MOLECULAR ECOLOGY

Molecular Ecology (2012) **21**, 834–850 doi: 10.1111/j.1365-294X.2011.05414.x

Gynodioecy in structured populations: understanding fine-scale sex ratio variation in *Beta vulgaris* ssp. *maritima*

I. DE CAUWER,[1] M. DUFAY, B. HORNOY,[2] A. COURSEAUX and J.-F. ARNAUD

Laboratoire de Génétique et Évolution des Populations Végétales, FRE CNRS 3268, Bâtiment SN2, Université des Sciences et Technologies de Lille – Lille 1, F-59655 Villeneuve d'Ascq Cedex, France

Abstract

Natural selection, random processes and gene flow are known to generate sex ratio variations among sexually polymorphic plant populations. In gynodioecious species, in which hermaphrodites and females coexist, the relative effect of these processes on the maintenance of sex polymorphism is still up for debate. The aim of this study was to document sex ratio and cytonuclear genetic variation at a very local scale in wind-pollinated gynodioecious *Beta vulgaris* ssp. *maritima* and attempt to elucidate which processes explained the observed variation. The study sites were characterized by geographically distinct patches of individuals and appeared to be dynamic entities, with recurrent establishment of distinct haplotypes through independent founder events. Along with substantial variation in sex ratio and unexpectedly low gene flow within study sites, our results showed a high genetic differentiation among a mosaic of genetically distinct demes, with isolation by distance or abrupt genetic discontinuities taking place within a few tens of metres. Overall, random founder events with restricted gene flow could be primary determinants of sex structure, by promoting the clumping of sex-determining genes. Such high levels of sex structure provide a landscape for differential selection acting on sex-determining genes, which could modify the conditions of maintenance of gynodioecy in structured populations.

Keywords: Founder events, gene flow, gynodioecy, pollen and seed dispersal, polymorphic mating systems, spatial genetic structure

Received 14 July 2011; revision received 8 November 2011; accepted 17 November 2011

Introduction

Populations of sexually polymorphic plants species typically vary in sex ratio. In the particular case of gynodioecious plants, in which both hermaphroditic and male-sterile (female) individuals can be observed in natural populations, sex ratio can differ considerably among localities (e.g. Medrano *et al.* 2005; Nilsson & Agren 2006; Dufay *et al.* 2009). It is generally acknowledged that variation in sex ratio among sexually polymorphic plant populations arises from (i) differential selection on the phenotypes in varying environments, (ii) deterministic oscillations because of frequency-dependent selection and/or (iii) nonselective processes, such as random founder events or spatially restricted migration (reviewed in McCauley & Bailey 2009). Understanding the causes of variation in the relative frequencies of the co-occurring sexual phenotypes should thus help gain insight into the mechanisms behind the maintenance of females and hermaphrodites in natural populations.

Although poorly understood in most sexually polymorphic species, the genetic basis of sex determination is well known for several gynodioecious species. It has been shown that it generally involves epistatic interactions between cytoplasmic male sterility genes (CMS)

Correspondence: Isabelle De Cauwer, Fax: +44 (0)1142220002; E-mail: i.decauwer@sheffield.ac.uk

[1]Present address: Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK

[2]Present address:: UMR CNRS 6553 ECOBIO, Université de Rennes 1, France

and nuclear genes that restore male fertility (e.g. Dommée *et al.* 1987; Boutin-Stadler *et al.* 1989; Koelewijn & van Damme 1995; Delph *et al.* 2007). To develop as a female, an individual must carry an unrestored CMS gene. To develop as a hermaphrodite, an individual must either carry a CMS gene in combination with the matching restoration alleles (restored CMS hermaphrodite) or carry a non-CMS cytoplasm (non-CMS hermaphrodite). The knowledge of the genetic basis of sex determination is an essential prerequisite for understanding the mechanisms that are responsible for sex ratio variation.

Several studies have suggested that sex ratio can vary from locality to locality as a local response to natural selection (Vaughton & Ramsey 2004; Nilsson & Agren 2006; Caruso & Case 2007). Typically, female plants are more frequent in sites subject to harsh conditions, probably due to their better seed production under stressful conditions compared to hermaphrodites (reviewed in Ashman 2006). In addition to the effect of natural selection in spatially varying environments, theoretical models also suggest that interactions between nuclear restorers and CMS genes can lead to cyclic variations in population sex ratio, independently of ecological factors (see Gouyon *et al.* 1991; Dufay *et al.* 2007). In this case, a direct relationship between the relative fitness of sexual phenotypes and population sex ratio may be weak and difficult to detect (Dufay *et al.* 2009).

In addition to the effects of natural selection, two nonselective mechanisms can also contribute to variation in sex ratio, by altering local frequencies of CMS genes and restorer alleles. First, it is well acknowledged that genetic drift, generated for instance by random founder events, can cause important variation in sex ratio. After a founder event, local allele frequencies, including frequencies of sex-determining genes, depend on the number of founding seeds colonizing an unoccupied habitat as well as on the origin of these seeds (Wade & McCauley 1988; Whitlock & McCauley 1990). The effects of founder events are expected to be even more important in species that are subject to important extinction/recolonization dynamics (e.g. Manicacci *et al.* 1996), because new populations may not persist long enough to allow substantial changes of founding allele frequencies through gene flow or selection. Founder effects are also expected to increase female frequencies rather than hermaphrodite frequencies, for three reasons. First, because females often produce more seeds than hermaphrodites (reviewed in Shykoff *et al.* 2003), population founders may be more likely to have female mothers. Second, because of dominance effects among nuclear restorer alleles, mating between related individuals or self-fertilization of restored CMS hermaphrodites that are heterozygous at restorer loci are expected to produce some homozygous recessive female offspring, when a new population is composed of few restored CMS hermaphrodites (Emery & McCauley 2002; Bailey & McCauley 2005). Finally, newly established populations are unlikely to contain all restorer alleles for the CMS types that are present in the founders, even if there are several founders coming from diverse sources (Nilsson & Agren 2006).

