; Figure 4, Additional file 5, Table S5). Differences in diversity among the duplicated tetraploid datasets (D1-22) were negligible, and there were no clear differences between taxa or ploidal levels. For *trnL-F*, the diversity was somewhat lower ($H_d = 0.60-0.86$), particularly for *arenosa*. For this marker the diversity was significantly higher in tetraploids than in diploids of both species ($p < 0.05$; Figure 4). There were no significant differences in nucleotide diversity (π) and average number of nucleotide differences (k) between ploidal levels or species for *trnL-F* and CHS. For *scADH*, however, both estimates of molecular diversity were significantly higher for *arenosa* than for *lyrata*, and this was the case for diploids and tetraploids (π : $p < 0.005$ for tetraploids and $p < 0.02$ for diploids, and k : $p < 0.001$ for tetraploids and $p < 0.005$ for diploids; Figure 4, Additional file 5, Table S5).

Isolation with migration results

For analyses of pairs of populations all runs reached ESS values > 1000 and mixing seemed good based on trend plots. Repeated runs indicated good convergence. The analyses with four populations did, however, not perform equally well and ESS values remained < 50 for several parameters. We will therefore in the first place base our conclusions on the pair wise runs, and only mention the results of the four population runs as indicative.

The main aim of this study was to assess evidence for gene flow between ploidal levels and species. For *arenosa*, the IM analysis revealed strong support for gene flow from diploids to tetraploids, but not in the other direction (Figure 5, Additional files 6, 7, Fig. S1, S2). The 95% highest posterior density intervals (HPD) for the migration rate from diploids to tetraploids excluded 0 for seven of eight datasets where it was estimated reliably (four variants of tetraploid genotypes \times two options of largest non-recombinant blocks; Additional file 2, Table S2), whereas the estimate of m was at the lowest value for migration from tetraploids to diploids in all cases. Estimates of the number of migrants from diploids to tetraploids were between 2.1 and 4.1, but HPD intervals were large and overlapped considerably among both migration directions (Additional file 2, Table S2). The posterior distribution for time since divergence did not go down towards 0 at the upper limit of the prior interval, independent of prior choice, and divergence time could thus not be properly estimated. Effective population size estimates were 1.5 to 2 times higher for tetraploids than for diploids (Additional file 6, Fig. S1), but as these are estimates of the effective number of genes, the estimated number of tetraploid individuals was in fact somewhat lower than for diploids.

For *lyrata*, there was also clear support for gene flow between the ploidal levels. HPD intervals for gene flow from diploids to tetraploids excluded 0 for five of eight datasets (Additional file 2, Table S2) and all estimates of m were larger than 0. For gene flow from tetraploids to diploids, HPD distributions did not reach low levels at high values for gene flow from tetraploids to diploids, making them unreliable. Still the HPD intervals excluded 0 for three out of four def datasets, and estimates of m were at the lowest point of the distribution only for one def dataset and two seq datasets, indicating gene flow from tetraploids to diploids. The discrepancies between the different datasets resulted from the fact that different parts of the sequences were kept by IMgc (Additional file 2, Table S2). The effective number of migrants per generation was estimated as 1.5 to 1.8 from diploids to tetraploids and as 1.2 to 1.4 from tetraploids to diploids. As for *arenosa*, the posterior distributions for time since divergence were not unimodal and did not provide any reliable estimates. Estimates of effective population sizes varied also somewhat between datasets and were in general not different between ploidal levels.

The results of the analysis including all four taxonomic/cytogenetic entities (Additional file 2, Table S2) were largely consistent with the results of the pair wise runs, although these estimates have to be considered unreliable due to poor performance of the

