MOLECULAR ECOLOGY

Molecular Ecology (2012) doi: 10.1111/j.1365-294X.2012.05460.x

'Missing link' species *Capsella orientalis* and *Capsella thracica* elucidate evolution of model plant genus *Capsella* (Brassicaceae)

HERBERT HURKA*, NIKOLAI FRIESEN†, DMITRY A. GERMAN‡ §, ANDREAS FRANZKE§ and BARBARA NEUFFER*

*Department of Botany, University of Osnabrück, Barbarastr. 11, D-49076 Osnabrück, Germany, †Botanical Garden of the University of Osnabrück, Albrechtstr. 29, D-49076 Osnabrück, Germany, ‡South-Siberian Botanical Garden, Altai State University, Lenina Str. 61, 656049 Barnaul, Russia, §Heidelberg Botanic Garden, Centre for Organismal Studies (COS) Heidelberg, Heidelberg University, Im Neuenheimer Feld 340, D-69120 Heidelberg, Germany

Abstract

To elucidate the evolutionary history of the genus *Capsella*, we included the hitherto poorly known species *C. orientalis* and *C. thracica* into our studies together with *C. grandiflora*, *C. rubella* and *C. bursa-pastoris*. We sequenced the ITS and four loci of noncoding cpDNA regions (*trn*L – F, *rps*16, *trn*H – *psb*A and *trn*Q – *rps*16). Sequence data were evaluated with parsimony and Bayesian analyses. Divergence time estimates were carried out with the software package BEAST. We also performed isozyme, cytological, morphological and biogeographic studies. *Capsella orientalis* (self-compatible, SC; 2*n* = 16) forms a clade (eastern lineage) with *C. bursa-pastoris* (SC; 2*n* = 32), which is a sister clade (western lineage) to *C. grandiflora* (self-incompatible, SI; 2*n* = 16) and *C. rubella* (SC; 2*n* = 16). *Capsella bursa-pastoris* is an autopolyploid species of multiple origin, whereas the Bulgarian endemic *C. thracica* (SC; 2*n* = 32) is allopolyploid and emerged from interspecific hybridization between *C. bursa-pastoris* and *C. grandiflora*. The common ancestor of the two lineages was diploid and SI, and its distribution ranged from eastern Europe to central Asia, predominantly confined to steppe-like habitats. Biogeographic dynamics during the Pleistocene caused geographic and genetic subdivisions within the common ancestor giving rise to the two extant lineages.

Keywords: biogeography, *Capsella*, cpDNA, isozymes, ITS, phylogeny age estimation

Received 29 August 2011; revision received 17 November 2011; accepted 26 November 2011.

Introduction

Wild relatives of the model organism *Arabidopsis* are increasingly in focus of contemporary evolutionary research programmes (Mitchell-Olds 2001; Koch *et al.* 2003; Hurka *et al.* 2005; Franzke *et al.* 2011). From all wild relatives of *Arabidopsis* currently used as study objects, *Capsella* is the most closely related genus. Molecular systematic studies confirm that both genera belong to the same tribe, Camelineae (Al-Shehbaz *et al.* 2006; Bailey *et al.* 2006; German *et al.* 2009; Warwick *et al.* 2010). Scientific research is focusing its attention increasingly on *Capsella* addressing such key issues as speciation, adaptation, mating systems and evolutionary developmental biology of plant form (Hurka & Neuffer 1997; Foxe *et al.* 2009; Guo *et al.* 2009; Paetsch *et al.* 2010; Neuffer 2011; Sicard *et al.* 2011; Theißen 2011). Additionally, sequencing of the *Capsella rubella* genome is currently being carried out by the Joint Genome Institute, United States Dept. of Energy. Many attempts to elucidate the evolutionary history of the genus *Capsella* in which one of the most widespread flowering plants on earth (*C. bursa-pastoris*) is included (Coquillat 1951) have already been undertaken (e.g. Shull 1929; Hurka & Neuffer 1997; Ceplitis *et al.* 2005; Slotte *et al.* 2006; St. Onge 2010), but, so far, no convincing hypothesis has been put forward. This has lead to controversy regarding, for example, phylogenetic relationships, mode of

Correspondence: Barbara Neuffer, Fax: +49 541 969 2845;
E-mail: neuffer@biologie.uni-osnabrueck.de

speciation, biogeographic origin and age estimations of the genus and its species.

Species delimination is difficult and controversial because of the enormous morphological variation within the genus. Chater (1993) list in Flora Europaea four *Capsella* species, which are commonly mostly accepted: *C. grandiflora* (Fauché & Chaub.) Boiss., *C. rubella* Reuter, *C. bursa-pastoris* (L.) Medik., including *C. thracica* Velen. as a subspecies, and *C. orientalis* Klokov. *Capsella grandiflora* and *C. rubella* are diploid ($2n = 2x = 16$), and *C. bursa-pastoris* is tetraploid ($2n = 4x = 32$). Interestingly, *Capsella orientalis* and *C. thracica* have never been the subject of experimental work, obviously due to the fact that no seed material was available. We included both taxa in our study and have, for the first time, explored the biosystematics and phylogenetics of these taxa.

The aim of this study was to reveal phylogenetic and biogeographic patterns within the genus *Capsella* covering all currently accepted taxa (Chater 1993). We analysed the nuclear internal transcribed spacers ITS1 and ITS2 including the 5.8 S gene, together with four different noncoding regions of the chloroplast genome. Shaw *et al.* (2007) provided an index of the relative levels of cpDNA variability. From among that list, we chose the less variable *trn*L – *trn*F intergenic spacer region and a highly variable cpDNA region, the *trn*Q – *rps*16 intergenic spacer, as well as two regions more or less intermediate in their levels of variation (*trn*H – *psb*A intergenic spacer, *rps*16 intron). We also performed isozyme analyses to study the genetic variation between and within species. The investigations were complemented by morphological, cytological and biogeographic studies. In the light of all the data presented in this study, it is obvious that *C. orientalis* and *C. thracica* hold a key position in our endeavours towards understanding the evolutionary history of the genus *Capsella*.

Material and methods

Origin of plant material

Seeds from *Capsella orientalis* were collected from single plants randomly taken from natural populations. The origin of the seed material is given in Table 1. Plants were cultivated from seeds either under greenhouse conditions or in the experimental garden of the Osnabrück University Botanical Garden and were used for phenotypic character analyses, cytology and isozyme studies. Herbarium specimens used for DNA sequencing and corresponding GenBank accession numbers are given in Table 2. Additional *Capsella* specimens were sequenced for ITS, and ITS sequences were also retrieved from GenBank, the origin or GenBank accession numbers of which are as follows: *C. grandiflora*: OSBU (Osnabrück University Herbarium) 12499; accession from seed genebank Gatersleben/Germany; sequence AM905718.1; *C. rubella*: OSBU 20858; *C. orientalis*: OSBU 10587; *C. bursa-pastoris*: OSBU 17229; OSBU 12500; sequences DQ310530.1; AF055196.1; AF128110; AF12811.1; *Neslia paniculata*: sequence AF137576.

Geographical distribution of Capsella orientalis

The geographical distribution of *C. orientalis* was established through literature surveys (Ebel 2002; German & Ebel 2009), our own field collections and by investigating herbarium collections. The following herbaria have been examined: ALTB (Altai State University, Barnaul, Russia); KW (Kholodny Institute of Botany, Kiev, Ukraine); LE (Komarov Botanical Institute, St. Petersburg, Russia); MHA (Moscow Main Botanical Garden, Russia); MW (Moscow State University, Russia); NS (Central Siberian Botanical Garden, Novosibirsk, Russia); OSBU (Botany Dept., University of Osnabrück, Germany); SVER (Institute of Plant and Animal Ecology, Jekaterinburg, Russia); TK (Tomsk State University, Russia); and without acronym: Pavlodar Pedagogical Institute (Pavlodar, Pavlodarskaya oblast, Kazakhstan).

Cytology and flow cytometry

Young flower buds were fixed overnight in Carnoy solution (acetic acid/ethanol = 1:3) at 4 °C, washed three times with ethanol (70%) and finally stored in ethanol (70%) at minus 20 °C. For preparation, the buds were washed twice with distilled water and three times with citrate buffer (pH 4.8). The material was digested with a pectolytic enzyme mix (cellulase, pectolyase, cytohelicase), and the buds were squeezed on glass slides with acetic acid, warmed to 50 °C and subsequently cooled with Carnoy solution and dried. Selected chromosome spreads of (pro)metaphase chromosomes of pollen mother cells were stained with 1–2 μg/mL DAPI (Roth, Karlsruhe), mounted in Vectashield and photographed at 1000-fold magnification using the Olympus BX-61 epifluorescence microscope system equipped with a Zeiss AxioCam HR CCD camera. To slow down bleaching of the fluorescence dye, a drop of DABCO solution (Roth, Karlsruhe, Germany) was applied. Pictures were viewed and processed with the photoshop software. At least five chromosome figures per slide and accession were analysed.

Flow cytometry was used to determine the relative DNA amount. Fresh leaf material was harvested, and *c.* 0.5 cm^2 leaf material was chopped with a sharp razor blade in a DAPI solution and filtered into a sample tube. Subsequent flow cytometry was performed on a Partec Ploidy Analyser-I (Partec, Münster, Germany).