The second nonselective mechanism that can play a fundamental role in sex ratio variation is dispersal, because it directly acts on the partitioning of genetic diversity within and among structured demes (Loveless & Hamrick 1984; Hamrick & Nason 1996; Ennos 2001). The magnitude and spatial patterns of gene flow through pollen and seed can thus have crucial effects on the distribution of sexes in space. Indeed, dispersal rates among established populations determine the probability of sex-determining genes (CMS genes and restorer alleles) to establish in populations, and theoretical models suggest that pollen and seed dispersal among structured populations can modify the dynamics expected under selection only (Dufay & Pannell 2010).

In most cases, the effects of natural selection, founder events, genetic drift and dispersal on sex ratio variation have been studied at large geographical scales, by comparing populations separated by several kilometres (e.g. McCauley 1998; Medrano *et al.* 2005) or even working at regional scales (e.g. Nilsson & Agren 2006; Caruso & Case 2007; Dufay *et al.* 2009). However, because several studies also document sex ratio variation at very local scales (within populations, e.g. Manicacci *et al.* 1996; Laporte *et al.* 2001; Olson *et al.* 2006), these processes could also act very locally.

The aim of this study was to document sex ratio variation at fine spatial scale in wind-pollinated gynodioecious *Beta vulgaris* ssp. *maritima* and attempt to understand the processes underlying the observed variation. *Beta vulgaris* constitutes a relevant model for understanding the processes responsible for sex ratio variation because the genetic basis of sex determination is well known: male sterility is associated with four particular mitochondrial types, called CMS *E*, *G*, *Svulg* and *H*. These sterilizing cytoplasms coexist with male-fertile cytoplasms, and these different cytotypes can be identified with molecular markers (Cuguen *et al.* 1994; Forcioli *et al.* 1998; Desplanque *et al.* 2000; Fénart *et al.* 2006). When coupling data on genotypes and sexual phenotypes, it is possible to directly measure female frequencies, the frequencies of the distinct sterilizing cytoplasms and the rates of restoration per sterilizing cytoplasm (by recording the proportion of hermaphroditic plants for each sterilizing cytoplasm). This information was used to fulfil the first aim of our study: exploring the occurrence and the magnitude of fine-scale sex structure in two *B. vulgaris* study sites.

Subsequently, this study aimed to enlighten the evolutionary processes responsible for the observed fine-scale sex structure. We chose to focus on the neutral evolutionary processes. Genetic markers are commonly used to determine the effects of genetic drift and gene flow on the structure of genetic diversity within and among populations (for a review, see Ennos 2001). Because the evolutionary processes giving rise to spatial structure for a set of cytoplasmic and nuclear markers are also likely to affect the spatial distribution of cytoplasmic male sterility genes and nuclear restorers, the study of neutral genetic structure will allow us to make inferences on the impact of neutral evolutionary processes—in particular random genetic drift and migration through seed and pollen dispersal—on sex ratio variation. In other words, the characterization of the genetic signatures left by drift and gene flow will be linked back to the distribution of sexual phenotypes in space. Using information on phenotypes, sex-determining genes and a set of neutral markers, we thus investigated to what extent sex structure can be caused by neutral processes such as random genetic drift owing to founder events and spatially restricted migration.

Materials and methods

The species

Wild sea beet, *Beta vulgaris* ssp. *maritima* is a diploid species ($2n = 18$) widely distributed along the western coasts of Europe and around the Mediterranean Sea. It is a short-lived perennial and wind-pollinated species (Letschert 1993). Despite the fact that *Beta vulgaris* is generally considered as a self-incompatible species in the wild, with up to four gametophytic S loci (Owen 1942; Larsen 1977), recent studies on mating system in crop-wild hybrids suggested the occurrence of a self-fertility factor (*Sf* gene) that offsets the self-incompatibility system (Arnaud *et al.* 2010, 2011). The *Sf* gene is a Mendelian self-fertility factor widely used by breeders to produce inbred lines. No data are available on the occurrence of this gene in the wild, but paternity analysis in natural conditions showed very low levels of selfing (De Cauwer *et al.* 2010b), which may be an indication that the *Sf* gene occurs at marginal frequencies in wild populations.

There is no vegetative reproduction, and dispersal thus only occurs through seeds and/or pollen movement. Seeds are aggregated in an irregular, dry body that contains 1–8 seeds and has no particular dispersal mechanism. This study was carried out in Brittany (western France), where sea beets colonize areas located along estuaries, above the upper tide level, on cliffs overhanging the sea and in other coastal habitats (Letschert 1993; Laporte *et al.* 2001; Arnaud *et al.* 2003; Viard *et al.* 2004).

Study sites, sampling and laboratory procedures

Exhaustive sampling was carried out in two study sites located in Brittany (France), within two isolated coves separated by 30 km. The first one, called MOR, is located near Planguenoual (N 48°34′168″; E −2°34′831″), stretches over approximately 300 m and comprised 1098 flowering individuals in 2007. The second one, named PAL, is located near Plouha (N 48°40′497″; E −2°52′911″), stretches over approximately 500 m and comprised 615 flowering individuals in 2007.