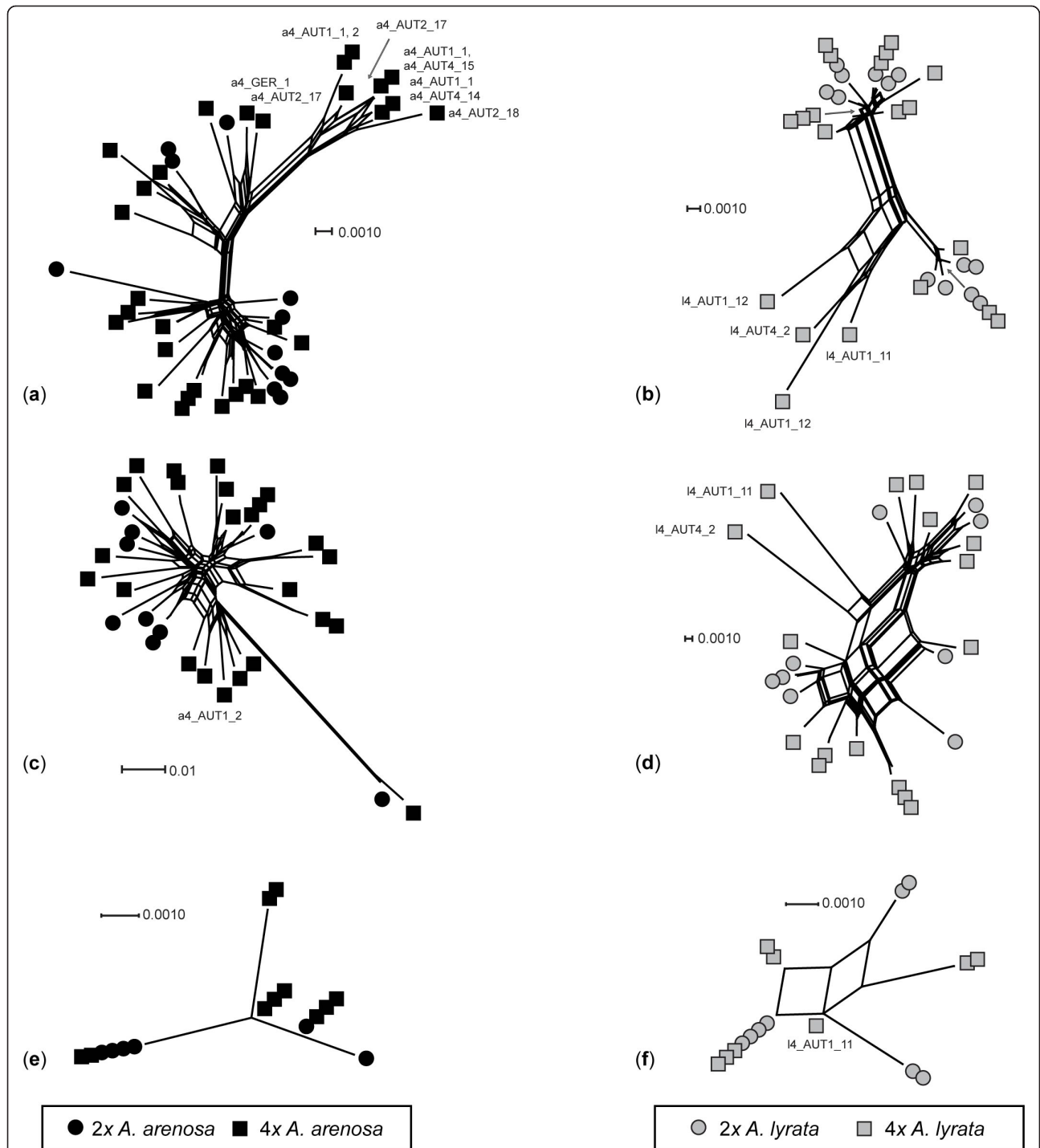
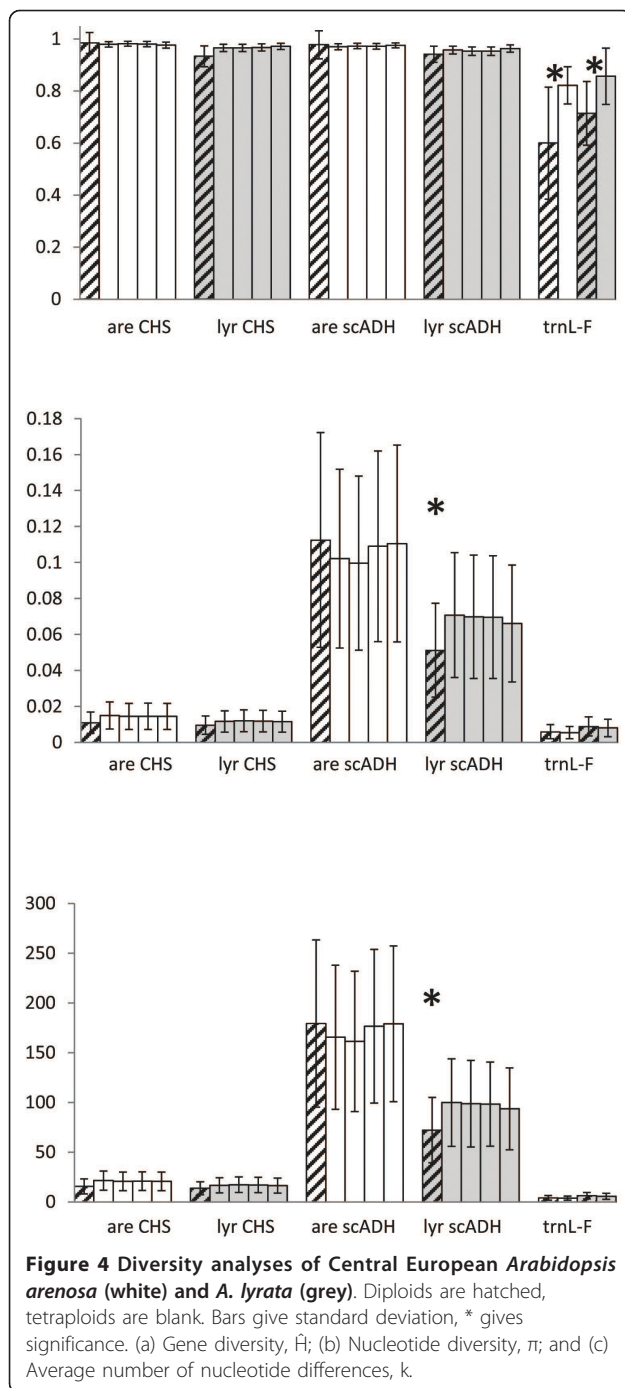


Figure 3 Neighbour network analyses of diploid and tetraploid Central European *Arabidopsis arenosa* (a, c, e) and *A. lyrata* (b, d, f) species. (a, b) are based on the low-copy nuclear CHS region; (c, d) are based on the low-copy nuclear scADH region; and (e, f) are based on the plastid region trnL-F. Only specimens identified as deviating from the majority within each taxon in Figure 2 are named. Taxon is given by colour: *A. arenosa* - black and *A. lyrata* - grey. Ploidal levels are given as circle - diploid and square - tetraploid.



