Molecular Ecology (2005) **14**, 4221–4233 doi: 10.1111/j.1365-294X.2005.02743.x

Bayesian inference of evolutionary history from chloroplast microsatellites in the cosmopolitan weed *Capsella bursa-pastoris* (Brassicaceae)

ALF CEPLITIS, YINGTAO SU* and MARTIN LASCOUX
Department of Conservation Biology and Genetics, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18 D, SE-75236 Uppsala, Sweden

Abstract

Besides showing an extraordinary degree of phenotypic variability, *Capsella bursa-pastoris* (Brassicaceae) is also one of the world's most common plant species and a serious weed in many countries. We have employed a coalescent-based Bayesian analysis of chloroplast microsatellite data to infer demographic and evolutionary parameters of this species. Two different demographic models applied to data from seven chloroplast microsatellite loci among 59 accessions show that the effective population size of *C. bursa-pastoris* is very small indicating a rapid expansion of the species, a result that is in accordance with fossil and historical data. Against this background, analysis of flowering time variation among accessions suggests that ecotypic differentiation in flowering time has occurred recently in the species' history. Finally, our results also indicate that mononucleotide repeat loci in the chloroplast genome can deteriorate in relatively short periods of evolutionary time.

Keywords: Bayesian inference, *Capsella bursa-pastoris*, chloroplast microsatellites, evolutionary history, flowering time, mutation rate

Received 2 June 2005; revision accepted 10 August 2005

Introduction

The cosmopolitan weed *Capsella bursa-pastoris* (shepherd's purse) is one of the world's most common plant species. Found in all continents, it thrives in a wide variety of climates except for the humid tropics (Coquillat 1951; Aksoy *et al*. 1998). Within *C. bursa-pastoris* there is a remarkable degree of variation for several morphological and life history traits (Neuffer & Hurka 1986a, b; Paoletti *et al*. 1991). In particular, there is a pronounced ecotypic differentiation in flowering time, a trait of great importance when colonizing new habitats (Neuffer & Hurka 1986a; Neuffer & Bartelheim 1989; Neuffer 1990). Due to its tremendous variability and colonizing ability, there has been a long-standing interest in *C. bursa-pastoris* as a model for weed ecology and evolution (Baskin & Baskin 1989; Hurka & Neuffer 1991; Neuffer & Hurka 1999; Neuffer & Linde 1999; Hawes *et al*. 2005).

It is generally believed that *C. bursa-pastoris* originated in the Middle East, including the eastern Mediterranean area (Hurka & Neuffer 1997; Aksoy *et al*. 1998). Fossil *Capsella* seeds dated to 5850–5600 BC have been found in Turkey, while the earliest records from central Europe date from 3500 BC (Willerding 1986). It is also known that *C. bursa-pastoris* colonized the Americas in post-Columbian times and reached Australia with European settlers after 1770 (Neuffer & Linde 1999). While fossil and historical records indicate that *C. bursa-pastoris* gained its present-day world-wide distribution rather recently, information from molecular markers on evolutionary or demographic parameters is lacking. Molecular markers have been used in previous studies on *C. bursa-pastoris* to describe possible routes of colonization of North and South America and to identify European source populations (Neuffer & Hurka 1999; Neuffer *et al*. 1999), but these studies have not attempted to estimate, for example, the effective size of the *C. bursa-pastoris* population or the time when the species' expansion began. By providing a background against which the evolution of

Correspondence: Alf Ceplitis, Present address: Department of Cell and Organism Biology, Lund University, Sölvegatan 29, SE-22362 Lund, Sweden. Fax: +46 46 147874; E-mail: alf.ceplitis@cob.lu.se.
*Present address: Department of Plant Biology and Forest Genetics, Swedish University of Agricultural Sciences, Dag Hammarskjölds väg 181, SE-750 07 Uppsala, Sweden

© 2005 Blackwell Publishing Ltd

adaptations can be interpreted, such information is of significant interest for understanding the evolution of invasive species (Lee 2002).

Traditionally, parameter inference in population genetics has relied on summary statistics as, for example, the use of the sample heterozygosity, H, to estimate the scaled mutation rate $\theta = 4N\mu$. However, summary statistics may ignore important information contained in the ancestral relationships in a sample (Felsenstein 1992). With the increasing use of DNA sequences and microsatellite markers in population genetics, likelihood methods based on coalescent theory have been developed to exploit the full information content of such data (Beaumont 1999; Stephens & Donnelly 2000). As a consequence, Bayesian inference, which is a likelihood-based inferential method, is currently enjoying a renaissance in several areas of genetics, including population genetics and evolution (Shoemaker *et al*. 1999; Beaumont & Rannala 2004).

In the Bayesian framework, both the parameters of interest, such as the effective population size or the mutation rate, and the data are treated as random variables. Inferences are then made by conditioning parameter probability distributions on the data, i.e. observations of genetic polymorphisms (e.g. Stephens 2001). Bayesian analytical methods have been particularly popular in studies of recent human history using mitochondrial or Y-chromosome data (reviewed in Beaumont 2004), but have also been used in studies of (other) animal species (Estoup *et al*. 2001; Storz & Beaumont 2002; Estoup & Clegg 2003). By contrast, Bayesian inference of demographic history has very rarely been carried out in plants; indeed, we are aware of only three such studies, all on tree species from the Pinaceae family (Marshall *et al*. 2002; Semerikov & Lascoux 2003; Robledo-Arnuncio *et al*. 2005).

In this study, we use coalescent-based Bayesian analysis of chloroplast microsatellite data obtained from *C. bursa-pastoris* accessions sampled mainly from the European-Mediterranean region to estimate the effective population size, mutation rates, and the time to the most recent common ancestor. We also investigate the spatial distribution of cytoplasmic variation. The results are discussed in relation to the recent evolution of *C. bursa-pastoris* with particular reference to ecotypic variation in flowering time. In addition, some general insights are also gained into the evolution of chloroplast microsatellites.

Materials and methods

Study species

The study species, *Capsella bursa-pastoris* (L.) Medik. (Brassicaceae), is an annual plant species that prefers to grow on open and disturbed ground (Aksoy *et al*. 1998). In many areas of the world, *C. bursa-pastoris* is a common and serious weed (Defelice 2001). The species is predominantly selfing with an estimated outcrossing rate of 1–2% under field conditions (Shull 1929) and 2–10% in a more benign greenhouse environment (Hurka *et al*. 1989). The genus *Capsella* is closely related to *Arabidopsis* (Koch *et al*. 2000; O'Kane & al-Shehbaz 2003), but the taxonomy of the genus itself has been the subject of some dispute. The current view is that *Capsella* comprises three species (Hurka & Neuffer 1997): the two diploids *Capsella rubella* and *Capsella grandiflora* ($2n = 16$) and the tetraploid *C. bursa-pastoris* ($4x = 2n = 32$). Occasionally, diploid forms of *C. bursa-pastoris* have been described (e.g. Svensson 1983), but these are presently treated as *C. rubella* (Neuffer & Eschner 1995).

Plant material

Seeds were collected from 74 naturally occurring *Capsella* plants. Since it is often difficult to distinguish between the tetraploid *C. bursa-pastoris* and the diploid *C. rubella* solely on the basis of morphological criteria, the ploidy level of each accession was determined by flow cytometry analysis carried out by Plant Cytometry Services (Schijndel, the Netherlands). Out of the initial sample, 59 accessions were identified as *C. bursa-pastoris* and thus included in this study. Roughly half of the *C. bursa-pastoris* accessions (33) originated from the European-Mediterranean region, while three originated from East Asia, one from sub-Saharan Africa, two from North America, and one from Australia (Table 1, Fig. 1). No attempt was made to hierarchically sample accessions (i.e. to subdivide the total sample into 'populations'), but at some localities sampling was more intense in order to study genetic variation at a finer scale. Seeds were germinated in Petri dishes and seedlings were transferred to individual pots and grown in the greenhouse as a fully randomized block with a 16-h photoperiod. From each accession, a single seed-derived offspring was chosen for microsatellite analysis. Total DNA was extracted from a small piece of fresh leaf using the DNeasy Plant DNA Kit (QIAGEN). DNA quality and concentration was checked on agarose gels.