Global positioning system coordinates of all adults were recorded using a GARMIN GPS map60CS (accuracy of 2–5 m), allowing us to map the location of all flowering plants within geographical patches and to calculate pairwise geographical distances between individuals. At both sites, individuals were clustered in geographically distinct patches (Fig. 1), with few isolated individuals. Individuals were considered as isolated when they had <10 individuals in a radius of 10 m. As pollen and seed flow are known to be mainly local in this species (De Cauwer *et al.* 2010b), considering these geographical clusters of individuals as a proxy of genetic units for all subsequent statistical analyses may be a good prerequisite: four and five geographical groups of individuals were therefore considered throughout this study for MOR and PAL, respectively (Fig. 1). The cytoplasmic and nuclear genotypic structure, the genetic differentiation among geographical clusters and the search of hidden genetic structure will allow us to validate the biological relevance of this geographical criterion. Within these sites, leaves were collected for genotyping on all individuals that flowered during the study year (i.e. all adult individuals). Additionally, the sexual phenotype was determined (female or hermaphrodite) for almost all individuals (98.5% in MOR and 97.4% in PAL).

DNA was extracted from dried leaves and purified using the NucleoSpin®96 Plant kit (Macherey-Nagel). All sampled individuals were genotyped using three different marker types. First, nine nuclear microsatellite loci chosen for high polymorphism were used to describe the neutral diversity in both sites: *Bmb6* (Cureton *et al.* 2002); *Bvm3* (Mörchen *et al.* 1996); *GTT1*, *CAA1* (Viard *et al.* 2002); *SB04*, *SB06*, *SB07*, *SB15* (Richards *et al.* 2004); and *FDSB1027* (McGrath *et al.* 2007). Second, cytoplasmic polymorphism was characterized using four mitochondrial minisatellite loci: *TR1*, *TR2*, *TR3* and *TR4* (Nishizawa *et al.* 2000). Each genotype combination for these four minisatellite loci will be analysed as a single haplotype, as the entire mitochondrial

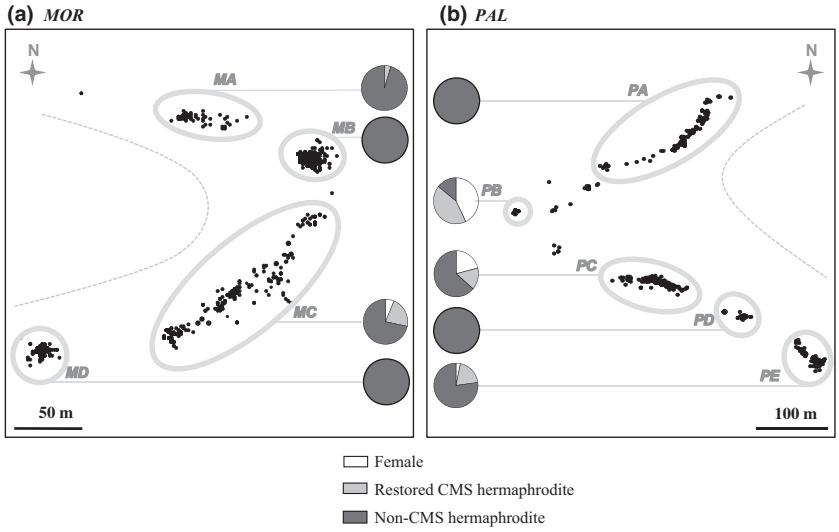

Fig. 1 Spatial distribution of *Beta vulgaris* ssp. *maritima* individuals at the two study sites, MOR (a) and PAL (b), located in Brittany (western France). The locations of the geographical patches (MA, MB, MC and MD in MOR and PA, PB, PC, PD and PE in PAL) are circled in grey, and individuals are represented by black dots. The grey dashed lines show the position of the shore line. The pie charts show the frequency of the three sexual phenotypes (females, restored CMS hermaphrodites and non-CMS hermaphrodites) within the different patches. CMS, cytoplasmic male sterility genes.

genome is generally inherited as a single linkage unit. Finally, diagnostic cytoplasmic PCR-RFLP markers were used to distinguish among the three main CMS types: two mitochondrial markers that correspond to the CMSs *Svulg* and *G* and a chloroplast marker for CMS *E* (Ran & Michaelis 1995; Dufay *et al.* 2008). In this study, we did not attempt to detect the occurrence of the fourth sterilizing cytoplasm (CMS *H*) because this cytotype is known to be absent from Brittany (Cuguen *et al.* 1994). Amplification procedures and detection of polymorphism for these three marker types have been detailed in previous studies (Fénart *et al.* 2007, 2008; Dufay *et al.* 2009; De Cauwer *et al.* 2010b). A total of 1713 samples were genotyped, and missing data rates were 2.3%, 1.6% and 0.6% for nuclear microsatellites, mitochondrial minisatellite haplotypes and cytoplasmic male sterility PCR-RFLP markers, respectively.

Statistical analyses

Nuclear and cytoplasmic diversity. The total number of sampled alleles at each locus for nuclear microsatellites ($A_{N\ Nuc}$) and the number of haplotypes obtained with cytoplasmic minisatellites ($A_{N\ Cyto}$) were counted within both study sites, as well as within each distinct geographical patch of individuals. We used the rarefaction method described in El Mousadik & Petit (1996) to calculate allelic richness ($A_{R\ Nuc}$ and $A_{R\ Cyto}$), which is a measure of the number of alleles (or haplotypes) that is independent of sample size (i.e. patch size in our study, as the sampling was exhaustive). To be able to compare allelic richness, values were standardized to a common sample size ($N = 18$ individuals, i.e. the size of the smallest geographical patch in this study), using FSTAT version 2.9.3 (Goudet 1995). Expected heterozygosities (H_e) as well as intra-population fixation indexes (F_{IS}) were also calculated for nuclear microsatellite loci using FSTAT version 2.9.3 (Goudet 1995). Departure from Hardy–Weinberg equilibrium within each site and each geographical patch was tested by comparing the F_{IS} values for the observed data set with the distribution for a randomized data set obtained after 10 000 permutations of alleles among individuals within geographical patches. Additionally, to compare the level of individual genetic diversity among the three sexual phenotypes (female, restored CMS hermaphrodites and non-CMS hermaphrodites), mean individual heterozygosity (HL) was computed following Aparicio *et al.* (2006) for each individual, using GENHET (an R function developed by Coulon 2009).