Table 1 Origin of *Capsella orientalis* seed samples

Pop. no.	Country of origin, locality, habitat	Coordinates	Collector/remarks
1718	MN; Bayan-Olgiy Aymag; eastern end of lake Hoton Nuur, weed in lawn, mixed stand with *C. bursa-pastoris*	48° 35′ N 88° 26′ E	H. Hurka, B. Neuffer; voucher OSBU 10588
1719	MN; Bayan-Olgiy Aymag; between lakes Hoton Nuur and Horgon Nuur, sheep paddock	48° 35′ N 88° 26′ E	B. Neuffer, H, Hurka; voucher OSBU 10587
1938	RU; Siberia, Altai Kraj; city of Barnaul, ruderal, mixed stand with *C. bursa-pastoris*	53° 20′ N 83° 45′ E	D.A. German; voucher OSBU 18247
1939	KZ; Pavlodarskaya Oblast, Pavlodar, 400 km north-north-east from Astana, ruderal in lawn	52° 16′ N 76° 57′ E	D.A. German; voucher OSBU 18248
1940	KZ; Pavlodarskaya Oblast, 300 km east of Astana, near Bayanaul, ruderal in steppe country	50° 47′ N 75° 41′ E	D.A. German; voucher OSBU 18249
1941	KZ; Vostochno-Kazakhstanskaya Oblast, 750 km east of Astana; northern foothills of Kalbinskij Mt. Range, 15 km south of village Gagarino, steppe slopes	49° 59′ N 81° 48′ E	S.V. Smirnov; voucher ALTB
1978	RU; Siberia, Altai Kraj; Tretjakovsk raion, river valley Beresovja, at the Gilevskoe water reservoir, ruderal in steppe country	51° 06′ N 81° 54′ E	D.A. German, N. Friesen voucher ALTB
1979	RU; Siberia, Altai Kraj; Loktevsk raion, village Gilevo, ruderal in village	51° 07′ N 81° 48′ E	D.A. German, N. Friesen; voucher OSBU 19372
1980	RU; Siberia, Altai Kraj; Loktevsk raion, river valley Tushkanchikha, western slopes of mountain range, steppe slopes	51° 10′ N 81° 40′ E	D.A. German, N. Friesen; voucher ALTB
1981	RU; Siberia, Altai Kraj; Loktevsk raion, village Ust'yanka, ruderal in village	51° 08′ N 81° 36′ E	D.A. German, N. Friesen; voucher ALTB
1982	RU; Siberia, Altai Kraj; Rubzovsk raion, city of Rubzovsk, ruderal	51° 30′ N 81° 13′ E	D.A. German, N. Friesen; voucher ALTB
1983	RU; Siberia, Altai Kraj; Smeinogorsk raion, Kolyvanskoe Lake, ruderal in steppe country	51° 22′ N 82° 12′ E	D.A. German, N. Friesen; voucher OSBU 19373
1984	RU; Siberia, Altai Kraj; city centrum of Barnaul, ruderal	53° 21′ N 83° 44′ E	D.A. German; voucher OSBU 19374
1985	RU; Siberia, Altai Kraj; city of Barnaul, north-western part, ruderal	53° 21′ N 83° 44′ E	D.A. German; voucher OSBU 19375
2005	CN; Xinjiang, Dzungaria, 485 km north of Urumchi, Mongolian Altai, Fuhai county, ruderal	48° 05′ N 88° 56′ E	D.A. German *et al.*; voucher ALTB: SRAE2007653
2006	CN; Xinjiang, Dzungaria, 390 km northwest of Urumchi; Jeminay county, Saur, valley of Tastykarasu, 55 km south-east of Jeminay, rocky steppe slopes	47° 09′ N 86° 07′ E	D.A. German *et al.*; voucher ALTB: SRAE2007399
2007	CN; Xinjiang, Dzungaria, 410 km northwest of Urumchi; Jeminay county, Saur, 30 km south of Jeminay, meadow steppe, roadside	47° 14′ N 85° 43′ E	D.A. German *et al.*; voucher ALTB: SRAE2007042
2008	CN; Xinjiang, Dzungaria, 400 km northeast of Urumchi; Qinghe county, 40 km east of Qinghe, Mongolian Altai, valley of Tsagan-gol, 15 km northeast of Dunfyn; ruderal at local forest station	46° 37′ N 90° 52′ E	D.A. German *et al.*; voucher ALTB: SRAE2007897; OSBU 18585

Pop. no. refers to the *Capsella* seed collection hold at the Botany Dept. of the University of Osnabrück; country codes: CN, China; KZ, Kazakhstan; MN, Mongolia; RU, Russia; samples are individual seed samples except for pop. 1941. ALTB: Herbarium Altai State University, Barnaul, Russia; OSBU: Herbarium Botany Dept., University Osnabrück, Germany.

Petroselinum crispum was used as an internal standard (2C-value of absolute DNA amount 4.46 pg, Yoyoka *et al.* 2000; 1C-value of absolute DNA amount for *C. rubella* 0.22 pg (2*C* = 0.44 pg) and 1C-value of absolute DNA amount for *C. bursa-pastoris* 0.4 pg (2*C* = 0.8 pg), Lysak *et al.* 2009).

Isozyme analyses

Isozyme investigations of *Capsella orientalis* and of *C. thracica* were carried out with progeny raised from the provenances listed in Table 1 or Table 2, respectively. Rosette leaves of single plants, and *c.* 10 weeks old, were harvested and stored at −80 °C. Electrophoresis was performed in a continuous system on vertical polyacrylamide gel slabs. The following enzyme systems were assayed: aspartate aminotransferase (AAT; EC 2.6.1.1), glutamate dehydrogenase (GDH; EC 1.4.1.4) and leucine aminopeptidase (LAP; 3.4.11.1). Buffer systems and other experimental details are given in Hurka *et al.* (1989) for AAT, in Hurka & Düring (1994) for GDH and in Neuffer & Hurka (1999) for LAP. The

Table 2 Provenances of *Capsella* and *Neslia* specimens used for DNA sequencing and GenBank accession numbers

Species	Country of origin, locality, coordinates	Voucher	GenBank accession numbers				
			ITS	*trn*Q-*rps*16 spacer	*rps*16 intron	*trn*H-*psb*A spacer	*trn*L-*trn*F spacer
C. grandiflora	GR; Prov. Joannina, Metsovo; N 39° 46′, E 21° 10′	OSBU 7339	FR773701	FR822325	FR822364	FR822352	FR822334
C. grandiflora	IT; Prov. Brescia, Pilzone/Lago Iseo; N 45° 41′, E 10° 05′	OSBU 18615	FR773702	FR822324	FR822365	FR822353	FR822335
C. rubella	CL; Región Biobío, near Concepción; S 36° 50′, W 73° 03′	OSBU 7334	FR773704	FR822322	FR822362	FR822350	FR822336
C. rubella	IT; Prov. Foggia, Mte. Gargano, Foresta Umbra; N 41° 49′, E 15° 59′	OSBU 20857	FR773703	FR822323	FR822363	FR822351	FR822337
C. bursa-pastoris	DE; North Rhine-Westphalia, north of Muenster; N 52° 19′, E 07° 56′	OSBU 14439	FR773707	FR822329	FR822358	FR822345	FR822341
C. bursa-pastoris	RU; Novosibirskaya Oblast, near Novosibirsk, N 52° 20′, E 82° 54′	OSBU 12815	FR773706	FR822330	FR822356	FR822346	FR822342
C. bursa-pastoris	TR; Prov. Antalya, Taurus Mts., Bey Daglari massif, N 36° 52′, E 30° 15′	OSBU 18590	FR773705	FR822331	FR822357	FR822344	FR822343
C. orientalis	KZ; Pavlodarskaya Oblast, Pavlodar; N 52° 16′, E 76° 57′	OSBU 18248	FR773709	FR822327	FR822360	FR822347	FR822339
C. orientalis	KZ; Pavlodarskaya Oblast, near Bayanaul; N 50° 47′, E 75° 41′	OSBU 18249	FR773710	FR822326	FR822361	FR822348	FR822340
C. orientalis	MN; Bayan-Olgiy Aymag; Lake Hoton Nuur, N 48° 35′, E 88° 26′	OSBU 10588	FR773708	FR822328	FR822359	FR822349	FR822338
C. thracica	BG; Sozopol, *c.* 20 km south-east from Burgas, N 42° 25′, E 27° 42′	OSBU 20859	HE575237 HE575238 HE575239 HE575240 HE575241	HE575225	HE575234	HE575231	HE575228
C. thracica	BG; Sozopol, *c.* 20 km south-east from Burgas, N 42° 26′, E 27° 42′	OSBU 20860	HE575242 HE575243	HE575226	HE575235	HE575232	HE575229
C. thracica	BG; Thracian Plain, Kurtovo Konare, N 42°05′, E 24°30′	OSBU 20875	HE575244	HE575227	HE575236	HE575233	HE575230
N. paniculata	DE; Bavaria, Frankonian mountain region; N 50° 06′, E 11° 01′	OSBU 6887	FR773711	FR822332	FR822355	FR822354	FR822333

OSBU, Herbarium of the Botany Dept. of the University of Osnabrück, Germany; country codes: BG, Bulgaria; CL, Chile; DE, Germany; GR, Greece; IT, Italy; KZ, Kazakhstan; MN, Mongolia; RU, Russia; TR, Turkey.

genetics of these enzyme systems in *Capsella* has been deciphered in the above-cited literature, and the previous nomenclature of the enzyme loci and their allozymes was adopted in this study. Isozyme data for the species *C. grandiflora*, *C. rubella* and *C. bursa-pastoris* either were previously published or are presented here for the first time.

DNA sequencing

The nuclear ribosomal internal transcribed spacers ITS1 and ITS2 including the 5.8 S region as well as four noncoding regions of the chloroplast genome have been analysed. Genomic DNA was sampled from herbarium specimens listed in Table 2 using the 'InnuPREPP Plant DNA kit' (Analytic Jena AG) according to the instructions of the manufacturer and was used directly in PCR amplifications.

Amplification and sequencing primers for ITS are given in German *et al.* (2009). Primers for the chloroplast regions were as follows: for the *trn*Q-*rps*16 region described in Shaw *et al.* (2007), for *rps*16 intron described in Oxelman *et al.* (1997), for *trn*L-*trn*F described in Taberlet *et al.* (1991) and for *trn*H-*psb*A described in Kress *et al.* (2005). Products of the cycle sequencing reactions were run on an ABI 377XL automated sequencer. Forward and reverse sequences from each individual were manually edited in CHROMAS Lite 2.1 (Technesylum Pty Ltd) and combined in single consensus sequences. The sequences of all samples were aligned with CLUSTAL X (Thompson *et al.* 1997) and subsequently corrected manually in MEGA 5 (Tamura *et al.* 2011).