Microsatellite analysis

A total of eight chloroplast microsatellite loci were screened in the entire material. Primer sequences for seven of the microsatellite loci (ATCP112, ATCP7905, ATCP28673, ATCP30287, ATCP46615, ATCP66701, and ATCP70189) are given by Provan (2000). An additional microsatellite locus (here denoted ATCP31017) was amplified using the forward primer 5′-GCCTACCGCATCGAAATAGA-3′ and the reverse primer 5′-CAAGAAAGTCGGCCAGAATC-3′. Amplification was performed in a 10-μL reaction volume containing 1× polymerase chain reaction (PCR) buffer

Table 1 Sampling locations of *Capsella bursa-pastoris* accessions. Negative latitudes and longitudes refer to locations south of the equator and west of the Greenwich meridian, respectively. See Materials and methods for measurement of flowering time and Table 3 for explanation of haplotype designations; n.a., measurement not available

Code	Location	Country	Latitude	Longitude	Flowering time	Haplotype
SE12	Härnösand*	Sweden	62.64	17.94	36.0	B
SE17	Härnösand*	Sweden	62.64	17.94	119.0	B
SE38	Uppsala	Sweden	59.87	17.63	87.0	B
SE39	Uppsala	Sweden	59.87	17.63	200	C
SE40	Uppsala	Sweden	59.87	17.63	200	C
SE41	Uppsala	Sweden	59.87	17.63	200	C
SE42	Uppsala	Sweden	59.87	17.63	81.0	A
SE43	Uppsala	Sweden	59.87	17.63	200	B
SE44	Uppsala	Sweden	59.87	17.63	94.5	A
SE45	Uppsala	Sweden	59.87	17.63	90.5	B
SE46	Uppsala	Sweden	59.87	17.63	200	C
SE47	Uppsala	Sweden	59.87	17.63	88.5	A
SE48	Uppsala	Sweden	59.87	17.63	77.0	B
SE37	Västervik	Sweden	57.75	16.63	80.8	B
SE35	Växjö	Sweden	56.88	14.82	78.8	C
SE33	Hässleholm	Sweden	56.15	13.77	78.3	C
SE30	Lund	Sweden	56.67	13.22	97.0	C
SE32	Hjärup	Sweden	55.67	13.12	92.7	C
SE31	Torup	Sweden	55.58	13.20	100.5	B
GB93	Durham†	Great Britain	54.77	–1.57	36.8	G
NL55	Meppen	the Netherlands	52.78	6.70	50.7	A
NL54	Amsterdam	the Netherlands	52.35	4.92	113.2	A
DE97	Osnabrück	Germany	52.27	8.05	90.9	A
DE05	Jena‡	Germany	50.93	11.58	40.9	B
CZ96	Brno§	Czech Republic	49.20	16.63	94.7	A
FR49	Colmar	France	48.08	7.37	119.0	A
FR50	Colmar	France	48.08	7.37	56.6	A
FR51	Ronchamp	France	47.70	6.65	86.3	A
AT101	Trins¶	Austria	47.08	11.42	n.a.	A
GR90	Rhodes**	Greece	36.17	28.00	52.6	H
TR71	Istanbul	Turkey	41.02	28.97	45.5	F
TR73	Istanbul	Turkey	41.02	28.97	44.5	A
TR75	Istanbul	Turkey	41.02	28.97	45.0	B
TR79	Istanbul	Turkey	41.02	28.97	44.0	A
TR83	Istanbul	Turkey	41.02	28.97	43.0	G
TR91	Izmir**	Turkey	38.30	21.32	46.0	D
SY68	Aleppo	Syria	36.20	37.17	42.1	D
SY69	Aleppo	Syria	36.20	37.17	46.3	A
SY70	Aleppo	Syria	36.20	37.17	38.4	G
SY67	Deir-es-Zur	Syria	35.33	40.15	36.2	G
SY66	Palmyra	Syria	34.55	38.38	41.8	G
SY61	Damascus	Syria	33.50	36.30	60.0	A
SY62	Damascus	Syria	33.50	36.30	67.0	G
SY63	Damascus	Syria	33.50	36.30	53.3	G
SY64	Damascus	Syria	33.50	36.30	40.3	G
SY65	Damascus	Syria	33.50	36.30	69.3	G
JO60	Jerash	Jordan	32.28	35.89	48.4	G
JO56	Amman	Jordan	31.97	35.98	36.4	G
JO58	Amman	Jordan	31.97	35.98	47.0	G
JO59	Amman	Jordan	31.97	35.98	37.4	G
AL88	Issoumar††	Algeria	36.77	5.08	42.6	G
AL87	Bainem††	Algeria	35.45	7.96	37.8	E
ET08	Addis Abeba‡‡	Ethiopia	9.03	38.70	43.5	A
US740	Reno§§	USA (California)	39.50	–119.78	91.0	A
US721	Shafter§§	USA (California)	35.42	–119.05	45.0	D

Table 1 *Continued*

Code	Location	Country	Latitude	Longitude	Flowering time	Haplotype
CHCD	Chengdu¶¶	China (Sichuan)	30.67	104.07	62.7	A
TWTY	Taoyuan¶¶	Taiwan	24.27	121.25	34.2	A
TWPL	Puli¶¶	Taiwan	23.97	120.95	33.3	A
AU100	Morven***	Australia (Queensland)	–26.42	147.12	48.0	G

Accessions were collected by the authors, except for the following, which were supplied by: *S. Holm, Mid Sweden University College, Sweden; †C. Anderung, Uppsala University, Sweden; ‡Jena Botanical Garden, Germany; §M. Jaarola, Lund University, Sweden; ¶L. Ghatnekar, Lund University, Sweden; **Tohokou University, Japan; ††A. Harfouche, Arboretum Bainem, Algeria; ‡‡T. Säll, Lund University, Sweden; §§H. Hurka & B. Neuffer, Osnabrück University, Germany; ¶¶Y.-W. Yang, Academia Sinica, Taiwan; ***A.-S. Albrekt, Lund University, Sweden.

(MBI Fermentas), 1.5 mm $MgCl_2$, 0.2 mm dNTP, 0.3 U *Taq* DNA polymerase (MBI Fermentas), 10 pmol of each primer, and 20 ng genomic DNA. All loci were amplified using the following PCR programme: 3 min at 94 °C, 30 cycles of 30 s at 94 °C, 30 s at 55 °C, and 30 s at 72 °C, with a final elongation step of 5 min at 72 °C. Amplification products were resolved on 6% denaturing polyacrylamide gels and visualized by silver staining. Allele sizes were scored by eye and categorized into size classes. Amplification and scoring of microsatellite loci was repeated until all loci had been reliably scored in all accessions. For each locus and allele, two accessions were chosen for sequencing in order to determine the number of mononucleotide repeats. Two replicate samples of each accession were sequenced in opposing directions using the amplification primers as sequencing primers. Sequence reactions were analysed on an ABI 3100 sequencing machine (Applied Biosystems) at the BM unit at Lund University, Sweden. Sequences were edited in the SEQUENCHER software package (Gene Codes Corp.). Sequences have been deposited in GenBank under accession nos DQ144475–DQ144500. The level of variation at each microsatellite locus was quantified by the number of alleles observed and the gene diversity (expected heterozygosity), H, calculated as $H = n(1 - \Sigma_i p_i^2)/(n - 1)$, where n is the sample size and p_i is the relative frequency of allele i (Nei 1978). Each accession was assigned a chloroplast haplotype based on the combined allelic information from the microsatellite loci.

Genetic data analysis

To visualize the phylogenetic relationship among chloroplast haplotypes, a median-joining network (Bandelt *et al*. 1999) was constructed using the NETWORK computer program www.fluxus-engineering.com). In a network of this kind, the number of connections between haplotypes, i.e. the degree of homoplasy in the data set, is influenced by adjusting the value of the parameter ε (Bandelt *et al*. 1999). In the present case, ε was set to zero so that the branches connecting the different haplotypes correspond to the minimum number of mutational differences between haplotypes.