Finally, data were tested for linkage disequilibrium for all locus pairs at both sites, as well as within each geographical patch, using exact tests based on a Markov chain method as implemented in Genepop version 4.0 (Raymond & Rousset 1995). Tests were conducted with dememorization number set to 10 000, for 1000 batches and 10 000 iterations. Significance of *P*-values was assessed after Bonferroni correction to eliminate significance by chance (Rice 1989).

Historical estimates of population structure and gene flow. We estimated pairwise population differentiation (F_{ST}) among the two study sites and among the distinct geographical patches within each site with 10 000 permutations of individuals between sites or patches, using a G test for significance of results (Goudet *et al.* 1996). Significance of *P*-values for pairwise F_{ST} values was assessed after Bonferroni correction (Rice 1989). It should be kept in mind that while pairwise estimates of F_{ST} provide some insight into the degree to which populations are historically connected by gene flow, they do not allow to determine whether that genetic

connectedness is a result of ongoing migration or of recent common ancestry (Holsinger & Weir 2009).

We further attempted to compare levels of gene flow through seed and pollen dispersal. Under the assumptions of the Wright's island model, Ennos (1994) demonstrated that, assuming migration–drift equilibrium, a ratio (r) of the amount of pollen migration (m_p) over the amount of seed migration (m_s) can be inferred from F-statistics estimated by both nuclear biparentally inherited markers (F_{STN}) and maternally inherited markers (F_{STC}) from the following equation:

$$r = \frac{m_p}{m_s} = \frac{\left(\frac{1}{F_{STN}} - 1\right)(1 + F_{IS}) - 2\left(\frac{1}{F_{STC}} - 1\right)}{\left(\frac{1}{F_{STC}} - 1\right)}$$

In this equation, F-statistics are computed according to the Weir & Cockerham (1984) procedure and F_{IS} refers to the mean multilocus estimates of F_{IS} calculated over the nine microsatellite loci. Some cautions have to be made when interpreting this indirect estimator of pollen migration over seed migration because (i) it supposes genetic equilibrium conditions under the island model, and this hypothesis is likely to be violated in natural populations, and (ii) the properties of this estimator are not known for gynodioecious populations, where the conditions of equal sex ratios are not respected.

Finally, the hierarchical approach developed by Yang (1998) and implemented in HIERFSTAT (R function developed by Goudet 2005) was used to carry out the estimation of variance components following a three-level hierarchical method: we estimated variance components within individuals, among individuals within haplotypes, among haplotypes within geographical patches and among geographical patches. We thus computed F_{IT} (nonrandom mating within each site), F_{IH} (nonrandom mating within haplotypes), F_{HP} (genetic differentiation among haplotypes within geographical patches) and F_{PT} (genetic differentiation among geographical patches). The effect of haplotype was assessed by permuting individuals among haplotypes within geographical patches (10 000 permutations). The effect of geographical patches was assessed by carrying out 10 000 permutations of haplotypes among geographical patches.

Bayesian estimate of population structure. Hidden population structure may confound estimates of genetic structure using classical F-statistics (Weir 1996). In our study, using geographical patches of individuals as the predefined population units may not necessarily accurately reflect true population structure. To test the assumption that distinct geographical patches represent well-defined genetic demes, we used a Bayesian model-based clustering algorithm to infer population structure and to probabilistically assign individuals to subpopulations within both study sites. We used no prior information on the geographical location in which the individuals were sampled. Pritchard *et al.*'s procedure (2000), implemented in Structure version 2.3.2., was used to cluster individuals in K subpopulations, using the multilocus genotypes of the individuals and minimizing departures from Hardy–Weinberg expectations and linkage disequilibria. A series of 10 independent runs were conducted, with different proposals for K, testing all values from 1 to 15. Each run was conducted assuming population admixture and correlation of allele frequencies (Falush *et al.* 2003) and included 100 000 iterations after a burn-in period of 10 000 iterations. To check for the convergence of the Markov chain Monte Carlo (MCMC), the consistency of results was checked for the 10 replicates performed for each values of K. Finally, the most probable number of clusters (K) was determined using the *ad hoc* statistic ΔK, based on the rate of change in the log probability of data between successive K values, as described in Evanno *et al.* (2005).

Estimates of recent migration rates among geographical patches. Evidence of recent migration events among geographical patches was assessed using a Bayesian multilocus procedure described in Wilson & Rannala (2003). Relative to indirect estimators of long-term gene flow, this method requires few assumptions and can be applied to populations that are far from equilibrium, for both animal and plant species (Friar *et al.* 2007; Munkacsi *et al.* 2008; Polato *et al.* 2010). This method uses individual multilocus genotypes to estimate rates of recent migration (i.e. within the last few generations) among populations, along with the posterior probability distribution of individual immigrant ancestries, population allele frequencies and population inbreeding coefficient. It should be kept in mind that this Bayesian approach does not allow disentangling the relative contributions of pollen and seed movement to gene flow, but has to be interpreted as an overall estimate of migration rate among plant populations. The MCMC was run for 2×10^6 iterations after a burn-in period of 10^6 iterations using BAYESASS (Wilson & Rannala 2003). Samples were collected every 2000 iterations to infer posterior probability distribution of parameters. The two data sets were independently run five times with different random seed values to verify the consistency of the results across runs. To examine the strength of the information in both data sets, 95% confidence intervals (CIs) were computed for migration rates and compared to a scenario where all proposed changes along the Markov chain are accepted (simulating a situation when there is no information in the data set).