MCMC (and are therefore not shown). In addition, these runs suggested gene flow between species, from diploid and tetraploid *arenosa* into tetraploid *lyrata*, and for one dataset also from diploid *lyrata* into tetraploid *arenosa*.

Analysing the two diploid species together (Additional file 2, Table S2) showed that there was no evidence what so ever for gene flow between them.

Discussion

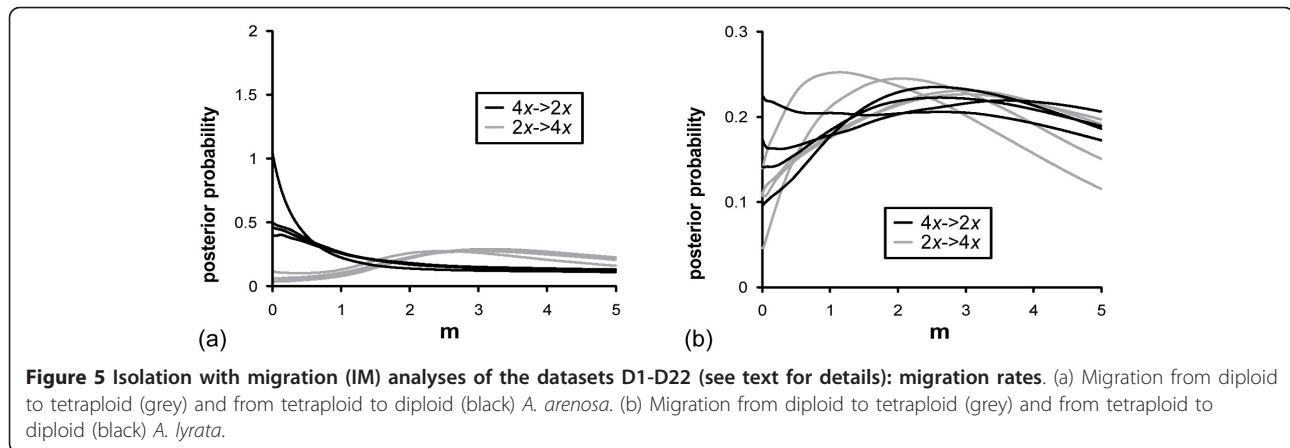
The overall picture indicated by our results is the following: Both the neighbour nets and the IM analysis show that there is no gene flow between the two diploid taxa, a result which is in agreement with expectations based on their presently well separated distribution areas [38]. There is, however, evidence for gene flow from diploids to tetraploids in *arenosa* and possibly for interploidal gene flow in both directions in *lyrata*. This is consistent with the intermingling of sequences of both ploidal levels revealed by the networks. The possibility of gene flow from tetraploids to diploids in *lyrata* is further supported by the triploid individual found in population l2_CZE. However, spontaneous occurrence of triploids in diploid populations is a common phenomenon and might be alternatively explained by low-frequency production of unreduced gametes. The networks including both species and ploidal levels clearly indicate mixing of lineages between species. As there seems to be no gene flow between diploids, we assume that this mixing results from gene flow into tetraploids. Consistent with this assumption, the IMA2 analysis of all four taxonomic/cytogenetic entities suggests gene flow from diploid and tetraploid *arenosa* into tetraploid *lyrata*, and possibly, with much lower frequency, also from diploid *lyrata* into tetraploid *arenosa*.

Interspecific gene flow

It has been suggested that genotypes of tetraploids are buffered against the shock of absorbing foreign genomes, and that extensive introgression often takes place at the tetraploid level between species that are isolated from each other at the diploid level [75,76]. Our analyses of *Arabidopsis* in Central Europe show that *arenosa* and *lyrata* are good biological species at the diploid level. The network analyses show no sharing of alleles, and the main splits are between the two taxa, which is in agreement with a comprehensive large-scale analysis of the genus [77]. Furthermore, isolation with migration (IM) analyses of diploids show no gene flow from $2x$ *arenosa* to *lyrata* or the other way around.

We do, however, find signs of introgression in the tetraploids on both sides; several tetraploid *lyrata* sequences cluster with *arenosa* in the network analyses (especially for the CHS region, but also for the other two regions), and several tetraploid *arenosa* sequences cluster with *lyrata*.

The pattern of gene flow that we observed could alternatively be explained by the fact that diploids of the two species are spatially separated, whereas the tetraploids largely overlap in distribution range. Because we lack proper population samples in the present study, we could not formally test to what extent a correlation



between genetic and geographic distance could explain our results (using for instance a Mantel test). However, the pattern resulting from the IM analyses cannot fully be explained by geography. IM indeed indicated unidirectional gene flow into tetraploids in *arenosa*, despite ploidal levels being at present allopatric. Gene flow from diploid *arenosa* into tetraploid *lyrata* was also suggested by some of the IM results. Therefore our results are very likely to reflect more than simple isolation by distance.