Isolation by distance among the accessions was investigated using Pearson's coefficient of correlation (e.g. Sokal & Rohlf 1995) between the genetic and geographic distance matrices. Genetic distances between accessions were estimated by Goldstein's $\delta\mu^2$ (Goldstein *et al*. 1995). Geographic distances were computed as $D_{i,j} = r[\cos(\mathrm{lat}_i) \cdot \cos(\mathrm{lon}_i) \cdot \cos(\mathrm{lat}_j) \cdot \cos(\mathrm{lon}_j) + \cos(\mathrm{lat}_i) \cdot \sin(\mathrm{lon}_i) \cdot \cos(\mathrm{lat}_j) \cdot \sin(\mathrm{lon}_j) + \sin(\mathrm{lat}_i) \cdot \sin(\mathrm{lat}_j)]$, where $D_{i,j}$ is the geographic distance between accessions i and j, r is the approximate radius of Earth (here 6377 km), and lat_k and lon_k are the latitude and longitude, respectively, of accession k expressed as radians. Distances between accessions with identical coordinates were either measured *in situ* or estimated from maps. The correlation was statistically evaluated by means of a Mantel test (Sokal & Rohlf 1995) using 1000 random permutations.

Inferences on population and mutation parameters were carried out with the BATWING computer program (Wilson *et al*. 2003). This program implements a Markov chain Monte Carlo (MCMC) algorithm to generate approximate random samples from posterior distributions for parameters of interest given the data and some specified prior probability distribution for the relevant parameters, i.e. the effective population size, the mutation rate(s), and the genealogy of the sample. We adopted two different demographic models: one assuming a constant population size and a second assuming a history of population growth; these models are henceforth referred to as the constant size model and the exponential growth model, respectively. The prior distribution for the genealogy is, in the former case, the standard coalescent (e.g. Nordborg 2001) and, in the latter case, the coalescent with exponential population growth at rate α each generation with growth starting at a time T_g prior to the present (Marjoram & Donnelly 1994; Wilson *et al*. 2003).

Under the coalescent, information about N and μ is available only through the product $\theta \propto N\mu$, but if some prior information about N and μ is provided, inference can

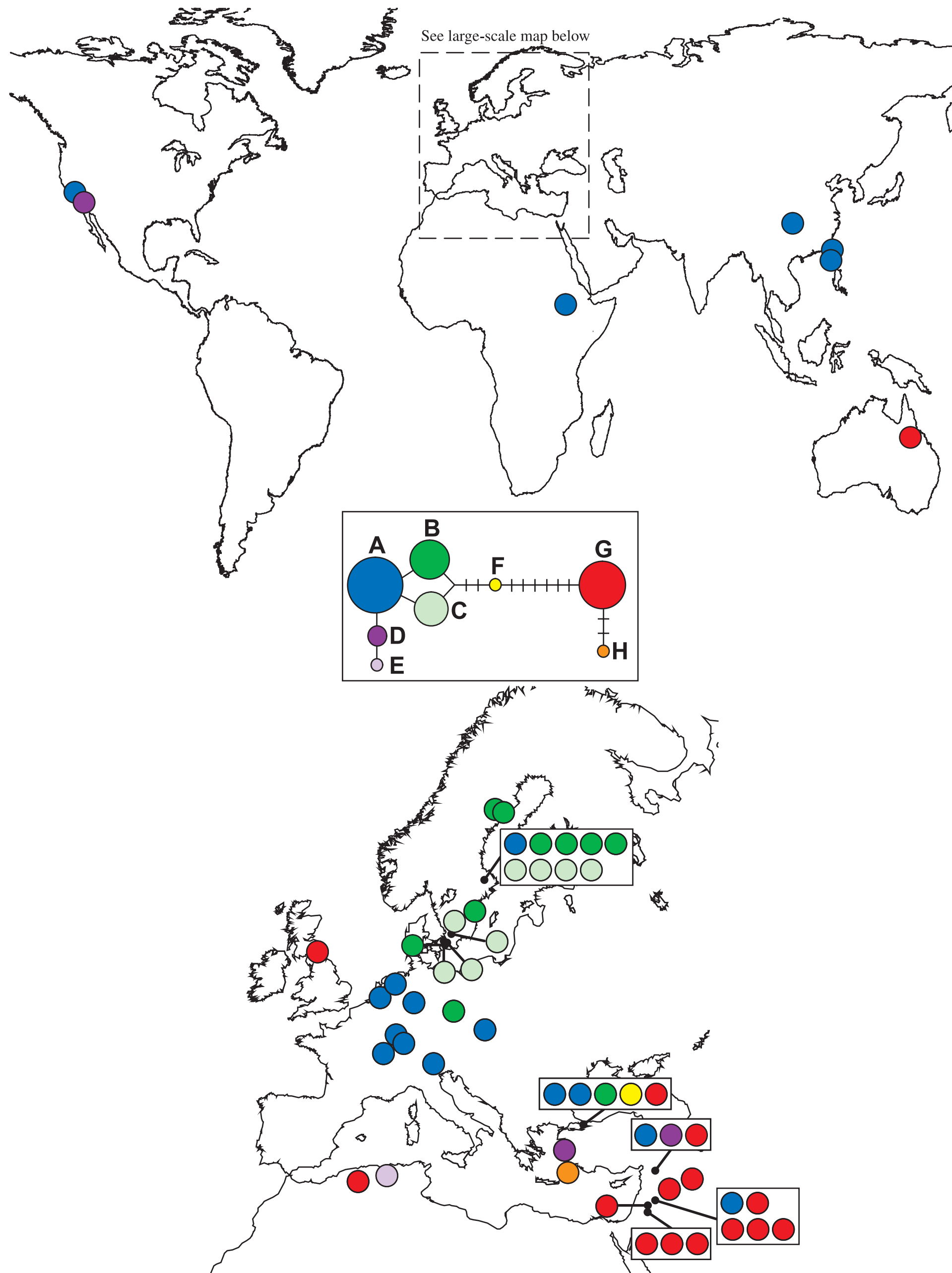

Fig. 1 Geographic distribution of chloroplast haplotypes identified by microsatellite markers among 59 accessions of *Capsella bursa-pastoris*. The insert shows the haplotype network with branch lengths proportional to the number of mutational steps between haplotypes.

be made about N and μ separately (Wilson & Balding 1998). Since we have little prior information about N and μ in the present case, we used diffuse priors (Shoemaker *et al*. 1999) covering most of the plausible values for N and μ. Thus, for the population size (i.e. the effective number of chloroplast genomes) we used an $N \in$ Uniform$[10^3,10^6]$ distribution. For the mutation rate, we specified two groups of microsatellite loci having independent, but identically distributed $\mu \in$ Uniform[0,0.01] mutation rates. The motivation for the separation of loci into two groups with respect to mutation rate is given below. The same set of priors was used for both demographic models; however, it should be noted that under the exponential growth model, N is the ancestral population size (i.e. the size prior to the start of the growth phase). In addition, for the exponential growth model we used an $\alpha \in$ Uniform[0,5] prior for the per-generation growth rate and a $T_g \in$ Uniform[0,40] for the scaled time since population growth began. Finally, we assumed a stepwise-mutation model for the microsatellite loci. Posterior distributions were generated for $\theta = 2N\mu$, N, μ, and

T_{MRCA} (the time in years to the most recent common ancestor of the haplotypes in the sample) for both models as well as for α and T_g (in years) for the exponential growth model. Since both T_{MRCA} and T_g are scaled in units of *N*, we multiplied the product of the posterior values of each parameter and *N* by the generation time to obtain values for T_{MRCA} and T_g in years (Wilson & Balding 1998; Wilson *et al*. 2003). We have assumed one generation per year for *C. bursa-pastoris*. For each model two independent runs of the chain were initiated with different starting conditions. Fifty thousand (50 000) parameter output values were sampled from each chain. Between successive samples there were 100 attempts to update the model parameters and between every such attempt there were 50 attempts to update the genealogy. After discarding the first 5000 values from each run as burn-in, the last 25 000 samples from the two chains were combined to give a total of 50 000 samples for each of the two demographic models (corresponding to 2.5×10^8 and 5×10^6 attempted updates of the genealogical tree and the model parameters, respectively). The median of the posterior distribution was used as a point estimate of the parameter of interest and the 95% highest probability density (HPD) interval — i.e. the shortest interval that contains a specified proportion of the posterior distribution — was computed. The behaviour of the MCMC algorithm was monitored by analysing sample autocorrelations to ensure proper mixing of the chains and by using the Gelman–Rubin statistic (Gelman & Rubin 1992) to check that the output of the chains converged to the desired posterior probability distributions. Density estimation and output diagnostics were performed in the BOA package (Smith 2005) implemented in R (www.r-project.org).