To test for multiple ITS copies within individuals of *C. thracica,* we also cloned PCR amplicons using the TOPOTA Cloning® kit (Invitrogen) according to the instructions of the manufacturer. The DNA of 16 clones was isolated with NucleoSpin plasmid kit (Macherey-Nagel, Düren, Germany) according to the instructions of the manufacturer and prepared for sequencing. Sequencing was performed on ABI 377XL automatic sequencer with universal M13 forward and reverse primers.

Phylogenetic analyses

Neslia paniculata (L.) Desv. has been chosen as an outgroup based on the analyses of Bailey *et al.* (2006) and Couvreur *et al.* (2010). Parsimony analysis was performed with PAUP* 4.0b10 (Swofford 2002) using heuristic searches with TBR and 100 random addition sequence replicates. Bootstrap support (BS; Felsenstein 1985) was estimated with 100 bootstrap replicates, each with 100 random addition sequence searches. Bayesian analyses were implemented with MrBayes 3.1.23 (Ronquist & Huelsenbeck 2003). Sequence evolution models were evaluated using the Akaike Information Criterion (AIC) with the aid of Modeltest 3.7 (Posada & Crandall 1998). Two independent runs each of eight chains, 10 million generations, sampling every 100 trees. 25% of initial trees were discarded as burn-in. The remaining 28 000 trees were combined into a single data set and a majority-rule consensus tree obtained. Bayesian posterior probabilities were calculated for that tree in MrBayes 3.1.23.

Divergence time estimates in Capsella

Divergence time estimates were carried out with the software package BEAST v1.4.8 (Drummond & Rambaut 2007) based on ITS sequences (ITS1 and ITS2 regions combined, 5.8 S gene region excluded). No intraspecific ITS variation was detected between five provenances of *Capsella grandiflora*; three of *C. rubella*; four of *C. orientalis*; and nine of *C. bursa-pastoris* (see chapter *Origin of plant material*). Therefore, for the BEAST analysis, the ITS data matrix was reduced to four taxon sequences. Branch length was calibrated using a mean published ITS substitution rate for herbaceous annual/perennial angiosperms of 4.13×10^{-9} subs/site/yr (Kay *et al.* 2006) under the GTR + I + G substitution model, the uncorrelated lognormal relaxed clock approach, the Birth-Death speciation process performing a chain length of 100 000 000. Stationarity of the MCMC chain and the effective sampling size (ESS) of each parameter were examined in Tracer v1.4.1 (Drummond & Rambaut 2007, available from http://beast.bio.ed.ac.uk/Tracer), and each ESS was above 1000.

Results

Morphology, cytology and geographical distribution of Capsella orientalis *and* Capsella thracica

Capsella orientalis. *Capsella orientalis* is morphologically very close to *C. bursa-pastoris* and often confused with it. Chromosome counts of $2n = 16$ for *C. orientalis* are cited by Dorofeyev (2002) but without reference. Krasnoborov *et al.* (1980) reported $2n = 16$ for '*C. bursa-pastoris*', a count that was probably based on *C. orientalis* and not on *C. bursa-pastoris*. Our data unambiguously prove diploidy for *C. orientalis* with $2n = 16$ (Fig. 1). Thus, in addition to morphological details, the most important difference between *C. orientalis* and *C. bursa-pastoris* is the ploidy level: *C. orientalis* is diploid with $2n = 2x = 16$, and *C. bursa-pastoris* is tetraploid with $2n = 4x = 32$ (Fig. 1). Flow cytometry suggests that,

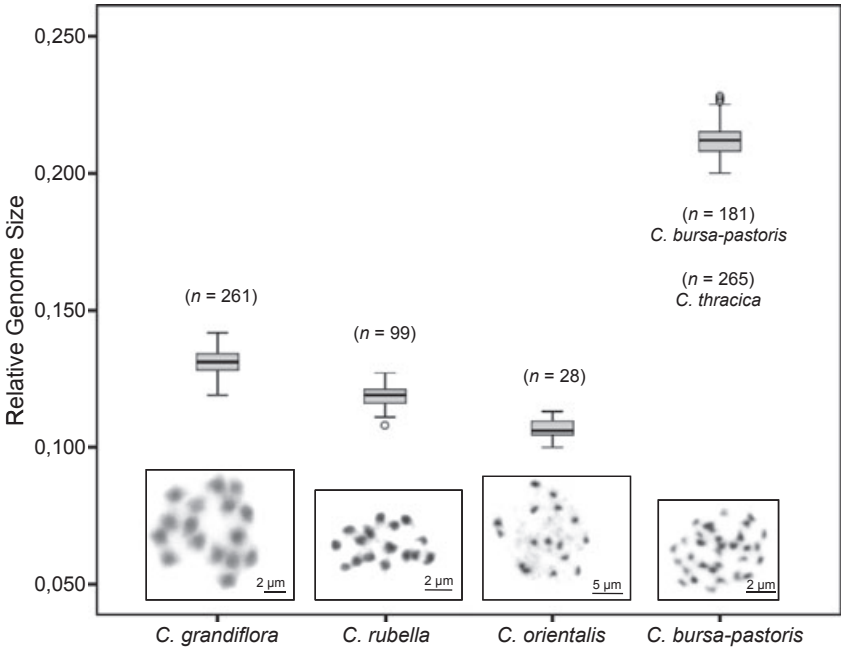

Fig. 1 Figuration of chromosomes and relative DNA amount of *Capsella* species: chromosome pictures are from metaphase plates from pollen mother cells. Relative DNA amount revealed by flow cytometry, standard: *Petroselinum crispum*; *n* = number of measured individuals.

despite equal chromosome numbers, the relative DNA content between *C. orientalis* and the other diploid species, *C. grandiflora* and *C. rubella*, is somewhat different between the three diploid species (Fig. 1). *Capsella orientalis* is fully self-compatible, as proven by our own greenhouse and field experiments. Our literature and herbarium survey revealed that *C. orientalis* has a much wider distribution area than hitherto reported (Fig. 2). It ranges from the middle Ukraine through the southern part of European Russia, the South Urals, northern Kazakhstan, south-west Siberia up to western Mongolia and north-western China (Xinjiang region). This distribution coincides noticeably with the middle and western part of the Eurasian steppe belt which stretches from south-eastern Europe to north-eastern China.

Capsella thracica. *Capsella thracica* is a Bulgarian endemic (Fig. 2) and, like *C. orientalis*, morphologically very close to *C. bursa-pastoris*. The main feature differentiating this species from *C. bursa-pastoris* is the elongated style. Just like *Capsella bursa-pastoris*, *C. thracica* is tetraploid as has been revealed by chromosome counts and

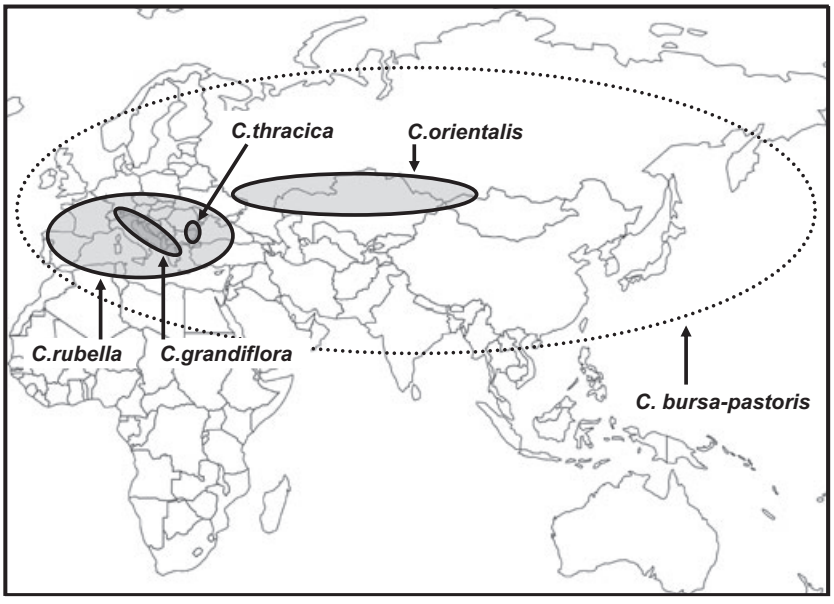

Fig. 2 Outline distribution map of *Capsella* species. *Capsella grandiflora*: western Balkan, northern Italy; *C. rubella*: circum Mediterranean; *C. orientalis*: eastern Europe to central Asia; *C. thracica*: Bulgaria. Putative native range of *C. bursa-pastoris* is shown by dotted line. The worldwide distribution of *C. bursa-pastoris* and colonized regions of *C. rubella* in the New World and Australasia are not indicated.

flow cytometry (Fig. 1) and is predominantly selfing as revealed by isozyme progeny analyses.

Phylogenetic analyses

ITS sequence data. Direct sequencing of the ITS PCR products produced unambiguous sequences, with the exception of *Capsella thracica* accessions. In *C. thracica*-12, we obtained different sequences using forward and reverse primers. The forward primer resulted in a sequence almost identical to *C. grandiflora*, and the reverse primer in a sequence identical to *C. bursa-pastoris/C. orientalis*. The two other *C. thracica* accessions, no. 11 and 13, displayed at ITS sequence positions 122–126, two identical peaks that can be translated as RWWW (R = A and G; W = A and T), showing that *C. thracica* has at least two different copies of rDNA in its genome. To confirm this, we cloned ITS PCR products of accession *C. thracica*-11. In the 16 sequenced clones, 14 sequences were identical with *C. bursa-pastoris* and two sequences almost identical to *C. grandiflora*; in *C. thracica,* one nucleotide was missing in a poly-T-motif. These additional copies were included in the analyses.