To our knowledge there are not many studies that have dealt with gene flow between sister species that contain two (or more) ploidal levels. Luttikhuizen *et al.* [23] found higher genetic diversity in autotetraploid *Rorippa amphibia* (L.) Besser compared to conspecific diploids using microsatellites, and suggested that introgression as well as multiple origins of the tetraploids might have contributed to the tetraploid diversity. Stift *et al.* [78] used crossing experiments to show that there are limited reproductive barriers between *R. amphibia* and the sympatric tetraploid *R. sylvestris* (L.) Besser, and concluded that gene flow between the two tetraploids is the probable reason for the high diversity found in tetraploid *R. amphibia* by Luttikhuizen *et al.* [23]. Kloda *et al.* [79] studied gene flow among diploid *Ononis* (*O. spinosa* L. and *O. intermedia* C.A.Mey. ex. A.K.Becker) and tetraploids (*O. repens* L. and *O. maritima* Dumort.) using microsatellites, and found that there were restrictions to gene flow between, but not within the ploidal levels. In the genus *Paeonia* L. homoploid hybrid species have been derived from allotetraploids, but not from the diploid progenitors, suggesting that chromosomal structural differences induced by polyploidy might create new opportunities for interspecific gene flow [80]. In line with this previous evidence, our study suggests that polyploids might tolerate introgression better than their diploid progenitors, as suggested by de Wet & Harlan [75] and Harlan & de Wet [76].

Interploidal gene flow

In general, high genetic diversity in polyploids, as our network and genetic diversity analyses indicated for both *arenosa* and *lyrata*, can be explained by multiple independent origins of the polyploids, continuing introgression from the diploids, introgression from other polyploid species (see above), or as a result of subsequent evolution following one or more old polyploidisation events [see e.g., [3,6]]. Our network analyses for each species show no apparent clustering according to ploidal levels; i.e. we do not see a separation of diploids and tetraploids. For *arenosa* our IM analyses suggest that the diploids and the tetraploids are two distinct groups, with some migration from the diploid to the tetraploid level subsequent to the separation of the two gene pools. As the diploids from the Carpathians and the tetraploids from the Eastern Austrian Forealps and Germany are no longer sympatric, the migration we observe is probably the result of recurrent origin/introgression in the past. For *lyrata* our IM analysis suggests that the gene flow is bidirectional. This is also consistent with earlier studies [77] considering a broad geographic and population sampling that demonstrate a deeper evolutionary split between *arenosa* and *lyrata* irrespective of ploidal level variation.

In our flow cytometry data, we find signs of mixed-ploidy populations: a triploid plant was detected in the mainly diploid l2_CZE population of *lyrata* from the Czech Republic, and a diploid plant was detected in the mainly tetraploid a4_AUT1 population of *arenosa* from the Eastern Austrian Foreland. Also, the Central European *Arabidopsis* populations often have restricted distributions, and populations of different ploidal levels are sometimes only a few kilometres apart. Thus, introgression is possible, especially along disturbed sites, e.g., along roads. However, the generally low migration rates suggest that introgression remains a rare event. If this is the case, the small genus *Arabidopsis* contains polyploids

with very different evolutionary fates when it comes to introgression. The allotetraploid *A. suecica* (Fr.) Norrl. ex O.E.Schulz has very low genetic diversity and has probably originated only once [34,81]. The allotetraploid *A. kamchatica* (Fisch. ex DC.) K.Shimizu & Kudoh has probably originated multiple times, and additionally experienced some later introgression from the parental diploids [14,35,82,83]. The tetraploid *arenosa* shows some signs of introgression, and the tetraploid *lyrata* shows signs of extensive ongoing introgression [cf. [45]]. As such, the genus *Arabidopsis* might be a good model system to use for developing the criteria and methods called for by Soltis et al. [6] for distinguishing between recurrent formation and introgression from progenitors as sources of diversity in polyploids. In this study we used IM analysis [67] in an attempt to make this distinction. As we were not able to obtain credible estimates of time divergence in our analyses, we could however not answer this question properly. Nevertheless, the results provide indications that different populations of the two species, *arenosa* and *lyrata*, are placed in different positions along the gradient of possible polyploid evolution models ranging from single event polyploidisation with subsequent reproductive isolation at one end of the gradient to polyploidisation with ongoing gene flow or recurrent polyploidisation at the other end.

Gene flow from diploids to tetraploid derivatives has long been acknowledged as relatively common [see e.g., [2,3,6] and references therein, [12]]. The question of gene flow in the opposite direction is more controversial. Stebbins [84] states that interploidal gene flow is usually unidirectional from diploids to tetraploids for two reasons: 1) offspring of triploid hybrids are usually tetraploid or close to it in chromosome number, and 2) diploids and tetraploids are often so highly incompatible that triploid offspring cannot be formed at all. However, studies involving natural triploids in euploid hybrid swarms show that triploids may produce 1x, 2x, and 3x gametes, and may therefore contribute to gene flow in both directions [19,20,85]. Indeed, gene flow from tetraploids to diploids has been observed in some taxa, e.g. *Dactylorhiza maculata* (L.) Soó [86] and *Betula* L [87]. In our study, we did not observe gene flow from tetraploid to diploid *arenosa* ($m = 0$; Figure 5), but as the different ploidal levels are allopatric, we cannot distinguish between genetic/genomic and geographic barriers. However, for *lyrata*, with diploids and polyploids in close proximity, gene flow seems to be bidirectional.