Flowering time determination and analysis

Flowering time, measured as the time from sowing to the appearance of the first flower, was determined on 5 to 10 offspring from each accession. Since *C. bursa-pastoris* is a facultative long-day species (i.e. flowering is inhibited during periods with short daylengths, Hurka *et al*. 1976) we used a 16-h photoperiod to promote flowering. The experiment was terminated after 200 days; plants that did not flower within this time limit were given a flowering time of 200 days. The flowering time of each accession was defined as the average flowering time over all offspring. Flowering time differences between chloroplast haplotypes were analysed by a one-way ANOVA of accession averages. To fulfil the normality requirement, values were arctangent transformed prior to analysis.

Results

Microsatellite polymorphism and haplotype structure

Among the 59 *Capsella bursa-pastoris* accessions included in the study, we found three polymorphic chloroplast microsatellite loci (ATCP66701, ATCP70189, and ATCP31017). Four of the loci (ATCP7905, ATCP28673, ATCP30287, and ATCP46615) were monomorphic, while one locus (ATCP112) failed to amplify altogether and is excluded from further consideration. Sequencing of microsatellite alleles revealed that at three of the monomorphic loci (ATCP7905, ATCP28673, and ATCP30287) the mononucleotide repeat array was interrupted by one or more nucleotide substitutions, as inferred from comparisons with the *Arabidopsis thaliana* plastid reference sequence (Table 2). For the polymorphic loci, we detected three alleles each at locus ATCP66701 and ATCP70189, and five alleles at locus ATCP31017 (Table 2). Gene diversity, *H*, at these loci was 0.067, 0.576, and 0.673, respectively (Table 2), while average *H* over all loci was 0.188. Together, data from the individual loci combined into eight different haplotypes (Table 3).

The haplotype network showed that haplotypes could be roughly categorized into two groups with haplotypes A–F falling into one group and haplotypes G and H in a

Table 2 Repeat motifs at seven chloroplast microsatellite loci in *Arabidopsis thaliana* and at the corresponding loci in *Capsella bursa-pastoris*

	Repeat motif			
Locus	*Arabidopsis thaliana*	*Capsella bursa-pastoris*	No. of alleles	Gene diversity
ATCP7905	$(A)_{13}$	$(A)_3C(A)_8$	1	0
ATCP28673	$(T)_{13}$	$(T)_4G(T)_5CA(T)_2$	1	0
ATCP30287	$(A)_{13}$	$(A)_4C(A)_5$	1	0
ATCP31017	$(T)_{13}$	$(T)_{10}(T)_{11}$ $(T)_{12}$ & $(T)_{17}$	4	0.673
ATCP46615	$(A)_{14}$	$(A)_8$	1	0
ATCP66701	$(T)_{16}$	$(T)_{11}$ $(T)_{12}$ & $(T)_{14}$	3	0.067
ATCP70189	$(A)_{13}$	$(A)_{13}$ $(A)_{16}$ & $(A)_{17}$	3	0.576

Repeat numbers for *A. thaliana* refer to the Columbia ecotype (GenBank Accession no. NC_000932). Repeat numbers for *C. bursa-pastoris*, together with the number of alleles and gene diversity, refer to observations from 59 *C. bursa-pastoris* accessions.

© 2005 Blackwell Publishing Ltd,

Table 3 Chloroplast haplotypes observed among 59 *Capsella bursa-pastoris* accessions and repeat lengths at the three constituent polymorphic microsatellite loci

	Locus				
Haplotype	ATCP66701	ATCP70189	ATCP31017	*n*	FT
A	$(T)_{11}$	$(A)_{17}$	$(T)_{11}$	20	70.3
B	$(T)_{11}$	$(A)_{17}$	$(T)_{10}$	10	87.7
C	$(T)_{11}$	$(A)_{16}$	$(T)_{11}$	8	143.4
D	$(T)_{11}$	$(A)_{17}$	$(T)_{12}$	3	44.4
E	$(T)_{11}$	$(A)_{13}$	$(T)_{10}$	1	37.8
F	$(T)_{12}$	$(A)_{17}$	$(T)_{12}$	1	45.5
G	$(T)_{11}$	$(A)_{13}$	$(T)_{17}$	15	45.9
H	$(T)_{14}$	$(A)_{13}$	$(T)_{17}$	1	52.6

n and FT are the frequency and mean flowering time of each haplotype, respectively.

second group (Fig. 1). Scandinavian and continental European accessions usually had the closely related A, B, or C haplotypes (Fig. 1), whereas haplotype G was most prevalent among the Middle Eastern accessions (Fig. 1). Nevertheless, very different haplotypes could be found within the same geographic area, e.g. haplotypes A and G occurred together in Istanbul (Turkey), Aleppo (Syria), and Damascus (Syria). At the same time, some haplotypes (notably haplotypes A, B, D, and G) were found at localities separated by large distances. Haplotype A was particularly widespread, occurring on four continents. No unique haplotypes were found outside the European-Mediterranean area. No isolation-by-distance effect was detected among the complete set of 59 accessions. The correlation between genetic and geographic distance was positive, but weak and nonsignificant ($r = 0.106$, $Z = 49\,020\,748$, $P = 0.085$). The correlation was, however, much stronger and highly significant among the subset of 52 accessions from the European-Mediterranean region ($r = 0.494$, $Z = 24\,546\,906$, $P < 0.001$).

Inference on population genetic parameters

From the finding of interrupted repeats at three chloroplast microsatellite loci in *C. bursa-pastoris* (Table 2) we decided to allow for mutation rate differences among microsatellite loci when setting mutation rate priors. In a study of variability at 164 chloroplast mononucleotide microsatellite loci among 15 *A. thaliana* accessions, M. Jakobsson *et al.* (unpublished) found that the most important factor in determining the level of polymorphism was the number of uninterrupted repeats, and that loci with less than seven continuous repeats were invariably monomorphic. Thus, in the present study, we defined two categories of loci: those having less than seven uninterrupted repeat units (loci ATCP28673 and ATCP30287) and those having seven or more uninterrupted repeat units (the remaining loci; see Table 2). Consequently, since only loci in the second category were polymorphic (Table 2), posterior probability distributions of $\theta = 2N\mu$ and μ were clearly different for the two categories of loci (Table 4, see also Fig. 2a, b) despite identical priors. Point estimates of θ and μ were an order of magnitude smaller for loci ATCP28673 and ATCP30287 compared to the remaining loci under both demographic models (Table 4). Also, under both models, the lower boundary of the 95% HPD interval for μ was nearly three orders of magnitude smaller for loci in the first category, whereas the upper limit of the interval was in the order of 10^{-4} for both categories of loci (Table 4). The exponential growth model gave lower values for θ and μ than the constant size model, particularly for loci with more than seven uninterrupted repeats (θ_2 and μ_2) which were more variable (Table 4).

Under both demographic models, the posterior distributions for both N and T_{MRCA} were strongly shifted towards smaller values compared with the flat priors (Fig. 2c, d). There were, however, only slight differences in posterior parameter distributions between the two models. For N, the effective population size, our data gave a point estimate of only 1900 under the constant size model and 1700

Table 4 Median and 95% interval boundaries (within parentheses) of the posterior distribution of $\theta = 2N\mu$, μ (mutation rate), N (population size), T_{MRCA} (time in years since most recent common ancestor), T_g (time in years since population growth), and α (the population growth rate per generation), respectively, for the two demographic models

	Parameter							
Model	$\theta_1 \times 10^2$	$\theta_2 \times 10^2$	$\mu_1 \times 10^6$	$\mu_2 \times 10^6$	$N \cdot 10^{-3}$	$T_{MRCA} \cdot 10^{-3}$	$T_g \cdot 10^{-3}$	α
Constant size	6.3	120	13	290	1.9	7.0	—	—
	(0.004, 31)	(54, 200)	(0.007, 92)	(6.2, 700)	(1.0, 13)	(1.2, 48)	—	—
Exponential growth	0.7	9.5	1.7	25	1.7	17	3.2	2.5
	(0.003, 4)	(0.003, 4)	(0.01, 12)	(3.7, 88)	(1.0, 6.2)	(3.9, 60)	(0.71, 11)	(0.01, 4.7)

Parameters θ_1 and μ_1 refer to microsatellite loci ATCP28673 and ATCP30287, while parameters θ_2 and μ_2 refer to the remaining five loci; T_g and α apply only to the exponential growth model. See Materials and methods for a description of the demographic models.