The alignment of combined ITS1 and ITS2 sequences, including the 5.8 S gene of the taxa listed in Table 2, generated a matrix of 640 characters, of which 10 were parsimony informative. For the Bayesian analyses, the substitution model K80 was chosen by AIC in Modeltest 3.7. Unweighted parsimony analysis of the 19 sequences resulted in a single most parsimonious tree of 60 steps (CI = 1.000; Fig. 3). *Capsella bursa-pastoris* and *C. orientalis* formed a clade supported by 98% bootstrap value and 1.00 Bayesian posterior probabilities. This clade is a sister group to the clade consisting of *C. grandiflora* and *C. rubella* (58% bootstrap support, 0.70 Bayesian posterior probabilities) (Fig. 3). Within the two sister clades, *C. orientalis* is resolved from *C. bursa-pastoris* by 62% bootstrap support and 0.95 Bayesian posterior probabilities, and *C. rubella* from *C. grandiflora* by 74% bootstrap and 0.98 Bayesian probabilities. The *C. thracica* accessions analysed (Table 2) displayed two different ITS sequence types, one from the *C. grandiflora/C. rubella* lineage and one from the *C. bursa-pastoris/C. orientalis* lineage (Fig. 3).

CpDNA sequence data. Phylogenetic analyses were conducted separately with each cpDNA region sequenced. The alignments generated matrices of 855 characters for the *rps*16 intron with 8 (0.93%) parsimony informative characters; 366 characters for the *trn*H-*psb*A region with 10 (2.73%) parsimony informative characters; 469 characters for the *trn*Q-*rps*16 region with 13 (2.77%) parsimony informative characters; and 756 characters for the

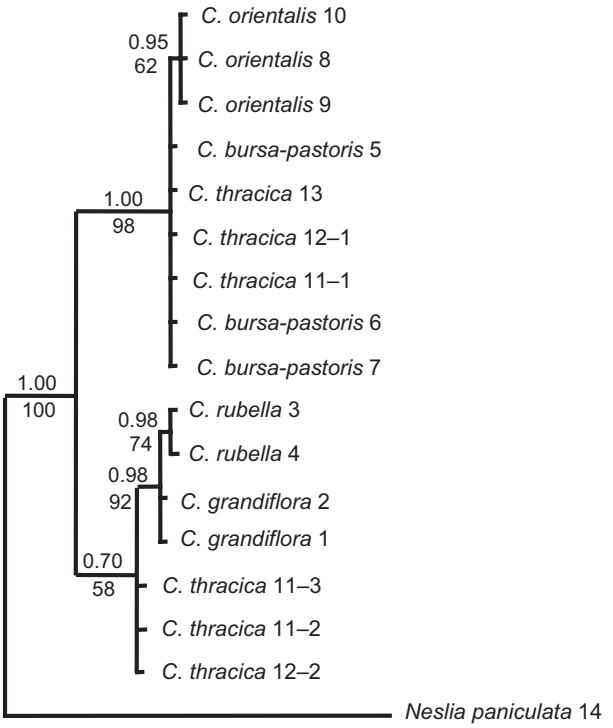

Fig. 3 Phylogenetic tree for *Capsella* species based on ITS: Bayesian posterior probabilities above branches, bootstrap support over 50% below branches. For *C. thracica* 13 only the original sequence with two peaks at positions 122–126 was included in the analyses. For further information, see in the chapter Results.

*trn*L-*trn*F region with 101 (13.35%) parsimony informative characters.

The *trn*L-F spacer region in *Capsella* displayed noticeable length variations caused by varying numbers of up to six repeats of 70–80 bp length. The repeats are characterized by a recurrent motif of *c.* 10 bp (GCTTTTTTTG), occasionally modified by single nucleotide and indel polymorphism. Excluding the gaps in the total alignment of 756 characters, *trn*L-F intergenic spacer length was 720 bp in *Capsella grandiflora* and *C. rubella*, and 703 bp in *C. bursa-pastoris*, *C. thracica* and *C. orientalis* accessions 8 and 10, whereas *C. orientalis* 9 had a length of only 562 bp because of complete or part loss of three of the six repeats. Following Koch *et al.* (2005, 2007), we interpret the repeats as *trn*F pseudogenes, which, according to the above-mentioned authors, cause extensive length variation of the *trn*L-F regions in many Brassicaceae. We removed the region with varying repeats (pseudogenes) from the total *trn*L-F alignment. The discarded fragment had a length of 432 characters (alignment positions 310–742) leaving a *trn*L-F alignment of 322 characters, which was implemented in the phylogenetic analysis.

As the phylogenetic trees for the single four cpDNA regions did not produce contradictory results (trees not shown), we combined the cpDNA sequences, generating a combined matrix of 2012 characters, of which 34 (1.7%) were parsimony informative. Parsimony analysis resulted in a single most parsimonious tree of 132 steps (CI = 0.992). For the Bayesian analysis, the substitution model TIM + I was selected by AIC in Modeltest 3.7. The resulting phylogenetic tree (Fig. 4) reflects the main features: the sister group relationship between the clade *C. bursa-pastoris/C. orientalis/C. thracica* on the one side and the clade *C. grandiflora/C. rubella* on the other is supported by high significance values. There are subgroups within the two clades, for example, one *C. orientalis* accession clustered with *C. bursa-pastoris*, and there is also clustering between the *C. bursa-pastoris* accessions. The subgroups in the combined DNA data set mirror corresponding variation in the *trn*Q-*rps*16 and *trn*H-*psb*A intergenic spacer regions, known to be highly variable noncoding cp DNA regions (Shaw *et al.* 2007).

Divergence time estimates with BEAST

Relaxed clock estimates using BEAST and a published ITS substitution rate for herbaceous/perennial angiosperms resulted in a crown age of the genus *Capsella* of 3.18 myr (95% HPD, 0.58 to 6.98 myr; HPD, highest posterior density intervals, is equivalent to confidence intervals). The split between *C. rubella* and *C. grandiflora* was dated 0.86 myr (95% HPD, 0.015–2.45 myr), and the divergence time of *C. bursa-pastoris* and *C. orientalis* was estimated at 0.87 myr (95% HPD, 0.006–2.44 myr).

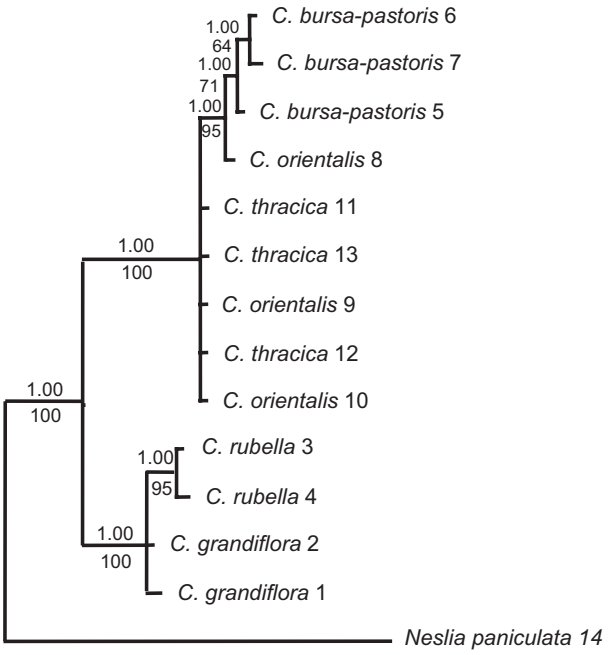

Fig. 4 Phylogenetic tree for *Capsella* species based on a combined cpDNA data set: *trn*L – *trn*F, *rps*16, *trn*H – *psb*A, *trn*Q – *rps*16 regions. Bayesian posterior probabilities above branches, bootstrap support below branches.

Isozyme analyses

Whereas allozyme frequencies within *C. grandiflora*, *C. rubella* and *C. bursa-pastoris* have been intensively studied (Hurka & Neuffer 1997; Neuffer & Hoffrogge 2000; Neuffer & Hurka 1999; Neuffer *et al.* 1999; Neuffer 2011; Neuffer & Hurka, unpublished), isozyme data for *Capsella orientalis* and *C. thracica* are documented here for the first time. *Capsella grandiflora* and *C. bursa-pastoris* share most of their allozymes, but the two alleles *Aat1-4* and *Aat3-5*, rather common in *C. bursa-pastoris*, have not been recorded for *C. grandiflora* and thus appear unique for *C. bursa-pastoris* (Fig. 5). All *C. orientalis* plants that we have analysed so far (123 individuals from 16 populations from Siberia, Kazakhstan, Mongolia and China, Table 1) were nearly monomorphic regarding the isozyme loci analysed. Only at the *Aat2* locus did we find two alleles, *Aat2-1* and *Aat2-7* (Fig. 5). The frequency of *Aat2-1* was $f = 0.77$ and that of *Aat2-7* was $f = 0.29$. Four heterozygotes between *Aat2-1* and *Aat2-7* have been detected so far. All alleles found in *C. orientalis* have also been recorded for the diploid *C. grandiflora* and the tetraploid *C. bursa-pastoris*, but *C. orientalis* displayed only a fraction of the allele spectrum discovered in the latter two species (Fig. 5). All allozymes recorded for *C. thracica* are also found in *C. bursa-pastoris*, and no private alleles for *C. thracica* have been detected so far.

Discussion

Molecular phylogeny of the genus Capsella

Two lineages within Capsella. The principle finding of our phylogenetic studies is evidence of two extant groups within the genus *Capsella*. The two diploid species *C. grandiflora* and *C. rubella* are a sister clade to a clade consisting of the diploid *C. orientalis* and the tetraploid *C. bursa-pastoris* (Fig. 3 and 4).

In these taxa, no intraspecific variation of the nuclear ribosomal ITS region was detected (Fig. 3), in contrast to the noncoding cpDNA (Fig. 4) analysed. The phylogenetic position of the tetraploid *C. thracica* is discussed later.