Conclusions

In this study, we looked at the effect of polyploidisation on interspecific introgression, and interploidal gene flow using Central European *Arabidopsis* as a model system. There was no evidence for interspecific gene flow

between 2x *arenosa* and 2x *lyrata*, which can be considered as good biological entities, but some support for gene flow into 4x *lyrata* and possibly 4x *arenosa*. Thus, whole genome duplication might decrease vulnerability to interspecific hybridisation and buffer negative effects of introgression. Interploidal gene flow was detected from 2x to 4x in both species, and from 4x to 2x in *lyrata*. For *arenosa*, the two ploidal levels are allopatric, and the lower level of gene flow could be the result of geographic as well as genetic barriers. In *lyrata*, however, where geographic barriers are limited, gene flow is bidirectional.

Additional material

Additional file 1: Table S1. Dataset summary with primers, GenBank numbers, alignment lengths and minimum number of recombination events per region.

Additional file 2: Table S2. Isolation with migration (IM) analyses: description of datasets.

Additional file 3: Table S3. Isolation with migration analyses: mutation rates per fragment per generation.

Additional file 4: Table S4. Signs of introgression among tetraploid *Arabidopsis arenosa* and *A. lyrata*.

Additional file 5: Table S5. Diversity indices.

Additional file 6: Figure S1. Isolation with migration analyses: effective population sizes and estimates of time for the def datasets.

Additional file 7: Figure S2. Isolation with migration analyses of the seq datasets.

Acknowledgements

The authors would like to thank Cecilie Mathiesen for assistance in the lab, and Barbara Mable, Ovidiu Paun, and the botanical group in Oslo for reading and commenting on the manuscript.

Author details

¹Centre for Ecological and Evolutionary Synthesis (CEES), Department of Biology, University of Oslo, P.O. Box 1066 Blindern, NO-0316 Oslo, Norway.

²Institute for Arctic and Marine Biology, University of Tromsø, NO-9037

Tromsø, Norway. ³Centre for Organismal Studies (COS) Heidelberg, Department of Biodiversity and Plant Systematics, University of Heidelberg, Im Neuenheimer Feld 345, D-69120 Heidelberg, Germany.

Authors' contributions

MHJ carried out the molecular work and statistical analyses, and drafted the manuscript. DE carried out the isolation with migration analyses and helped to draft the manuscript. RS and MAK contributed to the sampling design, provided the samples and contributed to draft the manuscript. AKB was project leader, contributed in the molecular work, and drafted the manuscript. All authors read and approved the final manuscript.

Received: 6 July 2011 Accepted: 29 November 2011

Published: 29 November 2011

References

1. Soltis DE, Soltis PS: Molecular data and the dynamic nature of polyploidy. *Critical Reviews in Plant Sciences* 1993, **12**(3):243-273.
2. Soltis DE, Soltis PS: Polyploidy: recurrent formation and genome evolution. *Trends in Ecology & Evolution* 1999, **14**(9):348-352.
3. Soltis DE, Soltis PS, Tate JA: Advances in the study of polyploidy since *Plant speciation*. *New Phytologist* 2003, **161**:173-191.