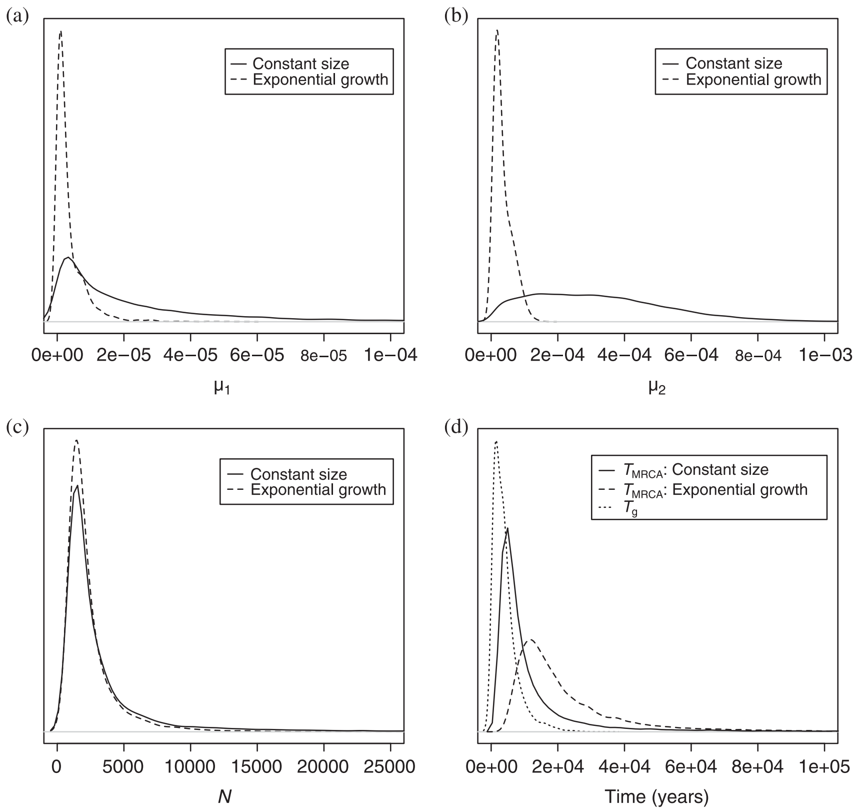

Fig. 2 Posterior distributions for (a) μ_1, the mutation rate for loci ATCP28673 and ATCP30287, (b) μ_2, the mutation rate for the remaining five loci, (c) *N*, the effective population size, and (d) T_{MRCA}, the time in years since the most recent common ancestor, together with T_g, the time since the start of population growth (for the exponential growth model). Filled and dashed lines refer to the constant size model and the exponential growth model, respectively. The dotted line in (d) refers to the exponential growth model. Note the different scales on the *x*-axes in (a) and (b). Prior distributions are flat for all parameters and are not shown. See Materials and methods for further details on models and priors. Medians and 95% HPD intervals of the posterior distributions are given in Table 4.

under the exponential growth model (Table 4). Under both models, the posterior distribution for *N* had a rather narrow 95% HPD interval, supporting values roughly between 10^3 and 10^4 (Table 4). The point estimate of T_{MRCA}, the time to the most recent common ancestor of the sample, was 7000 years under the constant size model and 17 000 years under the exponential growth model. Compared to the posterior distributions for *N* the 95% HPD interval for T_{MRCA} was somewhat wider under both models, giving some support to values up to about 6×10^4 years (Table 4). The exponential growth model gave an estimate of T_g, the time since population growth, of 3200 years with support for values up to 11 000 years (Table 4; Fig. 2d). In contrast to T_g, the posterior distribution for α, the population growth parameter, was flat and thus not obviously different from the prior (Table 4, diagram not shown). Thus, whereas the median point estimate of α was 2.5 (Table 4), no particular value obtained more support than any other.

Examination of the output indicated satisfactory behaviour of the MCMC algorithm. For all parameters, autocorrelation declined to negligible levels at a lag approximately equal to the square root of the sample size. Values of the Gelman–Rubin statistic were ≤ 1.05 for all model parameters indicating proper convergence of the chains (values up to 1.1 are considered acceptable, Gelman *et al*. 1995).

Flowering time variation

The 59 accessions included in this study displayed a wide range of flowering times, with accession averages between 33.3 days for the Taiwanese accession TWTY to over 200 days for some of the accessions from northern Sweden (Table 1). There was a significant difference in flowering time among chloroplast haplotypes ($F = 5.72$, d.f. = 7, 50, $P < 0.001$), with most of the late-flowering accessions belonging to haplotype C and to some extent to haplotypes A and B (Table 3).

Discussion

Levels of polymorphism in the chloroplast genome

Besides being an extremely common and widespread species, *Capsella bursa-pastoris* also displays striking phenoptyic variability. Indeed, so variable is the species that in an early series of studies, Almquist (1907, 1929) identified nearly 200 distinct subtaxa of *C. bursa-pastoris* on the basis of leaf and fruit characters. Later investigations have shown that there is a significant genetic component of variation for many morphological, developmental, and life history traits in *C. bursa-pastoris* (Neuffer & Hurka 1986a, b; Linde *et al*. 2001). Thus, it might be expected that the level of molecular genetic diversity in *C. bursa-pastoris* should be

equally high. However, as shown in the present study, molecular diversity in this species is rather limited, at least in the chloroplast genome. Using seven chloroplast microsatellite loci, we found from one to four alleles per locus and only eight haplotypes in a sample of 59 *C. bursa-pastoris* accessions from five continents (Tables 2 and 3). In comparison, Provan & Campanella (2003), using eight chloroplast microsatellite loci, found between two and seven alleles per locus and 17 haplotypes among 17 accessions of the closely related *Arabidopsis thaliana*, a species that shares many characteristics with *C. bursa-pastoris*, such as a mainly annual, selfing habit and a worldwide distribution. Since we have not carried out a hierarchical sampling of accessions in this study, we have no measure of diversity within and among 'populations'. Nevertheless, even though global chloroplast diversity was restricted, diversity may be rather high locally. For example, four of the eight haplotypes discovered in the global sample could be found in the city of Istanbul, and six haplotypes, including some of the most divergent ones, were found within an area from Istanbul to Rhodes (Fig. 1).

Evolutionary history of C. bursa-pastoris

The level and pattern of genetic diversity at the chloroplast microsatellite loci indicates that the effective population size of *C. bursa-pastoris* is very small (Table 4). Of course, even though we have data from seven microsatellite loci, the lack of recombination in the chloroplast genome means that inference is, in practice, based on a single locus. Sufficiently accurate parameter inference from a nonrecombining genome can be obtained with as few as five linked loci (Wilson & Balding 1998). Nevertheless, extrapolating conclusions, however robust, based on a single locus to encompass the whole genome will generally be highly uncertain, since stochastic variation among genealogies of unlinked loci can be large (e.g. Nordborg 2001). Furthermore, loci with uniparental inheritance often have a lower effective population size than biparentally inherited loci (Birky *et al*. 1983). For these reasons, data from cytoplasmic and nuclear loci might give very different results (see, e.g. Wall & Przeworski 2000). Genetic diversity in the nuclear genome of *C. bursa-pastoris* has been investigated using isozymes and random amplified polymorphic DNA (RAPD) markers (Neuffer 1996; Yang *et al*. 1998; Neuffer & Hurka 1999; Neuffer *et al*. 1999; Neuffer & Hoffrogge 2000), but these studies have been largely descriptive and no attempts have been made at parameter estimation. Despite the fact that we currently lack information from nuclear loci, we nevertheless find it reasonable to regard inferences from chloroplast data in *C. bursa-pastoris* as representative for the genome as a whole on the following grounds: First, the high selfing rate of *C. bursa-pastoris* means that genealogies of different loci — even loci on different chromosomes — will be correlated due to a reduced incidence of recombination (Nordborg 2000). Indeed, this is evident from observations of linkage disequilibrium among enzyme loci located on different linkage groups in *C. bursa-pastoris* (Neuffer & Hurka 1999; Neuffer *et al*. 1999; Linde *et al*. 2001). Second, since *C. bursa-pastoris* also is a hermaphroditic species, the effective population size of the cytoplasmic and nuclear genomes will be (nearly) equal (Birky *et al*. 1983; Pollak 1987). Inferences from chloroplast genome data should thus be more reliable in species like *C. bursa-pastoris*, than, for example, in outcrossing forest trees in which the majority of studies using chloroplast markers have been carried out.