Divergence time estimates

Published time estimates for Brassicaceae 'lineage I', to which *Arabidopsis* and *Capsella* belong (Beilstein *et al.*

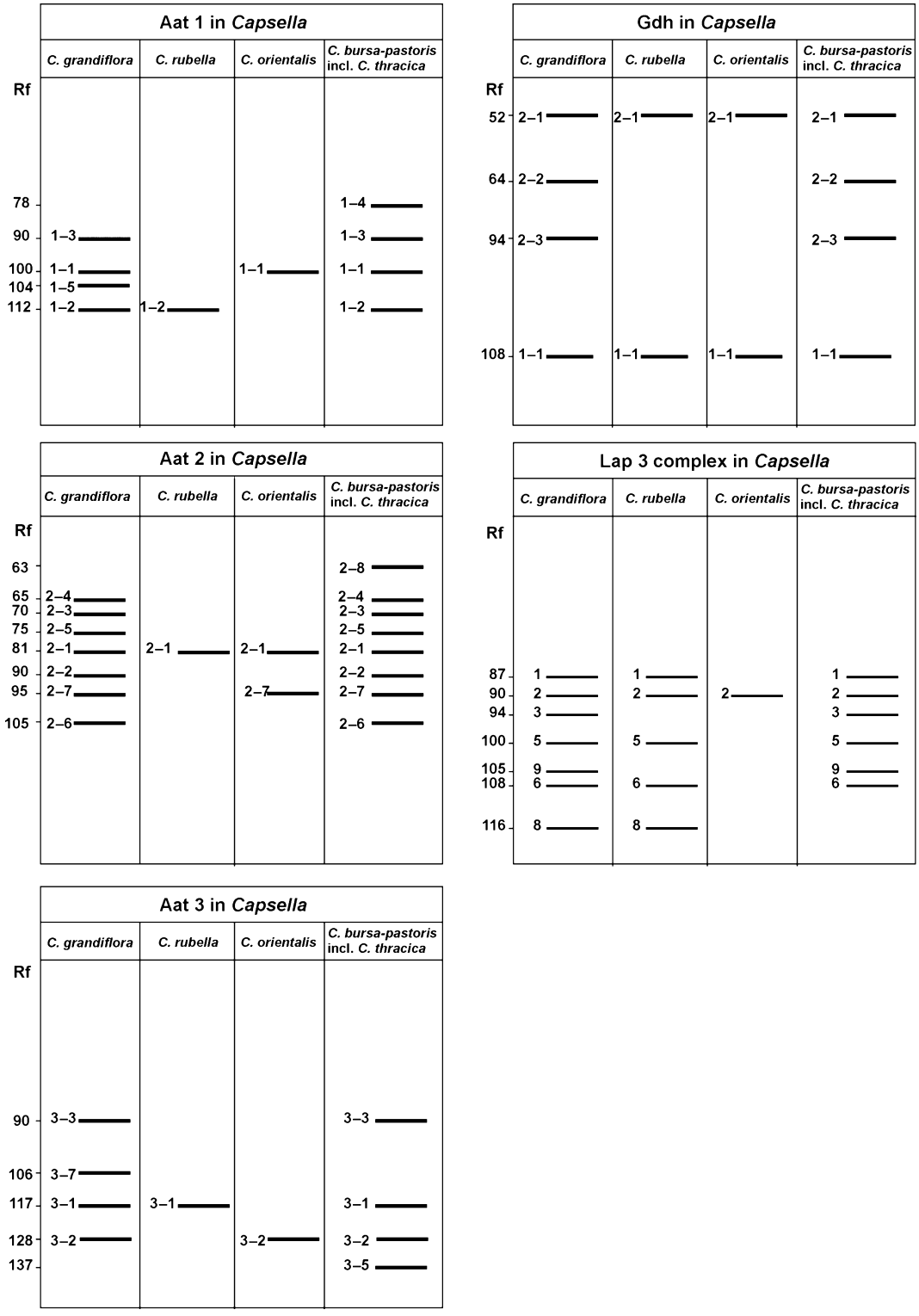

Fig. 5 Presence/absence allozyme profiles of *Capsella* species: isozyme loci are given at the head of the diagrams. Rf values refer to an internal standard allozyme band set at value 100. Individuals examined: *C. orientalis* $n = 123$ of 16 populations; *C. thracica* $n = 30$ of 3 populations; *C. grandiflora*, *C. rubella* $n > 1000$ for each of the species and *C. bursa-pastoris* $n > 20\,000$ covering the entire species ranges.

2006), are 19–13 myr (Koch *et al.* 2000, 2001), 19.0–8.0–0.5 myr (Franzke *et al.* 2009), 36.1–27.3–18.2 (Couvreur *et al.* 2010) and 42.8–35.6–28.5 myr (Beilstein *et al.* 2010). The age of the tribe Camelineae, which includes *Arabidopsis* and *Capsella*, is estimated to be 17.9–13.0–8.0 myr (Beilstein *et al.* 2010). The split between the *Arabidopsis* lineage and its sister clade that includes *Capsella* is estimated at 14.6–10–5.7 myr (Koch *et al.* 2000), and separation of *Arabidopsis* and *Capsella* is dated 9.8–6.2 myr by Ačarkan *et al.* (2000). Divergence between *Arabidopsis thaliana* and its close relatives is estimated at 9.0–5.0–3.1 myr by Koch *et al.* (2000), whereas Ossowski *et al.* (2010) advocate the separation of *Arabidopsis thaliana* (self-compatible) from *A. lyrata* (self-incompatible) 18 myr ago. Such a high age, in connection with the assumption that *A. thaliana* probably has been self-fertile since its separation from *A. lyrata* (Wright *et al.* 2002), appears to contrast with the statement of Tang *et al.* (2007) that selfing in *A. thaliana* most likely evolved a 'million years ago or more'. Thus, age estimates published for *Arabidopsis* and its close relative *Capsella* vary considerably, and it is well known that molecular date estimates may be full of substantial errors (Graur & Martin 2004; Welch & Bromham 2005; Pulquério & Nichols 2007). Nevertheless, lacking old *Capsella* fossils, we used published ITS substitution rates to provide rough estimates for dating divergences within the genus. Given the large range of the 95% highest posterior density intervals (HPD, equivalent of confidence intervals) of our analysis, we do not want to over-interpret our dating estimates. Our main conclusion from our dating analysis is that the genus *Capsella* is of pre-Pleistocene origin and that diversification within the genus which lead to its extant members most likely occurred during Pleistocene times. Thus, our date estimates are within the range of most published age estimates on *Capsella* and its close relatives.

Mode, time and place of origin of Capsella *species*

To avoid confusion of terminology, and in accordance with the recent relevant literature (Ramsey & Schemske 2002; Soltis *et al.* 2007), we have used the term autopolyploidy to denote origin of a polyploid taxon within or between populations of a single species, whereas allopolyploids are derived from interspecific hybridizations. Thus, autopolyploidy is synonymous with the intraspecific mode of origin and allopolyploidy with the interspecific mode of origin.

Capsella grandiflora *and* Capsella rubella. *Capsella grandiflora* is diploid and self-incompatible (SI) because of a sporophytic self-incompatibility system (Paetsch *et al.* 2006). Although the majority of extant *Capsella* species are self-compatible (SC), self-incompatibility should surely be regarded as the ancestral character state (e.g. Sherman-Broyles & Nasrallah 2008). As stated earlier, we conclude from our dating estimates that *C. grandiflora* and *C. rubella* are of Pleistocene age. Based on the present-day distribution of *C. grandiflora* and its sister taxon *C. rubella* (Fig. 2), we hypothesize that the place of origin for both species was the western part of a former larger distribution area of the most recent common ancestor as will be discussed below (Fig. 6).

The diploid, predominantly selfing, *C. rubella* is a derivative of the *C. grandiflora*-like most recent common

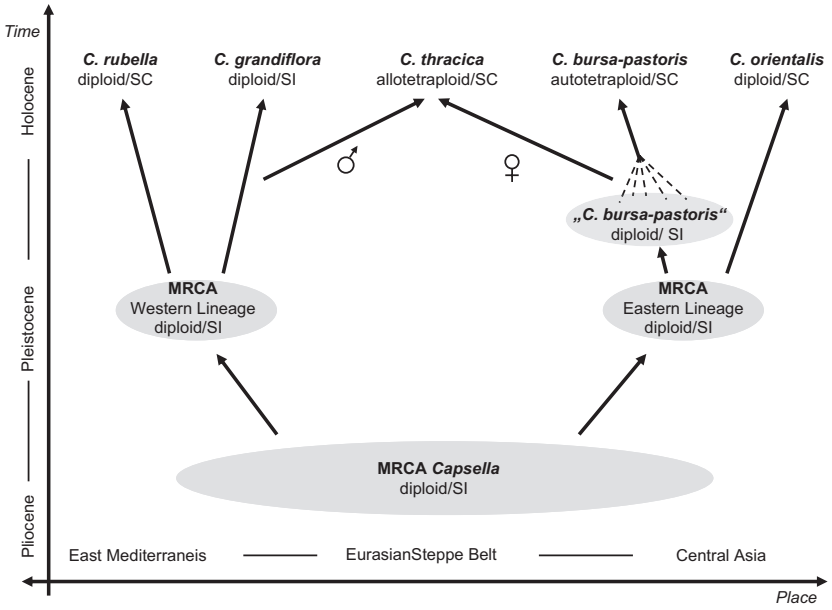

Fig. 6 Outline of the evolutionary history of the genus *Capsella*. Broken lines indicate multiple origins of *C. bursa-pastoris*.

ancestor (diploid and SI) of the western lineage. Associated with this speciation process was the transition from SI to SC (Hurka & Neuffer 1997; Foxe *et al.* 2009; Guo *et al.* 2009). *Capsella rubella* harvested only a fraction of the allozyme diversity of *C. grandiflora* (Fig. 3), which in connection with the findings of Guo *et al.* (2009) of only 1 or 2 alleles at most loci argues for a single origin. Foxe *et al.* (2009) and Guo *et al.* (2009) estimated that the two species, *C. grandiflora* and *C. rubella*, separated very recently, from less than 25 000 (Foxe *et al.* 2009) to 30 000 to 50 000 years ago (Guo *et al.* 2009). A Pleistocene origin of *C. rubella* and *C. grandiflora* is also indicated by our dating estimates (0.015–) 0.86 (–2.45) myr. A young age of *c.* 25 000–50 000 years as advocated by Foxe *et al.* (2009) and Guo *et al.* (2009) (transition from Pleistocene to Holocene) would imply unprecedented high ITS substitution rates, whereas the ITS substitution rates used in our analysis are in line with other accepted Quaternary ITS-based biographic scenarios for Brassicaceae taxa (Bleeker *et al.* 2002; Franzke *et al.* 2004; Mummenhoff *et al.* 2004). The place of origin of *C. rubella* was presumably the eastern Mediterranean region. Subsequently, *C. rubella* extended its range, colonized all Mediterranean countries and spread later with European colonists to North and South America and Australasia (Neuffer & Hurka 1999; Neuffer *et al.* 1999; Paetsch *et al.* 2010).