Testing for spatial genetic structure within geographical patches. The levels of spatial genetic structure within geographical patches were assessed by performing statistical correlation analyses between a genetic kinship estimate and pairwise geographical distance, for both nuclear data (microsatellites loci) and cytoplasmic data (minisatellite haplotypes) using SPAGeDi version 1.2. (Hardy & Vekemans 2002). Nason's kinship coefficient F_{ij} (Loiselle *et al.* 1995) was chosen as a pairwise estimator of genetic relatedness, as it has robust statistical properties (Vekemans & Hardy 2004). Kinship coefficients (F_{ij}) were regressed on the natural logarithm of geographical distance [ln (d_{ij})], thereby providing the regression slopes b. To visualize spatial genetic structure, F_{ij} values were averaged over a set of distance classes (10 distance classes, defined to obtain approximately the same number of individual pairs within each distance class) and plotted against geographical distances. Significance of the regression slopes and F_{ij} estimates was calculated by 10 000 permutations of individual locations for nuclear and cytoplasmic diversity. Finally, to compare the strength of spatial genetic structure between nuclear and cytoplasmic data, as well as between the different geographical patches, we used the statistic S_p described Vekemans & Hardy (2004), and defined as $S_p = -b/(1 - F_N)$, where F_N is the mean F_{ij} between neighbouring individuals ($d_{ij} < 2m$).

Results

Sex polymorphism and spatial sex ratio variation

In MOR, individuals were found in four geographical patches called MA, MB, MC and MD (mean number of individuals was 273, ranging from $N_{MIN} = 69$ to $N_{MAX} = 635$, see Table 1), with four isolated individuals. In PAL, individuals were clustered in five patches called PA, PB, PC, PD and PE (mean number of individuals was 118, ranging from $N_{MIN} = 18$ to $N_{MAX} = 300$, see Table 1), with 23 isolated individuals (Fig. 1). Overall sex ratio (proportion of females) was 1.6% in MOR and 12.3% in PAL, with important variation in sex ratios among patches (Table 1). At both study sites, we found gynodioecious patches (MC, PB, PC and PE) and patches where there were no females (MA, MB, MD, PA and PD). Among the nine studied geographical patches, the sex ratio (i.e. proportion of females) varied between 0% and 42.9% (see Table 1 and Fig. 1).

Male-sterilizing cytoplasms were detected at moderate frequencies, with 13% of the genotyped individuals carrying a CMS (8.4% and 25.7% in MOR and PAL, see Table 1). The most frequent CMS cytotype was CMS *E*, which was present at both sites and in five of the nine geographical patches. Local (i.e. within patch) CMS *E* frequencies ranged from 0% to 89%, and local restoration rates for this particular CMS gene ranged from 45% to 100%. CMS *G* occurred sporadically, with only 11 individuals found in one geographical patch in the *PAL* site (Fig. 4). Among those 11 individuals carrying CMS *G*, two expressed a hermaphroditic phenotype, suggesting that nuclear restorers of male fertility for that particular CMS gene were present in PAL. The last CMS type, CMS *Svulg*, was absent from our data set. In all gynodioecious patches, the genetic determination of sexual phenotypes was cytonuclear, with the coexistence of different cytotypes (associated with male sterility or not) and some individuals carrying CMS cytotypes expressing a hermaphroditic phenotype (and thus carrying nuclear restorers of male fertility). Among the geographical patches where females were absent, individuals carrying CMS genes were either all restored for male function (MA) or completely absent (MB, MD, PA and PD).

Cytonuclear diversity

Over the whole data set, nuclear microsatellite loci exhibited moderate to high levels of polymorphism, with the number of alleles ranging from 4 (*Gtt1*) to 19 (*Bmb6*) in MOR and from 3 (*Gtt1*) to 21 (*Bmb6*) in PAL. The total number of alleles was 124 (97 in MOR and 99 in PAL), and the mean number of alleles per locus (±SD) was 10.78 (±5.43) in MOR and 11 (±6.10) in PAL. The number of sampled alleles ($A_{N\ Nuc}$), allelic richness ($A_{R\ Nuc}$), expected heterozygosity (H_E) and estimated intra-population fixation index (F_{IS}) per study site and per geographical patch are given in Table 1.

At both study sites, nuclear microsatellites showed an overall significant deficit of heterozygotes compared to Hardy–Weinberg expectations ($F_{IS} = 0.052$ in *MOR* and $F_{IS} = 0.019$ in PAL; $P < 0.001$ in both cases, Table 1). As significant single-locus F_{IS} values were not specifically associated with one locus, heterozygote deficiency could not be attributed to the presence of null alleles (data not shown). When subdividing each study site into its geographical patches, fixation indexes remained significantly positive for all four geographical patches in MOR but for only two geographical patches in PAL (Table 1). These departures from Hardy–Weinberg expectations may be due to the presence of distinct genetic clusters and/or isolation by distance within geographical patches. Similarly, at the level of the study site, an important proportion of the 36 pairs of nuclear loci deviated from linkage equilibrium at $P < 0.05$ after Bonferroni correction (18 in MOR and 14 in PAL). As for fixation indexes, the proportion of locus pairs showing linkage disequilibrium decreased when subdividing the data sets into geographical patches (Fig. 2).