Both the constant size model and the exponential growth model gave a posterior point estimate of the effective population size of less than 2000 (Table 4; Fig. 2a), a number that certainly is several orders of magnitude smaller than the species' current census size. Species that have experienced a recent and rapid population expansion are expected to have a limited effective population size and thus to show little genetic variability (e.g. Nordborg & Innan 2002). For example, studies using Y-chromosome data have demonstrated that the human population has grown rapidly from a small number of individuals (reviewed in Beaumont 2004; see also Wilson *et al*. 2003). In this study, the most recent common ancestor of the chloroplast genome (and, presumably, for reasons given above, for the nuclear genome as well) was estimated to have existed 7000 or 17 000 years ago, depending on the model used (Table 4; Fig. 2c). The latter figure is probably closer to the truth since it is derived from the exponential growth model, which most likely better reflects the demographic history of *C. bursa-pastoris*. This model also suggests that *C. bursa-pastoris* has expanded from an effective population size of 1700 during the last 1000–11 000 years (Table 4; Fig. 2d). Together with information from fossil and historical records (Willerding 1986; Hurka & Neuffer 1997; Neuffer & Linde 1999), these results support the view that *C. bursa-pastoris* originated in the Middle East and spread with the diffusion of agriculture following the last glaciation in Eurasia. The fact that we found a significant isolation-by-distance effect among the European-Mediterranean accessions, but not in the sample as a whole, is probably a result of the very recent establishment of *C. bursa-pastoris* outside the European-Mediterranean area. A similar scenario has been proposed for *A. thaliana*, which is believed to have colonized Europe from Eurasia and the Iberian Peninsula (Sharbel *et al*. 2000). It should be pointed out that a small number of fossil seeds, identified as *C. bursa-pastoris*, from the middle Weichselian and late Wolstonian glaciations (40 000–100 000 years ago) have been found in England (Coope *et al*. 1961; West *et al*. 1964). While it is uncertain whether the taxonomic identity of these seeds can be accurately determined down to species level (H. Hurka,

personal communication), the present-day distributions of the three extant *Capsella* species indicate that the seeds probably belong to *C. bursa-pastoris*. Hence, although the evidence is circumstantial, the species might have a more ancient origin.

Apart from a rapid population expansion, other demographic and evolutionary phenomena can explain the apparently low effective population size of *C. bursa-pastoris*. First, frequent extinctions of local demes followed by repeated recolonization events may also reduce the effective population size (Pannell & Charlesworth 1999; Rousset 2003). Since *C. bursa-pastoris* is a poor competitor and local colonies are probably rather ephemeral, metapopulation dynamics may have a significant effect on the species' population size. Second, a recent selective sweep affecting the whole species or the recurrent purging of deleterious mutations (background selection) may lead to reductions in diversity across the entire genome — including cytoplasmic genomes — in highly selfing species (Maynard Smith & Haigh 1974; Charlesworth *et al*. 1993). For example, Fenster & Ritland (1992) found lower genetic diversity both for allozymes and for chloroplast DNA in the selfing plant *Mimulus micranthus* compared to the mixed-mating *Mimulus guttatus*. Background selection rather than population growth has been cited as responsible for low genomewide levels of polymorphism observed at nuclear microsatellite loci in the widespread and highly selfing nematode *Caenorhabditis elegans* (Sivasundar & Hey 2003). Due to the presumably low rate of recombination in *C. bursa-pastoris*, selection practically anywhere in the genome could be responsible for the species' small effective population size. We currently have no means of evaluating these possibilities since they depend on factors unknown to us, such as the extent of linkage disequilibrium and the distribution of selective effects across the genome in *C. bursa-pastoris* (factors that, it should be added, are unknown for the vast majority of species). However, we believe that the combined genetic, historical, and archaeological evidence support the notion of a recent expansion of *C. bursa-pastoris*.

We found a wide spectrum of flowering times among the 59 accessions included in this study (Table 1). Previous investigations have demonstrated a correlation between time to flowering and environmental characteristics in *C. bursa-pastoris*, indicating that flowering time represents an adaptation to local climatic conditions (Neuffer & Hurka 1986a; Neuffer & Bartelheim 1989; Neuffer 1990). In a study on North American *C. bursa-pastoris* accessions, Neuffer & Hurka (1999) concluded that the occurrence of early- and late-flowering ecotypes in California was the result of multiple introductions of pre-existing ecotypes from Europe rather than the outcome of adaptive evolution *in situ*. It has not been clear, however, if the same is true on a species-wide scale, i.e. whether the observed differentiation in flowering time has taken place recently or represents an ancient polymorphism. If local adaptation maintains distinct flowering-time lineages over evolutionary time, substantial genetic divergence may accumulate between flowering-time ecotypes. For example, using internal transcribed spacer (ITS) sequence data, Gustafsson & Lönn (2003) found large genetic differences between early- and late-flowering individuals of the orchid species *Gymnadenia conopsea* indicating an ancient split between the two variants. By contrast, the results of the present study argue against such a scenario in *C. bursa-pastoris*. First, there was no obvious genetic split between flowering-time ecotypes. We did find a significant difference in flowering time among chloroplast haplotypes, with most of the late-flowering accessions belonging to haplotype C, but the closely related haplotype B included some of the earliest flowering accessions (Table 1). Second, the small effective population size and the apparently recent origin of extant lineages indicate that flowering-time differences have evolved during a rather short period of time. This would, in turn, suggest that flowering time, despite being a continuously varying trait, is controlled by a small number of genes. Indeed, it has been shown that the difference in flowering time between Californian accessions of *C. bursa-pastoris* (accessions US721 and US740 in this study) is governed by only two to three major quantitative trait loci (QTL) (Linde *et al*. 2001; A. Ceplitis *et al*., unpublished). Even if the evidence is not conclusive, these results demonstrate that important adaptations can evolve rapidly even when effective population size is restricted (cf. McKay *et al*. 2001).

Mutation rates and evolution of chloroplast microsatellites

The mutation rate (μ) is a crucial parameter when making inferences about demographic history from population genetic data, since the effective population size (N) cannot be inferred independently from μ. Consequently, with more accurate prior information on μ, more precise estimates can be made of N. There is currently very little knowledge of μ for chloroplast microsatellites as no direct estimates are available. Provan *et al*. (1999) provided the first estimate, as far as we are aware, of chloroplast microsatellite mutation rates. In their study, Provan *et al*. (1999) found no polymorphism in any of 17 chloroplast microsatellite loci in a sample of 64 individuals of the conifer tree species *Pinus torreyana*. Assuming a starlike genealogy they calculated an upper $P = 0.05$ limit for μ around 5×10^{-5}. This figure is one to three orders of magnitude smaller than estimates for nuclear microsatellite loci in plants (Udupa & Baum 2001; Thuillet *et al*. 2002). Later studies have reported estimates of μ in the order of 10^{-3} based on models of constant population size (Marshall *et al*. 2002; Semerikov & Lascoux 2003; Robledo-Arnuncio *et al*. 2005). In the present study, the posterior point estimate

of μ for *C. bursa-pastoris* chloroplast microsatellite loci with more than seven continuous repeat units was 2.9×10^{-4} when assuming a constant population size, and 2.5×10^{-5} assuming exponential growth (Table 4). Interestingly, when making similar assumptions about demography, our estimate of μ comes close to that of Provan *et al*. (1999). Despite the disparity between these estimates, they imply that both the nucleotide substitution rate (Wolfe *et al*. 1987) and the microsatellite mutation rate are lower in the chloroplast genome than in the nuclear genome.