4. Fawcett JA, Maere S, Van de Peer Y: **Plants with double genomes might have had a better chance to survive the Cretaceous-Tertiary extinction event.** *Proceedings of the National Academy of Sciences of the United States of America* 2009, **106**(5737-5742).
5. Wood TE, Takebayashi N, Barker MS, Mayrose I, Greenspoon PB, Rieseberg LH: **The frequency of polyploid speciation in vascular plants.** *Proceedings of the National Academy of Sciences of the United States of America* 2009, **106**(33):13875-13879.
6. Soltis DE, Buggs RJA, Doyle JJ, Soltis PS: **What we still don't know about polyploidy.** *Taxon* 2010, **59**(5):1387-1403.
7. Stift M, Berenos C, Kuperus P, Tienderen PHv: **Segregation models for disomic, tetrasomic and intermediate inheritance in tetraploids: a general procedure applied to *Rorippa* (yellow cress) microsatellite data.** *Genetics* 2008, **179**:2113-2123.
8. Soltis DE, Soltis PS, Schemske DW, Hancock JF, Thompson JN, Husband BC, Judd WS: **Autopolyploidy in angiosperms: have we grossly underestimated the number of species?** *Taxon* 2007, **56**(1):13-30.
9. O'Kane SL, Al-Shehbaz IA: **A synopsis of *Arabidopsis* (Brassicaceae).** *Novon* 1997, **7**:323-327.
10. Doyle JJ, Doyle JL, Rauscher JT, Brown AHD: **Evolution of the perennial soybean polyploid complex (*Glycine* subgenus *Glycine*): a study of contrasts.** *Biological Journal of the Linnean Society* 2004, **82**:583-597.
11. Hendry AP: **Evolutionary biology: speciation.** *Nature* 2009, **458**:162-164.
12. Chapman MA, Abbott RJ: **Introgression of fitness genes across a ploidy barrier.** *New Phytologist* 2010, **186**:63-71.
13. Kim S-T, Sultan SE, Donoghue MJ: **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* 2008, **105**(34):12370-12375.
14. Shimizu-Inatsugi R, Lihová J, Iwanaga H, Kudoh H, Marhold K, Savolainen O, Watanabe K, Yakubov VV, Shimizu KK: **The allopolyploid *Arabidopsis kamchatica* originated from multiple individuals of *Arabidopsis lyrata* and *Arabidopsis halleri*.** *Molecular Ecology* 2009, **18**(19):4024-4048.
15. Symonds VV, Soltis PS, Soltis DE: **Dynamics of polyploid formation in *Tragopogon* (Asteraceae): recurrent formation, gene flow, and population structure.** *Evolution* 2010, **64**(7):1984-2003.
16. Soltis DE, Soltis PS: **Genetic consequences of autopolyploidy in *Tolmiea* (Saxifragaceae).** *Evolution* 1989, **43**:586-594.
17. Segraves KA, Thompson JN, Soltis PS, Soltis DE: **Multiple origins of polyploidy and the geographic structure of *Heuchera grossularifolia*.** *Molecular Ecology* 1999, **8**:253-262.
18. Parisod C, Bernard G: **Glacial *in situ* survival in the Western Alps and polytopic autopolyploidy in *Biscutella laevigata* (Brassicaceae).** *Molecular Ecology* 2007, **16**:2755-2767.
19. Husband BC: **The role of triploid hybrids in the evolutionary dynamics of mixed-ploidy populations.** *Biological Journal of the Linnean Society* 2004, **82**:537-546.
20. Husband BC, Sabara HA: **Reproductive isolation between autotetraploids and their diploid progenitors in fireweed, *Chamerion angustifolium* (Onagraceae).** *New Phytologist* 2003, **161**:703-713.
21. Slotte T, Huang H, Lascoux M, Ceplitis A: **Polyploid speciation did not confer instant reproductive isolation in *Capsella* (Brassicaceae).** *Molecular Biology and Evolution* 2008, **25**(7):1472-1481.
22. Abbott RJ, Ireland HE, Rogers HJ: **Population decline despite high genetic diversity in the new allopolyploid species *Senecio cambrensis* (Asteraceae).** *Molecular Ecology* 2007, **16**:1023-1033.
23. Luttikhuisen PC, Stift M, Kuperus P, Tienderen PHv: **Genetic diversity in diploid vs. tetraploid *Rorippa amphibia* (Brassicaceae).** *Molecular Ecology* 2007, **16**:3544-3553.
24. Facey PD, Lee PLM, Smith MNE, Hipkin CR: **Conservation of genetic diversity in British populations of the diploid endemic *Coincya monensis* ssp. *monensis* (Isle of Man Cabbage): the risk of hybridisation with the tetraploid alien, *Coincya monensis* ssp. *cheiranthos*.** *Conservation Genetics* 2007, **8**:1029-1042.
25. Koch MA: **Genetic differentiation and speciation in prealpine *Cochlearia* (Brassicaceae): Allohexaploid *Cochlearia bavarica* (Brassicaceae) compared to its diploid ancestor *Cochlearia pyrenaica* in Germany and Austria.** *Plant Systematics and Evolution* 2002, **232**(1-2):35-49.