Table 1 Major characteristics of the different geographical patches, occurrence of gynodioecy and measures of genetic diversity on nuclear and cytoplasmic data within the two study sites, MOR (a) and PAL (b), and within the different geographical patches of *Beta vulgaris* ssp. *maritima*

	(a) MOR				
	MA	MB	MC	MD	Overall
Patch description					
N_{TOT}	69	635	281	109	1094 (4)
D (ind/m^2)	0.057	0.843	0.081	0.138	0.529
Male sterility and restoration					
Frequency of CMS genes	0.088	0	0.296	0	0.084
Sex ratio	0	0	0.063	0	0.016
NF/NH_{CMS}/NH_{NCMS}	0/3/62 (4)	0/0/633 (2)	17/60/195 (9)	0/0/106 (3)	17/63/996 (18)
Frequency of restoration	1.000	–	0.779	–	0.788
Multilocus estimates of nuclear diversity					
Mean number of alleles per locus, $A_{N\ Nuc}$	5.444	6.889	9.222	5.111	10.778
Mean allelic richness per locus, $A_{R\ Nuc}$	4.542	4.500	6.153	3.911	5.896
Expected heterozygosity, H_E	0.512	0.562	0.614	0.487	0.544
Fixation index, F_{IS}	0.057**	0.038***	0.091***	0.014*	0.052***
S_p statistic and significance of the slope (b)	0.0086**	0.0133***	0.0420***	0.0041^{NS}	–
Cytoplasmic diversity					
Number of haplotypes, $A_{N\ Cyto}$	3	4	6	1	8
Haplotype richness, $A_{R\ Cyto}$	2.98	2.759	5.199	1	6.271
S_p statistic and significance of the slope (b)	0.4864***	0.1336***	1.1357***	–	–

	(b) PAL					
	PA	PB	PC	PD	PE	Overall
Patch description						
N_{TOT}	146	18	300	33	95	592 (23)
D (ind/m^2)	0.034	0.019	0.176	0.028	0.053	0.109
Male sterility and restoration						
Frequency of CMS genes	0	0.889	0.383	0	0.228	0.257
Sex ratio	0	0.429	0.210	0	0.033	0.123
NF/NH_{CMS}/NH_{NCMS}	0/0/142 (4)	6/6/2 (4)	61/45/185 (9)	0/0/32 (1)	3/18/71 (3)	70/69/432 (21)
Frequency of restoration	–	0.500	0.425	–	0.857	0.496
Multilocus estimates of nuclear diversity						
Mean number of alleles per locus, $A_{N\ Nuc}$	5.889	4.333	8.778	4.667	6.111	11
Mean allelic richness per locus, $A_{R\ Nuc}$	4.231	4.333	5.129	4.114	5.220	5.704
Expected heterozygosity, H_E	0.497	0.566	0.604	0.511	0.589	0.554
Fixation index, F_{IS}	0.072***	−0.090*	0.008**	−0.060*	0.028^{NS}	0.019***
S_p statistic and significance of the slope (b)	0.0266***	0.0158^{NS}	0.0185***	0.0202***	0.0695***	–
Cytoplasmic diversity						
Number of haplotypes, $A_{N\ Cyto}$	3	2	6	1	2	6
Haplotype richness, $A_{R\ Cyto}$	2.187	2	4.535	1	2	4.345
S_p statistic and significance of the slope (b)	0.4318***	0.0640^{NS}	0.0314**	–	2.8462***	–

N_{TOT} is the total number of flowering individuals (with the numbers in parentheses corresponding to isolated individuals, growing outside the geographical patches). D corresponds to the mean density of individuals (ind/m^2). The frequency of CMS genes corresponds to the proportion of individuals carrying a cytoplasmic male sterility, and the sex ratio is the proportion of females. NF, NH_{CMS} and NH_{NCMS} are the number of females, restored CMS hermaphrodites and non-CMS hermaphrodites (with the number between brackets corresponding to nonphenotyped individuals). The frequency of restoration is the proportion of restored hermaphrodites in CMS individuals. Allelic richness values ($A_{R\ Nuc}$ and $A_{R\ Cyto}$) were standardized to a common sample size (N = 18 individuals, i.e. the size of the smallest geographical patch in this study). Significance of multilocus F_{IS} estimates per site and per geographical patch was tested using 10 000 random permutations of alleles among individuals within geographical patches. S_p values estimate the intensity of spatial genetic structure in each geographical patch.
CMS, cytoplasmic male sterility genes; NS, nonsignificant.
*P < 0.05; **P < 0.01; ***P < 0.001.

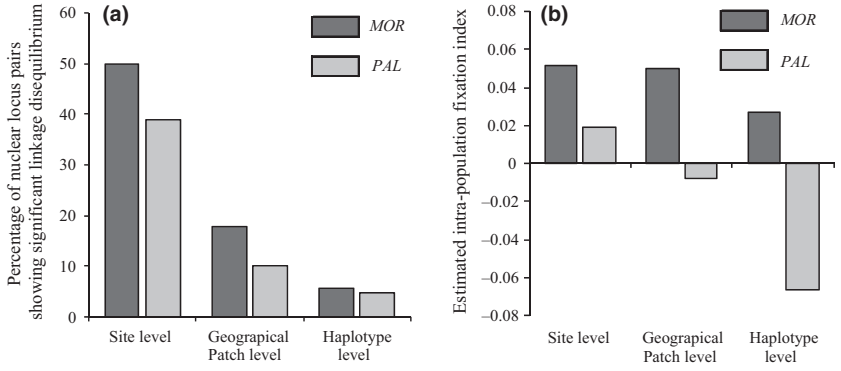

Fig. 2 Linkage disequilibrium and multilocus F_{IS} estimates for different levels of population subdivision. (a) Percentage of nuclear locus pairs showing significant linkage disequilibrium ($P < 0.05$ after Bonferroni correction) and (b) estimated intra-population fixation indices (F_{IS}) at the site level, the geographical patch level (mean values) and the haplotype level (mean values). MOR and PAL refer to the two study sites.