At two of the chloroplast microsatellite loci the repeat array was interrupted by point mutations leaving the length of the longest continuous repeat array less than seven units (Table 2). Neither of these loci showed any variation among the 59 *C. bursa-pastoris* accessions. There are a number of reports in the literature on monomorphism at chloroplast microsatellite loci in which point mutations have occurred within the repeat array. In a study of European *Calluna vulgaris* (heather) populations, Rendell & Ennos (2002) found no variation at any of six chloroplast microsatellite loci among 205 individuals from 25 populations. Sequencing revealed that all of the loci had interrupted repeat arrays whose longest continuous segment was seven units or less. Likewise, three chloroplast microsatellite loci with repeat arrays shorter than six units were monomorphic in a sample of European populations of the common ivy (*Hedera* sp.; Grivet & Petit 2002), and similar findings were made in the common ash (*Fraxinus excelsior*; Heuertz *et al*. 2004). In an extensive survey of 164 mononucleotide repeat loci across the chloroplast genome of *A. thaliana*, M. Jakobsson *et al*. (unpublished) found a threshold of seven repeat units below which loci were invariably monomorphic. Analogous results have been reported for nuclear microsatellite loci in *A. thaliana* (Symonds & Lloyd 2003). These findings clearly indicate that the occurrence of point mutations splitting up repeat arrays into shorter segments slows down the evolutionary rate of microsatellite loci (cf. Rolfsmeier & Lahue 2000). For the present study, we chose seven *A. thaliana* chloroplast microsatellite loci essentially at random, i.e. without prior knowledge of repeat structure or levels of polymorphism in *C. bursa-pastoris*. We found a total of five point mutations at three of the seven loci (Table 2), a number that may seem high considering the close relationship between *Capsella* and *Arabidopsis*. However, a rough calculation shows that the probability of observing five or more substitutions among a total of 100 nucleotide sites (i.e. the approximate sum of the repeat array lengths at the seven microsatellite loci) is actually 0.37 assuming a Poisson-distributed number of substitutions, a substitution rate of 2×10^{-9} per year (cf. Wolfe *et al*. 1987), and a time since the separation of the *Capsella* and *Arabidopsis* lineages of 10 million years (Koch *et al*. 2000, 2001). Thus, chloroplast microsatellite loci can deteriorate in relatively short periods of evolutionary time, a phenomenon which may explain why loci amplified by the currently popular 'universal' chloroplast microsatellite primers seldom display widespread intraspecific variability (Weising & Gardner 1999; Provan *et al*. 2001). Furthermore, since monomorphic loci are informative in likelihood-based analyses, it is clear that care should be taken to determine the molecular background to microsatellite monomorphism (e.g. by sequencing) since demographic inference depends critically on assumptions about mutation rates.

In conclusion, the results of the present study show that *C. bursa-pastoris*, one of the world's most common plant species, is characterized by a very small effective population size, most likely resulting from a recent expansion across the globe. The species' expansion appears to have been accompanied by a rapid evolutionary differentiation of flowering time ecotypes, the genetic basis of which we are presently investigating in more detail. Our study also highlights some of the complexities of chloroplast microsatellite evolution, a topic that certainly merits further study.

Acknowledgements

We thank Barbara Neuffer, Herbert Hurka and Klaus Mummenhoff for stimulating discussions and valuable comments. The manuscript benefited greatly from comments of Marcus Koch and two anonymous reviewers. Support was provided by the Swedish Research Council (VR), the Swedish Research Council for Environment, Agricultural Sciences, and Spatial Planning (FORMAS), the Erik Philip Sörensen Foundation, the Carl Trygger Foundation, and the Nilsson-Ehle Foundation.

References

Aksoy A, Dixon JM, Hale WG (1998) Biological flora of the British Isles. *Journal of Ecology*, **86**, 171–186.

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

Almquist E (1929) Studien über die *Capsella bursa-pastoris* (L.) II. *Acta Horti Bergiani*, **7**, 41–95.

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

Baskin JM, Baskin CC (1989) Germination responses of buried seeds of *Capsella bursa-pastoris* exposed to seasonal temperature changes. *Weed Research*, **29**, 205–212.

Beaumont MA (1999) Detecting population expansion and decline using microsatellites. *Genetics*, **153**, 2013–2029.

Beaumont MA (2004) Recent development in genetic data analysis: what can they tell us about human demographic history? *Heredity*, **92**, 365–379.

Beaumont MA, Rannala B (2004) The Bayesian revolution in genetics. *Nature Reviews Genetics*, **5**, 251–261.

Birky CW Jr, Maruyama T, Fuerst P (1983) An approach to population and evolutionary genetic theory for genes in mitochondria and chloroplasts and some results. *Genetics*, **103**, 513–527.

Charlesworth B, Morgan MT, Charlesworth D (1993) The effect of deleterious mutations on neutral molecular variation. *Genetics*, **134**, 1289–1303.

Coope GR, Shotton FW, Strachan I (1961) A late Pleistocene fauna and flora from Upton Warren, Worcestershire. *Philosophical Transactions of the Royal Society of London. Series B, Biologcal Sciences*, **244**, 379–421.

Coquillat M (1951) Sur les plantes les plus communes de la surface du globe. *Bulletin Mensuel de la Societé Linnéenne Lyon*, **20**, 165–170.

Defelice MS (2001) Shepherd's purse, *Capsella bursa-pastoris* (L.) Medic. *Weed Technology*, **15**, 892–895.

Estoup A, Clegg SM (2003) Bayesian inferences on the recent island colonization history by the bird *Zosterops lateralis*. *Molecular Ecology*, **12**, 657–674.

Estoup A, Wilson IJ, Sullivan C, Cornuet J-M, Moritz C (2001) Inferring population history from microsatellite and enzyme data in serially introduced cane toads, *Bufo marinus*. *Genetics*, **159**, 1671–1687.

Felsenstein J (1992) Estimating effective population size from samples of sequences: a bootstrap Monte Carlo integration method. *Genetical Research*, **60**, 209–220.

Fenster CB, Ritland K (1992) Chloroplast DNA and isozyme diversity in two *Mimulus* species (Scrophulariaceae) with contrasting mating systems. *American Journal of Botany*, **79**, 1440–1447.

Gelman A, Rubin DB (1992) Inference from iterative simulation using multiple sequences. *Statistical Science*, **7**, 457–511.

Gelman A, Carlin JB, Stern HS, Rubin DB (1995) *Bayesian Data Analysis*. Chapman & Hall, London.

Goldstein DB, Ruiz Linares A, Cavalli-Sforza LL, Feldman MW (1995) Genetic absolute dating based on microsatellites and the origin of modern humans. *Proceedings of the National Academy of Sciences, USA*, **92**, 6723–6727.

Grivet D, Petit RJ (2002) Phylogeography of the common ivy (*Hedera* sp.) in Europe: genetic differentiation through space and time. *Molecular Ecology*, **11**, 1351–1362.

Gustafsson S, Lönn M (2003) Genetic differentiation and habitat preference of flowering-time variants within *Gymnadenia conopsea*. *Heredity*, **91**, 284–292.

Hawes C, Begg GS, Squire GR, Iannetta PPM (2005) Individuals as the basic accounting unit in studies of ecosystem function: functional diversity in shepherd's purse, *Capsella*. *Oikos*, **109**, 521–534.

Heuertz M, Fineschi S, Anzidei M *et al*. (2004) Chloroplast DNA variation and postglacial recolonization of common ash (*Fraxinus excelsior* L.) in Europe. *Molecular Ecology*, **13**, 3437–3452.

Hurka H, Neuffer B (1991) Colonising success in plants: genetic variation and phenotypic plasticity in life-history traits in *Capsella bursa-pastoris*. In: *Modern Ecology: Basic and Applied Aspects* (eds Esser G, Overdieck D), pp. 77–96. Elsevier, Amsterdam.

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

Hurka H, Krauss R, Reiner T, Wöhrmann K (1976) Das Blüverhalten von *Capsella bursa-pastoris* (Brassicaceae). *Plant Systematics and Evolution*, **125**, 87–95.

Hurka H, Freudner 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.

Koch M, 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 from coding plastid *matK* and nuclear *Chs* sequences. *American Journal of Botany*, **88**, 534–544.

Lee CE (2002) Evolutionary genetics of invasive species. *Trends in Ecology & Evolution*, **17**, 386–391.

Linde M, Diel S, Neuffer B (2001) Flowering ecotypes of *Capsella bursa-pastoris* (L.) Medik. (Brassicaceae) analysed by a cosegregation of phenotypic characters (QTL) and molecular markers. *Annals of Botany*, **87**, 91–99.

Marjoram P, Donnelly P (1994) Pairwise comparisons of mitochondrial DNA sequences in subdivided populations and implication for early human evolution. *Genetics*, **136**, 673–683.