Capsella orientalis *and* Capsella bursa-pastoris. *Capsella orientalis* is, as is *C. rubella*, a diploid and predominantly selfing species (SC) with very low allozyme variability (Fig. 5). However, the distribution areas of the two diploid species appear to be mutually exclusive (Fig. 2), and the phylogenetic roots of the two species are different as clearly shown by ITS and cpDNA data (Figs 3 and 4).

The split between the sister species *C. orientalis* and the tetraploid self-compatible *C. bursa-pastoris* was estimated by us to be (0.006-) 0.87 (–2.44) myr ago (Pleistocene), which is the same as has been estimated for the split between *C. grandiflora* and *C. rubella*. The present-day distribution area of *C. orientalis* (Fig. 2) suggests that the species split between *C. orientalis* and *C. bursa-pastoris* has occurred in the more eastern parts of the Eurasian distribution belt (Figs 2 and 6). The DNA variation detected in *C. orientalis* and *C. bursa-pastoris* (Fig. 4) might argue for multiple origins of both species.

Our present data on nuclear and chloroplast DNA variation demonstrate that *C. bursa-pastoris* is not, as was argued earlier, a derivative species of *C. grandiflora* (Figs 3 and 4) (Hurka & Neuffer 1997; Slotte *et al.* 2006, 2008; St. Onge 2010), nor does this uphold an argument in favour of single origin (Slotte *et al.* 2006, 2008). Instead, cpDNA variation data (Fig. 4), high isozyme polymorphism (Fig. 5), as well as RAPD (Neuffer 1996) and AFLP data (Hameister *et al.* 2009) support the assumption of multiple origin of *C. bursa-pastoris*, as does the enormous morphological polymorphism (Almquist 1907, 1921). Presence/absence data on allozymes reveal that *C. grandiflora* and *C. bursa-pastoris* share most of their allozymes (Fig. 5). As there is no progenitor–derivative relationship between the two species (Figs 3 and 4), we interpret the concurrence of the allozymes, which are low mutation markers, in these two species as an ancient polymorphism inherited from the most recent common ancestor. It is highly unlikely that the shared allozymes are because of convergence.

Polyploidy in Capsella bursa-pastoris. There is no clear evidence for an allopolyploid origin of the tetraploid *C. bursa-pastoris*. Attributes of *C. bursa-pastoris*, like disomic inheritance, shown for allozymes (Hurka *et al.* 1989; Hurka & Düring 1994; Neuffer & Hurka 1999) and morphological characters (Shull 1929), and 'fixed heterozygosity' (true-breeding multiple banded isozyme patterns, Hurka *et al.* 1989; Hurka & Düring 1994), may argue for allopolyploid origin. However, it is well known that autopolyploids often behave cytologically like allopolyploids (Ramsey & Schemske 2002). Allopolyploids should retain a degree of hybrid character of their genomes (Ramsey & Schemske 2002), which could not as yet be demonstrated for *C. bursa-pastoris*. The occasional findings of *C. rubella* nuclear haplotypes in *C. bursa-pastoris* in southern Europe, where the *C. grandiflora/rubella* lineage and the *C. orientalis/bursa-pastoris*-lineage are sympatric, are probably due to introgression (Slotte *et al.* 2006, 2008). This interpretation is supported by the lack of such haplotypes in *C. bursa-pastoris* from China, where neither *C. grandiflora* nor *C. rubella* occur (Slotte *et al.* 2008). In agreement with previous studies (Hurka & Neuffer 1997; Slotte *et al.* 2006, 2008; St. Onge 2010), we thus again argue for an autopolyploid origin of *C. bursa-pastoris*. However, it should be kept in mind that signals indicating the hybrid nature of a species may be eradicated with time.

The ancestor that gave rise to *C. orientalis* and *C. bursa-pastoris* was most probably diploid and self-incompatible (SI). The shift from SI to SC in *C. bursa-pastoris* might have coincided with the polyploidization process leading to the extant tetraploid *C. bursa-pastoris*. Although the multiple origin of *C. bursa-pastoris* may imply origin not only at different places but also at different times, we nevertheless argue that polyploidization occurred in the Middle/Late Pleistocene times. Such a scenario is in accordance with recent coalescence analyses. Based on microsatellite data, the most recent common ancestor for the chloroplast genome of

C. bursa-pastoris has been estimated at 7000–17 000 years ago by Ceplitis *et al.* (2005) (late Pleistocene to Holocene), whereas Slotte *et al.* (2006), basing their estimate on cpDNA sequence data, date this occurrence between 43 000 and 430 000 years ago (Pleistocene). Tetraploid *Capsella bursa-pastoris* would then be another prime example of colonization success of a polyploid plant species. A middle to late Pleistocene origin of tetraploid *C. bursa-pastoris* is also in line with fossil records. Macrofossils (seeds) of *Capsella* have been reported from the interglacial deposits at Ilford, Essex, England, and have been identified as *C. bursa-pastoris* (West *et al.* 1964). The sediments are deemed to be Ipswichian (Eemian of continental Europe) and thus correlate with MIS (Marine Isotope Stage) 5e (Shackleton *et al.* 2003). More recently, however, it has been argued that the Ilford deposits belong to the penultimate interglacial complex (Hoxne = Holstein Interglacial) and correlate to MIS 7 (Turner 2000). Estimations for the duration of MIS 5e are *c.* 125 000–110 000 years BP (late Pleistocene), and for MIS 7, from 245 000 to 185 000 years BP (middle Pleistocene). In any case, there is evidence of a pre-(last) glacial occurrence of *Capsella* in western Europe, and *Capsella* might already have colonized western Europe in the middle Pleistocene. This does not contradict or deny postglacial anthropogenic introduction.

Based on several arguments, we hypothesize that the place of origin of *C. bursa-pastoris* is eastern Europe/-western to central Asia. (i) The main distribution area of *C. orientalis*, the sister species of *C. bursa-pastoris*, is eastern Europe (Transvolga) through North Kazakhstan to southwest Siberia, northwest China and western Mongolia. Allozyme *Aat2-7* that had a considerably high frequency of $f = 0.29$ in *C. orientalis* was also detected in *C. bursa-pastoris*, but only in accessions from eastern Europe (Russia: Moscow region, Voronezh/Don, Astrakhan, Teberda/Caucasus) and central Asia (Kirgistan: Tian Shan and Pamir Alai). (ii) Some alleles were unique for *C. bursa-pastoris* including the very common alleles *Aat1-4* and *Aat3-5* (Fig. 5). It is unlikely that we missed these alleles in *C. grandiflora* because of undersampling, because we sampled *C. grandiflora* throughout its distribution area intensively but could find no evidence of these alleles. It would appear that these allozymes private for *C. bursa-pastoris* were also acquired from the most recent common ancestor, postulating that the allozymes concerned had an eastern distribution within the common ancestor's distribution area. Alternatively, they might have been lost in *C. grandiflora* because of bottleneck effects.

Capsella thracica. *Capsella thracica* has been described by Velenovsky (1893) from Bulgaria. It is sometimes given species rank (e.g. Chater 1964) and sometimes treated as a subspecies of *C. bursa-pastoris* (Chater 1993), a view also adopted by Ančev (2007). It is a Bulgarian endemic reported from the Thracian lowlands, Black Sea coast and the Rhodopes Mts. (Ančev 2007). The main feature discriminating this species from *C. bursa-pastoris* is the presence of an elongated style in *C. thracica*. To date, no chromosome numbers have been documented, neither are detailed studies concerning that taxon available. We included *C. thracica* in our studies, and although details of this will be given elsewhere, we report on some of the main features here. *Capsella thracica* is tetraploid as revealed by its genome size (Fig. 1) and shares its cpDNA regions with *C. bursa-pastoris* (Fig. 4). The ITS sequences of the *C. thracica* accessions analysed (Table 2), however, are characterized by two different copies, one from *C. bursa-pastoris* and one from *C. grandiflora/C. rubella* (Fig. 3), indicating a hybrid origin of *C. thracica*. The place of origin of *C. thracica* would appear to be Bulgaria. We argue that the pollen recipient parent species was *C. bursa-pastoris,* as indicated by cpDNA sequences, and the pollen donator was *C. grandiflora* or its progenitor, indicated by the ITS sequences and the length of the style – only *C. grandiflora* and *C. thracica* have an elongated style (Neuffer, unpublished). Interspecific hybridization by fusion of an unreduced diploid *C. grandiflora* (or progenitor) pollen with a normally reduced egg cell of the autotetraploid *C. bursa-pastoris* would lead to the allotetraploid *C. thracica*. Alternatively, an unreduced pollen gamete of *C. grandiflora* (or progenitor) and an unreduced egg cell of hypothesized 'diploid' *C. bursa-pastoris* may have fused.