Interestingly, this decrease was even greater when dividing each geographical patch into the distinct haplotypes present within each patch. At the haplotype level, the few pairs of loci that were still in linkage disequilibrium involved different locus combinations.

Mean individual heterozygosity (Mean HL ± SD) was 0.399 (±0.174) in MOR and 0.358 (±0.178) in PAL. Using a general linear model, we simultaneously tested the effects of the study site (two levels, MOR and PAL), geographical patch (nine levels: PA, PB, PC, PD, PE, MA, MB, MC, MD), local density (number of individuals within a radius of 10 m) and sexual phenotype (three levels, female, restored CMS hermaphrodite and non-CMS hermaphrodite) on the level of individual heterozygosity using PROC GLM in SAS (version 9.1, SAS Institute Inc., Cary, NC, USA). Study site, geographical patch and phenotype all had a significant effect (Table 2). *Post hoc* Tukey pairwise comparisons showed that non-CMS hermaphrodites had lower individual heterozygosity compared to females and restored CMS hermaphrodites ($P < 0.05$). Females and restored CMS hermaphrodites showed similar levels of individual heterozygosity ($P = 0.6869$).

Table 2 Results of the general linear model testing simultaneously for the effects of study site (MOR and PAL), geographical patch (MA to MD and PA to PE), local density (number of individuals within a radius of 10 m) and sexual phenotype (female, restored CMS hermaphrodite or non-CMS hermaphrodite) on the level of individual heterozygosity (HL) in *Beta vulgaris* ssp. *maritima*

Source of variation	d.f.	MS	*F*	*P*
Population	1	0.133	4.58	0.0325
Geographical patch (population)	7	0.182	6.25	<0.0001
Density	1	0.006	0.19	0.6630
Sexual phenotype	2	0.308	10.57	<0.0001
Error	1610	46.839		

All the interactions between main factors were nonsignificant ($P > 0.1$) and were dropped from the analysis.
CMS, cytoplasmic male sterility genes.

In contrast to nuclear loci, cytoplasmic mitochondrial minisatellites displayed low levels of polymorphism and yielded a total of 10 distinct haplotypes of which eight were present in MOR and six were present in PAL. The number of sampled haplotypes ($A_{N\ Cyto}$) and the haplotypic richness ($A_{R\ Cyto}$) per geographical patch of individuals is given in Table 1. As expected, each male-sterilizing cytoplasm detected in our study sites (CMS *E* and CMS *G*) was exclusively associated with a unique minisatellite haplotype. A striking result was the spatial clumping of sex-determining cytoplasms, as shown in the Fig. 4, where the two minisatellite haplotypes associated with CMS *E* and CMS *G* were clearly not distributed at random in space. We observed a positive association between nuclear allelic richness ($A_{R\ Nuc}$) and haplotypic richness ($A_{R\ Cyto}$) found within geographical patches (Spearman's Rho = 0.75; $P_{\text{two-tailed}} = 0.002$).

Patterns of genetic differentiation

Strong spatial differentiation between geographical patches was found using mitochondrial minisatellite markers (overall $F_{STC} = 0.606$ in *MOR* and 0.429 in *PAL*; $P < 0.001$ in both cases, see Table 3). The spatial genetic structure was less pronounced for the nine nuclear microsatellites but remained highly significant (overall $F_{STN} = 0.107$ in MOR and 0.082 in PAL; $P < 0.001$ in both cases, see Table 3). Within sites, pairwise F_{ST} among patches were all significant and ranged from 0.154 to 0.914 for cytoplasmic data and from 0.039 to 0.217 for nuclear data (Table 3). The levels of differentiation calculated between the two study sites were lower than overall intra-site estimations ($F_{STC} = 0.278$ and $F_{STN} = 0.068$).

Based on these contrasted levels of genetic differentiation between cytoplasmic and nuclear markers, the *r*-value (estimating the ratio between pollen and seed

Table 3 Genetic differentiation (F_{ST}) estimated for all the pairs of geographical patches of *Beta vulgaris* ssp. *maritima*, using nuclear data (F_{STN}, upper half of the matrices) and cytoplasmic data (F_{STC}, lower half of the matrices), as well as overall F_{ST} estimates within the two study sites (MOR and PAL)

	MOR					PAL				
	MA	MB	MC	MD		PA	PB	PC	PD	PE
MA	–	0.072***	0.112***	0.217***	PA	–	0.116***	0.090***	0.137***	0.122***
MB	0.683***	–	0.076***	0.180***	PB	0.842***	–	0.039***	0.163***	0.060***
MC	0.348***	0.547***	–	0.130***	PC	0.338***	0.219**	–	0.091***	0.043***
MD	0.832***	0.817***	0.356***	–	PD	0.884***	0.914***	0.373***	–	0.073***
					PE	0.762***	0.554***	0.229***	0.154*	–
Overall F_{STN}					Overall F_{STN}					
	0.107***					0.082***				
Overall F_{STC}					Overall F_{STC}					
	0.606***					0.429***				

Significance of genetic differentiation was tested with 10 000 random permutations of individuals between geographical patches, using a G test for significance of results (Goudet *et al.* 1996).
NS, nonsignificant.
*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

gene flow) was equal to 11.5 in MOR and 6.6 in PAL. Under the assumptions of Wright's island model at migration–drift equilibrium, this result suggests that gene flow occurs predominantly through pollen dispersal at both study sites.