Marshall HD, Newton C, Ritland K (2002) Chloroplast phylogeography and evolution of highly polymorphic microsatellites in lodgepole pine (*Pinus contorta*). *Theoretical and Applied Genetics*, **104**, 367–378.

Maynard Smith J, Haigh J (1974) The hitch-hiking effect of a favorable gene. *Genetical Research*, **23**, 23–35.

McKay JK, Bishop JG, Lin JZ, Richards JH, Sala A, Mitchell-Olds T (2001) Local adaptation across a climatic gradient despite small effective population size in the rare sapphire rockcress. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **268**, 1715–1721.

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, **89**, 583–590.

Neuffer B (1990) Ecotype differentiation in *Capsella*. *Vegetatio*, **89**, 165–171.

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, Bartelheim S (1989) Gen-ecology of *Capsella bursa-pastoris* from an altitudinal transect in the Alps. *Oecologia*, **81**, 521–527.

Neuffer B, Eschner S (1995) Life-history traits and ploidy levels in the genus *Capsella* (Brassicaceae). *Canadian Journal of Botany*, **73**, 1354–1365.

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 del Jardin Botanico de Madrid*, **57**, 299–315.

Neuffer B, Hurka H (1986a) Variation of development time until flowering in natural populations of *Capsella bursa-pastoris* (Cruciferae). *Plant Systematics and Evolution*, **152**, 277–296.

Neuffer B, Hurka H (1986b) Variation in growth form parameters in *Capsella* (Cruciferae). *Plant Systematics and Evolution*, **153**, 265–279.

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, Linde M (1999) *Capsella bursa-pastoris* — colonisation and adaptation; a globe-trotter conquers the world. In: *Plant Evolution in Man-Made Habitats: Proceedings of the VIIth International IOPB Symposium* (eds van Raamsdonk L, den Nijs H), pp. 49–72. Hugo de Vries Laboratory, University of Amsterdam, Amsterdam.

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

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

Nordborg M (2001) Coalescent theory. In: *Handbook of Statistical Genetics* (eds Balding DJ, Bishop M, Cannings C), pp. 179–212. John Wiley & Sons, Chichester.

Nordborg M, Innan H (2002) Molecular population genetics. *Current Opinion in Plant Biology*, **5**, 69–73.

O'Kane SL, Al-Shehbaz IA (2003) Phylogenetic position and generic limits of *Arabidopsis* (Brassicaceae) based on sequences of nuclear ribosomal DNA. *Annals of the Missouri Botanical Garden*, **90**, 603–612.

Pannell JR, Charlesworth B (1999) Neutral genetic diversity in a metapopulation with recurrent local extinction and recolonization. *Evolution*, **53**, 664–676.

Paoletti C, Pigliucci M, Serafini M (1991) Microenvironmental correlates of phenotypic variation in *Capsella bursa-pastoris* (Cruciferae). *Canadian Journal of Botany–Revue Canadienne de Botanique*, **69**, 1637–1641.

Pollak E (1987) On the theory of partially inbreeding finite populations. I. Partial selfing. *Genetics*, **117**, 353–360.

Provan J (2000) Novel chloroplast microsatellites reveal cytoplasmic variation in *Arabidopsis thaliana*. *Molecular Ecology*, **9**, 2183–2185.

Provan J, Campanella JJ (2003) Patterns of cytoplamic variation in *Arabidopsis thaliana* (Brassicaceae) revealed by polymorphic chloroplast microsatellites. *Systematic Botany*, **28**, 578–583.

Provan J, Soranzo N, Wilson NJ, Goldstein DB, Powell W (1999) A low mutation rate for chloroplast microsatellites. *Genetics*, **153**, 943–947.

Provan J, Powell W, Hollingsworth PM (2001) Chloroplast microsatellites: new tools for studies in plant ecology and evolution. *Trends in Ecology & Evolution*, **16**, 142–147.

Rendell S, Ennos RA (2002) Chloroplast DNA diversity in *Calluna vulgaris* (heather) populations in Europe. *Molecular Ecology*, **11**, 69–78.

Robledo-Arnuncio JJ, Collada C, Alia R, Gil L (2005) Genetic structure of montane isolates of *Pinus sylvestris* L. in a Mediterranean refugial area. *Journal of Biogeography*, **32**, 595–605.

Rolfsmeier ML, Lahue RS (2000) Stabilizing effects of interruptions on trinucleotide repeat expansions in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology*, **20**, 173–180.

Rousset F (2003) Effective size in simple metapopulation models. *Heredity*, **91**, 107–111.

Semerikov VL, Lascoux M (2003) Nuclear and cytoplasmic variation within and between Eurasian *Larix* (Pinaceae) species. *American Journal of Botany*, **90**, 1113–1123.

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.

Shoemaker JS, Painter IS, Weir BS (1999) Bayesian statistics in genetics. *Trends in Genetics*, **15**, 354–358.

Shull GH (1929) Species hybridization among old and new species of Shepherd's purse. *Proceedings of the International Congress of Plant Sciences*, **1**, 837–888.

Sivasundar A, Hey J (2003) Population genetics of *Caenorhabditis elegans*: the paradox of low polymorphism in a widespread species. *Genetics*, **163**, 147–157.

Smith BJ (2005) *Bayesian Output Analysis (BOA) Program, Version 1.1.5 for R*. Available at www.public-health.uiowa.edu/boa/.

Sokal RR, Rohlf FJ (1995) *Biometry: The Principles and Practice of Statistics in Biological Research*. 3rd edn. W.H. Freeman, New York.

Stephens M (2001) Inference under the coalescent. In: *Handbook of Statistical Genetics* (eds Balding DJ, Bishop M, Cannings C), pp. 213–238. John Wiley & Sons, Chichester.

Stephens M, Donnelly P (2000) Inference in molecular population genetics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **62**, 605–635.

Storz JF, Beaumont MA (2002) Testing for genetic evidence of population expansion and contraction: an empirical analysis of microsatellite DNA variation using a hierarchical Bayesian model. *Evolution*, **56**, 154–166.

Svensson S (1983) Chromosome numbers and morphology in the *Capsella bursa-pastoris* complex (Brassicaceae) in Greece. *Willdenowia*, **13**, 267–276.

Symonds VV, Lloyd AM (2003) An analysis of microsatellite loci in *Arabidopsis thaliana*: mutational dynamics and application. *Genetics*, **165**, 1475–1488.

Thuillet AC, Bru D, David J *et al.* (2002) Direct estimation of mutation rate for 10 microsatellite loci in durum wheat, *Triticum turgidum* (L.) Thell. ssp *durum* desf. *Molecular Biology and Evolution*, **19**, 122–125.

Udupa SM, Baum M (2001) High mutation rate and mutational bias at $(TAA)_n$ microsatellite loci in chickpea (*Cicer arietinum* L.). *Molecular Genetics and Genomics*, **265**, 1097–1103.

Wall JD, Przeworski M (2000) When did the human population size start increasing? *Genetics*, **155**, 1865–1874.

Weising K, Gardner RC (1999) A set of conserved PCR primers for the analysis of simple sequence repeat polymorphisms in chloroplast genomes of dicotyledonous angiosperms. *Genome*, **42**, 9–19.

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

Willerding U (1986) *Zur Geschichte der Unkräuter Mitteleuropas*. Karl Wachholtz Verlag, Neumünster.

Wilson IJ, Balding DJ (1998) Genealogical inference from microsatellite data. *Genetics*, **150**, 499–510.

Wilson IJ, Weale ME, Balding DJ (2003) Inference from DNA data: population histories, evolutionary processes and forensic match probabilities. *Journal of the Royal Statistical Society. Series A, (Statistics in Society)*, **166**, 155–201.

Wolfe KH, Li WH, Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences, USA*, **84**, 9054–9058.

Yang YW, Kuo WHJ, Wong TH (1998) Genetic polymorphism of seven populations of *Capsella bursa-pastoris* based on RAPD markers. *Botanical Bulletin of Academia Sinica*, **39**, 17–21.

Alf Ceplitis is currently assistant professor at Lund University. He is a plant population geneticist, primarily interested in ecological genomics and the evolution of reproductive systems. Yingtao Su is a PhD student at the Swedish University of Agricultural Sciences. Martin Lascoux is professor at Uppsala University with a broad interest in theoretical and empirical population genetics.