Evolutionary history of the genus Capsella*, conclusions*

Based on our results and present knowledge, we hypothesize the following scenario outlined in Fig. 6. The genus *Capsella* is of Eurasian origin and comprises two evolutionary lineages, the western lineage (*C. grandiflora*, *C. rubella*) and the eastern lineage (*C. bursa-pastoris*, *C. orientalis*, see Figs 2, 3 and 4). Their common ancestor was diploid and self-incompatible, and its distribution ranged from eastern Europe to western or even central Asia, predominantly confined to Mediterranean and steppe-like climates. Such a continuous steppe belt from central Asia to south-eastern Europe formed, at the latest, at the end of the Pliocene, 2.5–1.6 million years ago (Kamelin 1998; Velichko 1999). Several climatic macrocycles with glacial and interglacial phases during the Pleistocene are associated with latitudinal range shifts of the steppe belt. The steppe belt also

faced significant longitudinal splits during the ice ages (for more detailed discussion, see Franzke *et al.* 2004). These biogeographic dynamics caused geographic and genetic subdivisions within the common ancestor into an eastern and a western lineage, as has also been demonstrated for the Brassicacean Eurasian steppe plant *Clausia aprica* (Franzke *et al.* 2004) and for many other organisms (Hewitt 2001, 2004). The eastern lineage gave rise to *C. bursa-pastoris* and *C. orientalis*, whereas in the western part of the common ancestor's distribution belt, populations gave rise to *C. grandiflora* and *C. rubella*. The current areal of *C. grandiflora* might be regarded as a relict areal. Later, range expansions of *C. bursa-pastoris* to the West led to contact zones with the western lineage species. This facilitated introgression of western lineage genetic material into the eastern genomes (Slotte *et al.* 2006, 2008) on the one side and led to hybrid speciation on the other, giving rise to the allotetraploid species *C. thracica* in Bulgaria (see Fig. 3 and the Discussion chapter). The place of the hybrid zones in Bulgaria, which is the south-western boundary of the Eurasian steppe belt, indicates that *C. grandiflora* or its progenitor once had a wider range than today, which is in line with our hypothesis of a relict areal of *C. grandiflora*. Also, the location of the secondary contact zones in middle and western Europe, as indicated by the introgression and hybridization zones, supports the view that *C. bursa-pastoris* colonized Europe from Asia. A similar scenario has been demonstrated for *Arabidopsis thaliana* (Sharbel *et al.* 2000). The time estimate for the origin of the *Capsella* species is, therefore, compatible with the historical biogeographic events outlined earlier.

The inclusion of the so far 'missing link' species *C. orientalis* and *C. thracica* into our phylogenetic and biogeographic concept will greatly expand the possibilities of using *Capsella* as a model plant genus.

Acknowledgements

The authors wish to thank Ulrike Coja, Claudia Gieshoidt and Rudi Grupe for technical assistance in sequencing, allozyme analyses and cultivation of plants; and Sara Mayland-Quellhorst and Carina Titel for chromosome counting and flow cytometry analyses. We thank Minčo Ančev, Sofia, for help in collecting *Capsella thracica* in Bulgaria. We are thankful to Lucille Schmieding for correcting the English text. Financial support by the Deutsche Forschungsgemeinschaft DFG and by the Deutscher Akademischer Austauschdienst DAAD is greatly acknowledged.

References

Ačarkan A, Roßberg M, Koch M, Schmidt R. (2000) Comparative genome analysis reveals extensive conservation of genome organisation for *Arabidopsis thaliana* and *Capsella rubella*. *Plant Journal*, **23**, 55–62.

Almquist E (1907) Studien über die Capsella bursa pastoris (L.). *Acta Horti Bergiani*, **4**, 1–92. Stockholm.

Almquist E (1921) Studien über die Capsella bursa pastoris II. *Acta Horti Bergiani*, **7**, 41–95. Stockholm.

Al-Shehbaz IA, Beilstein MA, Kellogg EA (2006) Systematics and phylogeny of the Brassicaceae (Cruciferae): an overview. *Plant Systematics and Evolution*, **259**, 89–120.

Ančev M (2007) Catalogue of the family Brassicaceae (Cruciferae) in the flora of Bulgaria. *Phytologia Balcanica*, **13**, 153–178.

Bailey CD, Koch MA, Mayer M *et al.* (2006) Toward a global phylogeny of the Brassicaceae. *Molecular Biology and Evolution*, **23**, 2142–2160.

Beilstein MA, Al-Shehbaz IA, Kellogg EA (2006) Brassicaceae phylogeny and trichome evolution. *American Journal of Botany*, **93**, 607–619.

Beilstein MA, Nagalingum NS, Clements MD, Manchester SR, Mathews S (2010) Dated molecular phylogenies indicate a Miocene origin for *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 18724–18728.

Bleeker W, Weber-Sparenberg C, Hurka H (2002) Chloroplast DNA variation and biogeography in the genus *Rorippa* Scop. (Brassicaceae). *Plant Biology*, **4**, 104–111.

Ceplitis A, Su Y, Lascoux M (2005) Bayesian inference of evolutionary history from chloroplast microsatellites in the cosmopolitan weed *Capsella bursa-pastoris* (Brassicaceae). *Molecular Ecology*, **14**, 4221–4233.

Chater AO (1964) *Capsella Medicus*. In: *Flora Europaea*, Vol. **1**, 1st edn (eds Tutin TG, Heywood VH, Burges NA, Valentine DH, Walters SM and Webb DA), p. 316. Cambridge University Press, Cambridge.

Chater AO (1993) *Capsella Medicus*. In: *Flora Europaea*, Vol. **1**, 2nd edn (eds Tutin TG, Burges NA and Chater AO *et al.*), pp. 381–382. Cambridge University Press, Cambridge.

Coquillat M (1951) Sur les plantes les plus communes a la surface du globe. *Bulletin mensuel de la Société linnéenne de Lyon*, **20**, 165–170.

Couvreur TLP, Franzke A, Al-Schehbaz IA, Bakker FT, Koch MA, Mummenhoff K (2010) Molecular phylogenetics, temporal diversifications, and principles of evolution in the mustard family (Brassicaceae). *Molecular Biology and Evolution*, **27**, 55–71.

Dorofeyev VI (2002) Cruciferae of European Russia. *Turczaninowia*, **5**, 5–114. (In Russian)

Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, **7**, 214.

Ebel AL (2002) New data on distribution of Brassicaceae species in South Siberia and East Kazakhstan. *Turczaninowia*, **5**, 60–68. (In Russian)

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, **39**, 783–791.

Foxe JP, Slotte T, Stahl EA, Neuffer B, Hurka H, Wright SI (2009) Recent speciation associated with the evolution selfing in *Capsella*. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 5241–5245.

Franzke A, Hurka H, Janssen D *et al.* (2004) Molecular signals for late tertiary/early quaternary range splits of an Eurasian

steppe plant: *Clausia aprica* (Brassicaceae). *Molecular Ecology*, **13**, 2789–2795.

Franzke A, German D, Al-Shehbaz IA, Mummenhoff K (2009) *Arabidopsis* family ties: molecular phylogeny and age estimates in Brassicaceae. *Taxon*, **58**, 1–13.

Franzke A, Lysak MA, Al-Shehbaz IA, Koch MA, Mummenhoff K (2011) Cabbage family affairs: the evolutionary history of Brassicaceae. *Trends in Plant Science*, **16**, 108–116.

German DA, Ebel AL (2009) Some interesting records of Cruciferae in Asia. *Animadversiones Systematicae ex Herbario Kryloviano Universitatis Tomskiensis nomive Kyibyschevi*, **101**, 5–11. (In Russian).

German DA, Friesen N, Neuffer B, Al-Shehbaz IA, Hurka H (2009) Contribution to ITS phylogeny of the Brassicaceae, with special reference to some Asian taxa. *Plant Systematics and Evolution*, **283**, 33–56.

Graur D, Martin W (2004) Reading the entrails of chickens: molecular timescales of evolution and the illusion of precision. *Trends in Genetics*, **20**, 80–86.

Guo Y-L, Bechsgaard JS, Slotte T *et al.* (2009) Recent speciation of *Capsella rubella* from *Capsella grandiflora*, associated with loss of self-incompatibility and an extreme bottleneck. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 5246–5251.

Hameister S, Neuffer B, Bleeker W (2009) Genetic differentiation and reproductive isolation of a naturally occurring floral homeotic mutant within a wild-type population of *Capsella bursa-pastoris* (Brassicaceae). *Molecular Ecology*, **18**, 2659–2667.

Hewitt GM (2001) Speciation, hybrid zones and phylogeography – or seeing genes in space and time. *Molecular Ecology*, **10**, 537–549.

Hewitt GM (2004) Genetic consequences of climatic oscillations in the quaternary. *Philosophical transactions of the Royal Society of London. Series B*, **359**, 183–195.

Hurka H, Düring S (1994) Genetic control of plastidic L-glutamate dehydrogenase isozymes in the genus *Capsella* (Brassicaceae). *Heredity*, **72**, 126–131.

Hurka H, Neuffer B (1997) Evolutionary processes in the genus *Capsella* (Brassicaceae). *Plant Systematics and Evolution*, **206**, 295–316.

Hurka H, Freundner S, Brown AHD, Plantholt U (1989) Aspartate aminotransferase isozymes in the genus *Capsella* (Brassicaceae): subcellular location, gene duplication and polymorphism. *Biochemical Genetics*, **27**, 77–90.

Hurka H, Paetsch M, Bleeker W, Neuffer B (2005) Evolution within the Brassicaceae. *Nova Acta Leopoldina NF 92*, **342**, 113–127.

Kamelin RV (1998) *Materials on the History of the Flora of Asia: the Altai Mountain Country*. Altai University Press, Barnaul. (in Russian).

Kay KM, Whittall JB, Hodges SA (2006) A survey of nuclear ribosomal internal transcribed spacer substitution rates across angiosperms: an approximate molecular clock with life history effects. *BMC Evolutionary Biology*, **6**, 36. doi: 10.1186/1471-2148-6-36.

Koch MA, Haubold B, Mitchell-Olds T (2000) Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis*, *Arabis*, and related genera (Brassicaceae). *Molecular Biology and Evolution*, **17**, 1483–1498.