26. Koch MA, Bernhardt K-G: **Comparative biogeography of the cytotypes of annual *Microthlaspi perfoliatum* (Brassicaceae) in Europe using isozymes and cpDNA data: refugia, diversity centers, and postglacial colonization.** *American Journal of Botany* 2004, **91**(1):115-124.
27. Hedrén M, Nordström S, Ståhlberg D: **Polyploid evolution and plastid DNA variation in the *Dactylophiza incamata/maculata* complex (Orchidaceae) in Scandinavia.** *Molecular Ecology* 2008, **17**:5075-5091.
28. Paun O, Greilhuber J, Temsch EM, Hörandl E: **Patterns, sources and ecological implications of clonal diversity in apomictic *Ranunculus carpaticola* (*Ranunculus auricomus* complex, Ranunculaceae).** *Molecular Ecology* 2006, **15**:897-910.
29. Paun O, Stuessy TF, Hörandl E: **The role of hybridization, polyploidization and glaciation in the origin and evolution of the apomictic *Ranunculus cassubicus* complex.** *New Phytologist* 2006, **171**:223-236.
30. Jørgensen MH, Elven R, Tribsch A, Gabrielsen TM, Stedje B, Brochmann C: **Taxonomy and evolutionary relationships in the *Saxifraga rivularis* complex.** *Systematic Botany* 2006, **31**(4):702-729.
31. Stebbins GL: **Polyploidy and the distribution of the arctic-alpine flora: new evidence and a new approach.** *Botanica Helvetica* 1984, **94**(1):1-13.
32. Shimizu KK, Fujii S, Marhold K, Watanabe K, Kudoh H: ***Arabidopsis kamchatica* (Fisch. ex DC.) K.Shimizu & Kudoh and *A. kamchatica* subsp. *kawasakiana* (Makino) K.Shimizu & Kudoh, new combinations.** *Acta Phytotaxonomica et Geobotanica* 2005, **56**(2):163-172.
33. O'Kane SL, Schaal BA, Al-Shehbaz IA: **The origins of *Arabidopsis suecica* (Brassicaceae) as indicated by nuclear rDNA sequences.** *Systematic Botany* 1996, **21**(4):559-566.
34. Jakobsson M, Hagenblad J, Tavaré S, Säll T, Halldén C, Lind-Halldén C, Nordborg M: **A unique recent origin of the allotetraploid species *Arabidopsis suecica*: evidence from nuclear DNA markers.** *Molecular Biology and Evolution* 2006, **23**(6):1217-1231.
35. Schmickl R, Jørgensen MH, Brysting AK, Koch MA: **The evolutionary history of the *Arabidopsis lyrata* complex: A hybrid in the amph-Beringian area closes a large distribution gap and builds up a genetic barrier.** *BMC Evolutionary Biology* 2010, **10**:98.
36. Berkutenko AN, Tzytlenok SI, Pulkina SV: **Chromosome numbers and dispersal of the Brassicaceae family in the Magadan district.** *Botanicheskii Zhurnal* 1984, **69**(1):75-80.
37. Elven R, Murray DF, Razzhivin VY, Yurtsev BA: **Checklist of the Panarctic Flora (PAF) vascular plants.** Oslo: National Centre for Biosystematics, University of Oslo; 2008.
38. Koch MA, Wernisch M, Schmickl R: ***Arabidopsis thaliana*'s wild relatives: an updated overview on systematics, taxonomy and evolution.** *Taxon* 2008, **57**(3):933-943.
39. Polatschek A: **Cytotaxonomische Beiträge zur Flora der Ostalpenländer, I.** *Österreichische botanische Zeitschrift* 1966, **113**:1-46.
40. Zhukova PG, Petrovsky VV, Plieva TV: **The chromosome numbers and taxonomy of some plant species from Siberia and Far East.** *Botanicheskii Zhurnal* 1973, **58**:1331-1342.
41. Zhukova PG, Petrovsky VV: **Chromosome numbers and taxonomy of some plant species from the northern Asia regions.** *Botanicheskii Zhurnal* 1987, **72**(12):1617-1624.
42. Hu TT, Pattyn P, Bakker EG, Cao J, Cheng J-F, Clark RM, Fahlgren N, Fawcett JA, Grimwood J, Gundlach H, et al: **The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change.** *Nature Genet* 2011, **43**(5):476-481.
43. Dobeš C, Vitek E: **Documented chromosome number checklist of Austrian vascular plants.** Wien: Verlag des Naturhistorischen Museums Wien; 2000.
44. Schmickl R: **Reticulate evolution in glacial refuge areas - the genus *Arabidopsis* in the eastern Austrian Danube Valley (Wachau).** *PhD thesis* Heidelberg: Ruperto-Carola University of Heidelberg; 2009.
45. Schmickl R, Koch MA: ***Arabidopsis* hybrid speciation processes.** *Proceedings of the National Academy of Sciences of the United States of America* 2011, **108**(34):14192-14197.
46. Doležel J, Suda J, eds: **Flow Cytometry with Plant Cells.** Weinheim: Wiley VCH Verlag GmbH & Co. KGaA; 2007.
47. Hall TA: **BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT.** *Nucleic Acids Symposium Series* 1999, **41**:95-98.
48. Schmickl R, Kiefer C, Dobeš C, Koch MA: **Evolution of *trnF*(GAA) pseudogenes in cruciferous plants.** *Plant Systematics and Evolution* 2008, **282**(3-4):229-240.