Koch M, Haubold B, Mitchell-Olds T (2001) Molecular systematics of the Brassicaceae: evidence form coding plastidic matK and nuclear Chs sequences. *American Journal of Botany*, **88**, 534–544.

Koch M, Al-Shehbaz IA, Mummenhoff K (2003) Molecular systematics, evolution, and population biology in the mustard family (Brassicaceae). *Annals of the Missouri Botanical Garden*, **90**, 151–171.

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, Dobeš C, Kiefer C, Schmickl R, Klimeš L, Lysak MA (2007) Supernetwork identifies multiple events of plastid *trn*F $_{(GAA)}$ pseudogene evolution in the Brassicaceae. *Molecular Biology and Evolution*, **24**, 63–73.

Krasnoborov IM, Rostovtseva TS, Ligus SA (1980) Chromosome numbers of some plant species of South Siberia and the Far East. *Botaničeskij Žurnal (Moscow & Leningrad)*, **65**, 659–668. (In Russian).

Kress W-J, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH (2005) Use of DNA barcodes to identify flowering plants. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 8369–8374.

Lysak MA, Koch MA, Beaulieu JM, Meister A, Leitch IJ (2009) The dynamic ups and downs of genome size evolution in Brassicaceae. *Molecular Biology and Evolution*, **26**, 85–98.

Mitchell-Olds T (2001) *Arabidopsis thaliana* and its wild relatives: a model system for ecology and evolution. *Trends in Ecology & Evolution*, **16**, 693–700.

Mummenhoff K, Linder P, Friesen N, Bowman JL, Lee J-Y, Franzke A (2004) Molecular evidence for bicontinental hybridogenous genomic constitution in *Lepidium* sensu stricto (Brassicaceae) species from Australia and New Zealand. *American Journal of Botany*, **91**, 254–261.

Neuffer B (1996) RAPD analyses in colonial and ancestral populations of *Capsella bursa-pastoris* (L.) Med. (Brassicaceae). *Biochemical Systematics and Ecology*, **24**, 393–403.

Neuffer B (2011) Native range variation in *Capsella bursa-pastoris* (Brassicaceae) along a 2500 km latitudinal transect. *Flora*, **206**, 107–119.

Neuffer B, Hoffrogge R (2000) Ecotypic and allozyme variation of *Capsella bursa-pastoris* and *C. rubella* (Brassicaceae) along latitude and altitude gradients on the Iberian Peninsula. *Anales Jardín Botánico de Madrid*, **57**, 299–315.

Neuffer B, Hurka H (1999) Colonization history and introduction dynamics of *Capsella bursa-pastoris* (Brassicaceae) in North America: isozymes and quantitative traits. *Molecular Ecology*, **8**, 1667–1681.

Neuffer B, Hirschle S, Jäger S (1999) The colonizing history of *Capsella* in Patagonia (South America) – Molecular and adaptive significance. *Folia Geobotanica*, **34**, 435–450.

Ossowski S, Schneeberger K, Lucas-Lledó JI *et al.* (2010) The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. *Science*, **327**, 92–94. Supporting Online Material.

Oxelman B, Lidén M, Berglund D (1997) Chloroplast *rps16* intron phylogeny of the tribe Sileneae (Caryophyllaceae). *Plant Systematics and Evolution*, **206**, 393–410.

Paetsch M, Mayland-Quellhorst S, Neuffer B (2006) Evolution of the self-incompatibility system in the Brassicaceae:

identification of S-locus receptor kinase (SRK) in self-incompatible *Capsella grandiflora*. *Heridity*, **97**, 283–290.

Paetsch M, Mayland-Quellhorst S, Hurka H, Neuffer B (2010) Evolution of the mating system in the genus *Capsella* (Brassicaceae). In: *Evolution in Action* (ed. Glaubrecht M), pp. 77–100. Springer, Berlin, Heidelberg.

Posada D, Crandall KA (1998) Modeltest: testing the model of DNA substitution. *Bioinformatics*, **14**, 817–818.

Pulquério MJF, Nichols RA (2007) Dates from the molecular clock: how wrong can we be? *Trends in Ecology and Evolution*, **22**, 180–184.

Ramsey J, Schemske DW (2002) Neopolyploidy in flowering plants. *Annual Review of Ecology and Systematics*, **33**, 589–639.

Ronquist R, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572–1574.

Shackleton NJ, Sánchez-Goni MF, Pailler D, Lancelot Y (2003) Marine isotope substage 5e, and the Eemian Interglacial. *Global and Planetary Change*, **36**, 151–155.

Sharbel TF, Haubold B, Mitchell-Olds T (2000) Genetic isolation by distance in *Arabidopsis thaliana*: biogeography and postglacial colonization of Europe. *Molecular Ecology*, **9**, 2109–2118.

Shaw J, Lickey EB, Schilling EE, Small RL (2007) Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. *American Journal of Botany*, **94**, 275–288.

Sherman-Broyles S, Nasrallah JB (2008) Self-incompatibility and evolution of mating systems in Brassicaceae. In: *Self-Incompatibility in Flowering Plants* (ed. Franklin-Tong VE), pp. 123–147. Springer, Berlin, Heidelberg.

Shull GH (1929) Species hybridizations among old and new species of shepherd's purse. *Proc. Int. Cong. Plant Sci.*, **1**, 837–888. Collegiate Press, George Banta Publ. Co.

Sicard A, Stacey N, Hermann K *et al.* (2011) Genetics, evolution, and adaptive significance of the selfing syndrome in the genus *Capsella*. *The Plant Cell*. Available from http://www.plantcell.org/cgi/doi/10.1105/tpc.111.088237.

Slotte T, Ceplitis A, Neuffer B, Hurka H, Lascoux M (2006) Intrageneric phylogeny of *Capsella* (Brassicaceae) and the origin of the tetraploid *C. bursa-pastoris* based on chloroplast and nuclear DNA sequences. *American Journal of Botany*, **93**, 1714–1724.

Slotte T, Huang H, Lascoux M, Ceplitis A (2008) Polyploid speciation did not confer instant reproductive isolation in *Capsella* (Bassicaceae). *Molecular Biology and Evolution*, **25**, 1472–1481.

Soltis DE, Soltis PS, Schemske DW *et al.* (2007) Autopolyploidy in angiosperms: have we grossly underestimated the number of species? *Taxon*, **56**, 13–30.

St. Onge K (2010) *Demography and Polyploidy in Capsella*. Acta Universitatis Upsaliensis, Uppsala. Digital Comprehensive Summaries of Uppsala Dissertations, Faculty of Science and Technology 725.

Swofford DL (2002) *PAUP*: Phylogenetic Analysis using Parsimony (* and Other Methods). Version 4*. Sinauer Associates, Sunderland, Massachusetts.

Taberlet P, Gielly L, Pantou G, Bouvet J (1991) Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology*, **17**, 1105–1109.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, **28**, 2731–2739.

Tang C, Toomajian C, Sherman-Broyles S *et al.* (2007) The evolution of selfing in *Arabidopsis thaliana*. *Science*, **317**, 1070–1072.

Theißen G (2011) The genetics of *Capsella*. In: *Genetics and Genomics of the Brassicaceae* (eds Schmidt R and Bancroft I), pp 373–387. Plant Genetics and Genomics: Crops and Models, **9**. Springer, New York.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The ClustalX window interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*, **25**, 4876–4882.

Turner C (2000) The Eemian interglacial in the North European plain and adjacent areas. *Netherland Journal of Geosciences*, **79**, 217–231.

Velenovsky J (1893) Dritter Nachtrag zur Flora von Bulgarien. Sitzungsberichte der königlichen böhmischen Gesellschaft der Wissenschaften. *Mathematisch-Naturwissenschaftliche Classe*, pp. 11–12.

Velichko AA, ed. (1999) *Climate and Environment Changes during the last 65 Million Years; Cenozoic: from Paleocene to Holocene*. GEOS, Moscow. (in Russian).

Warwick SI, Mummenhoff K, Sauder CA, Koch MA, Al-Shehbaz IA (2010) Closing the gaps: phylogenetic relationships in the Brassicaceae based on DNA sequence data of nuclear ribosomal ITS region. *Plant Systematics and Evolution*, **285**, 209–232.

Welch JJ, Bromham L (2005) Molecular dating when rates vary. *Trends in Ecology and Evolution*, **20**, 320–327.

West RG, Lambert CA, Sparks BW (1964) Interglacial deposits at Ilford, Essex. *Philosophical transactions of the Royal Society of London. Series B, Biological Science*, **247**, 185–212.

Wright SI, Lauga B, Charlesworth D (2002) Rates and patterns of molecular evolution in inbred and outbred *Arabidopsis*. *Molecular Biology and Evolution*, **19**, 1407–1420.

Yoyoka K, Roberts A, Motley J, Lewis R, Brandham P (2000) Nuclear DNA amounts in roses. *Annals of Botany*, **85**, 557–561.

H.H. is especially interested in the evolution of Brassicaceae and in its biogeography with a focus on the Florogenesis of Eurasia. N.F. works on phytogeography of Amaryllidaceae (genera *Allium* and *Galanthus*), Ranunculaceae and Brassicaceae with molecular and cytological methods as well as DNA taxonomy and barcoding. D.G. is interested in taxonomy, systematics, phylogeny and phylogeography of Cruciferae of Asia. A.F.'s research deals with molecular systematics, phylogeny and biogeography of the Brassicaceae. B.N. is working on speciation processes and evolution of the mating system of Brassicaceae.

Data accessibility

1 DNA sequences: Genbank accessions FR773701–FR773711; FR822322–FR822365; HE575225–HE575244 (see Table 2).

2 Final DNA sequence assembly: alignments are provided as supporting information.

3 Sample locations: for *Capsella orientalis* see Table 1, and for the specimens used for DNA sequencing Table 2.

Supporting information

Additional supporting information may be found in the online version of this article.

Appendix S1. ITS sequences.

Appendix S2. cpDNA sequences.

Appendix S3. cp DNA alignment.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.