49. Koch MA, Dobeš C, Matschinger M, Bleeker W, Vogel J, Kiefer M, Mitchell-Olds T: 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* 2005, **22**(4):1032-1043.
50. Bradley RD, Hillis DM: Recombinant DNA sequences generated by PCR amplification. *Molecular Biology and Evolution* 1997, **14**(5):592-593.
51. Judo MSB, Wedel AB, Wilson C: Stimulation and suppression of PCR-mediated recombination. *Nucleic Acids Research* 1998, **26**(7):1819-1825.
52. Popp M, Oxelman B: Inferring the history of the polyploid *Silene aegaea* (Caryophyllaceae) using plastid and homoeologous nuclear DNA sequences. *Molecular Phylogenetics and Evolution* 2001, **20**(3):474-481.
53. Cronn R, Cedroni M, Haselkorn T, Grover C, Wendel JF: PCR-mediated recombination in amplification products derived from polyploid cotton. *Theoretical and Applied Genetics* 2002, **104**:482-489.
54. Brysting AK, Mathiesen C, Marcussen T: Challenges in polyploid phylogenetic reconstruction: a case story from the arctic-alpine *Cerastium alpinum* complex. *Taxon* 2011, **60**(2):333-347.
55. Qiu X, Wu L, Huang H, McDonel PE, Palumbo AV, Tiedje JM, Zhou J: Evaluation of PCR-generated chimeras, mutations and heteroduplexes with 16S rRNA gene-based cloning. *Applied and Environmental Microbiology* 2001, **67**(2):880-887.
56. Kelly LJ, Leitch AR, Clarkson JJ, Hunter RB, Knapp S, Chase MW: Intra-genic recombination events and evidence for hybrid speciation in *Nicotiana* (Solanaceae). *Molecular Biology and Evolution* 2010, **27**(4):781-799.
57. Schierup MH, Hein J: Consequences of recombination on traditional phylogenetic analysis. *Genetics* 2000, **156**:879-891.
58. Hudson RR, Kaplan NL: Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* 1985, **111**:147-164.
59. Librado P, Rozas J: DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 2009, **25**:1451-1452.
60. Bryant D, Moulton V: Neighbor-net: an agglomerative method for the construction of phylogenetic networks. *Molecular Biology and Evolution* 2004, **21**(2):255-265.
61. Huson DH, Bryant D: Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* 2006, **23**(2):254-267.
62. Simmons MP, Ochoterena H: Gaps as characters in sequence-based phylogenetic analyses. *Systematic Biology* 2000, **49**(2):369-381.
63. Müller K: SeqState: Primer design and sequence statistics for phylogenetic DNA datasets. *Applied Bioinformatics* 2005, **4**(1):65-69.
64. De Silva HN, Hall AJ, Rikkerink E, McNeilage MA, Fraser LG: Estimation of allele frequencies in polyploids under certain patterns of inheritance. *Heredity* 2005, **95**:327-334.
65. Excoffier L, Laval G, Schneider S: Arlequin ver 3.11 An Integrated Software Package for Population Genetics. Bern: Computational and Molecular Population Genetics Lab (CMPG), Institute of Zoology, University of Berne; 2007.
66. Hey J, Nielsen R: Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proceedings of the National Academy of Sciences of the United States of America* 2007, **104**(8):2785-2790.
67. Hey J: Isolation with migration models for more than two populations. *Molecular Biology and Evolution* 2010, **27**(4):905-920.
68. Hudson RR: Estimating the recombination parameter of a finite population model without selection. *Genetical Research* 1987, **50**:245-250.
69. Woerner AE, Cox MP, Hammer MF: Recombination-filtered genomic datasets by information maximization. *Bioinformatics* 2007, **23**:1851-1853.
70. Strasburg JL, Rieseberg LH: How robust are "isolation with migration" analyses to violations of the IM model? A simulation study. *Molecular Biology and Evolution* 2010, **27**(2):297-310.
71. Wolfe KH, Li W-H, Sharp PM: Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences of the United States of America* 1987, **84**:9054-9058.
72. Wright SI, Lauga B, Charlesworth D: Subdivision and haplotype structure in natural populations of *Arabidopsis lyrata*. *Molecular Ecology* 2003, **12**:1247-1263.
73. Koch MA, Haubold B, Mitchell-Olds T: Molecular systematics of the Brassicaceae: evidence from coding plastidic *MatK* and nuclear *Chs* sequences. *American Journal of Botany* 2001, **88**(2):534-544.
74. Documentation for IMA2. [http://genfaculty.rutgers.edu/hey/software].
75. De Wet JMJ, Harlan JR: Apomixis, polyploidy, and speciation in *Dichanthium*. *Evolution* 1970, **24**(270-277).
76. Harlan JR, De Wet JMJ: On Ö. Winge and a prayer: the origins of polyploidy. *The Botanical Review* 1975, **41**(4):361-390.
77. Koch MA, Matschinger M: Evolution and genetic differentiation among relatives of *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America* 2007, **104**(15):6272-6277.
78. Stift M, Bregman R, Oostermeijer JGB, van Tienderen PH: Other tetraploid species and conspecific diploids as sources of genetic variation for an autotetraploid. *American Journal of Botany* 2010, **97**(11):1858-1866.
79. Kloda JM, Dean PDG, Maddren C, MacDonald DW, Mayes S: Using principal component analysis to compare genetic diversity across polyploidy levels within plant complexes: an example from British Restharrow (*Ononis spinosa* and *Ononis repens*). *Heredity* 2008, **100**:253-260.
80. Ferguson D, Sang T: Speciation through homoploid hybridization between allotetraploids in peonies (*Paeonia*). *Proceedings of the National Academy of Sciences of the United States of America* 2001, **98**(7):3915-3919.
81. Lind-Hallén C, Hallén C, Säll T: Genetic variation in *Arabidopsis suecica* and its parental species *A. arenosa* and *A. thaliana*. *Heredity* 2002, **136**:45-50.
82. Schmickl R, Jørgensen MH, Brysting AK, Koch MA: Phylogeographic implications for the North American boreal-arctic *Arabidopsis lyrata* complex. *Plant Ecology & Diversity* 2008, **1**(2):245-254.
83. Wang W-K, Ho C-W, Hung K-H, Wang K-H, Huang C-C, Araki H, Hwang C-C, Hsu T-W, Osada N, Chiang T-Y: Multilocus analysis of genetic divergence between outcrossing *Arabidopsis* species: evidence of genome-wide admixture. *New Phytologist* 2010, **188**(2):488-500.
84. Stebbins GL: Chromosomal evolution in higher plants. London: Edward Arnold (Publishers) Ltd; 1971.
85. Henry IM, Dilkes BP, Young K, Watson B, Wu H, Comai L: Aneuploidy and genetic variation in the *Arabidopsis thaliana* triploid response. *Genetics* 2005, **170**:1979-1988.
86. Ståhlberg D: Habitat differentiation, hybridization and gene flow patterns in mixed populations of diploid and autotetraploid *Dactylorhiza maculata* s.l. (Orchidaceae). *Evolutionary Ecology* 2009, **23**:295-328.
87. Thórsson ÆT, Salmela E, Anamthawat-Jónsson K: Morphological, cytogenetic, and molecular evidence for introgressive hybridisation in birch. *The Journal of Heredity* 2001, **92**(5):404-408.

doi:10.1186/1471-2148-11-346

Cite this article as: Jørgensen et al.: Interspecific and interploidal gene flow in Central European *Arabidopsis* (Brassicaceae). *BMC Evolutionary Biology* 2011 **11**:346.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