Finally, multilocus hierarchical *F*-statistics showed evidence for hierarchical structuring of nuclear genetic variation within both study sites. In MOR, F_{HP} (differentiation among haplotypes within geographical patches) was 0.0472 and F_{PT} (differentiation among geographical patches) was 0.0808 (both at $P < 0.001$). In PAL, F_{HP} was 0.0285 and F_{PT} was 0.0666 (both at $P < 0.001$). Nuclear genetic differentiation thus involves both a geographical component and a cytoplasmic component.

Bayesian analysis of population structure

The results of the Bayesian analysis suggest a hierarchical structure within both study sites, as the K vs. ΔK distribution was multimodal, with one mode at $K = 2$ for both data sets and another mode at $K = 5$ in PAL and at $K = 6$ in MOR (Fig. 3). The obtained results appeared to be geographically meaningful, as the inferred genetic clusters corresponded very closely to the existing geographical patches (Fig. 4). When $K = 2$ at MOR, the algorithm clustered individuals growing in the northern part of the site in the first group (MA and MB) and individuals growing in the southern part of the site in the second group (MC and MD). Neighbouring geographical patches were thus clustered together. Similarly, for $K = 2$ at PAL, one group was comprised of individuals growing in the southern part of the study site (PB, PC, PD and PE) and the other one included individuals located in the northern part of the site (PA). The second clustering solution ($K = 5$ in PAL and $K = 6$ in MOR) also showed geographically consistent results (Fig. 4). Some of the predefined geographical patches of individuals corresponded to homogeneous genetic clusters (MA, MD, PA and PD). In contrast, a mosaic structure was found within MC and PE, with contiguous groups of genetically distinct individuals. Finally, individuals growing in MB and PC were mainly admixed individuals, with a trend for clinal variation at the latter. Besides, one salient feature was the clumping of individual carrying CMS-associated minisatellite haplotypes with neighbouring non-CMS individuals (Fig. 4), indicating a close genetic affinity and suggesting spatially restricted pollen exchanges. Including spatial information in the prior distribution on individual admixture coefficients, as suggested by Durand *et al.* (2009), did not improve the resolution of Bayesian assignment and yielded very similar results, which strengthened the biological relevance of the depicted geographical partitioning.

Estimates of recent migration rates among geographical patches

Recent migration rates (i.e. within the last few generations) were estimated among the distinct geographical patches within both study sites using Wilson & Rannala's method (2003). CIs obtained from both data sets were considerably smaller than those obtained from the null hypothesis, suggesting that the data sets contained a sufficient amount of information to support the results. Within all geographical patches, more than 90%

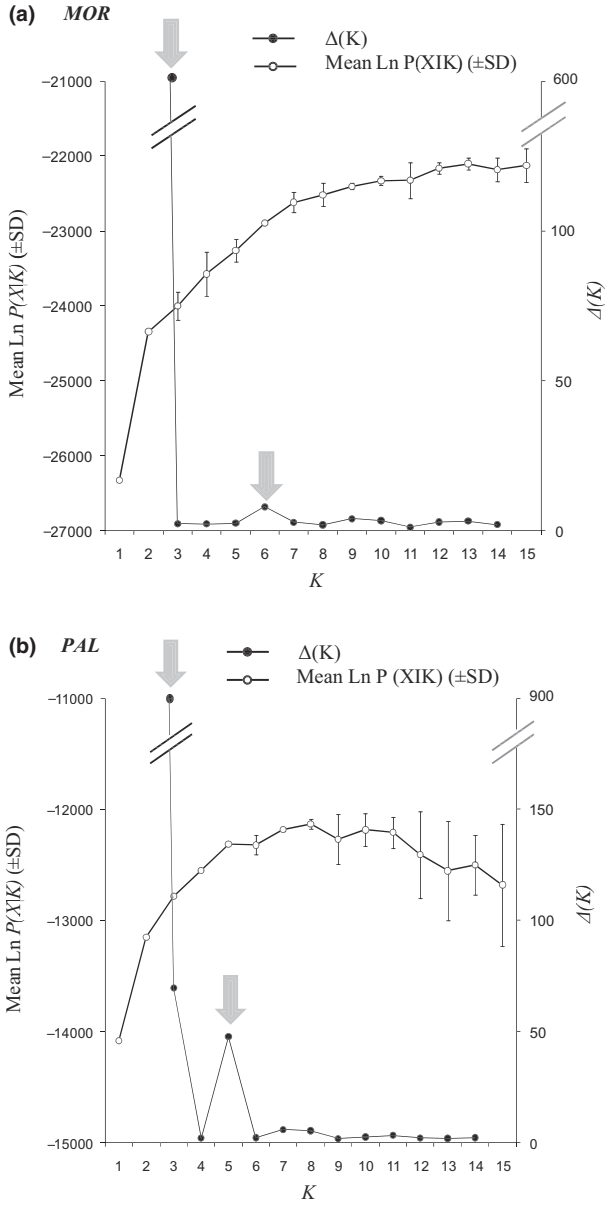

Fig. 3 Estimated number of populations assessed with the clustering method described in Pritchard *et al.* (2000), with mean (±SD) probabilities of the data ln $P(X|K)$ over 10 replicated runs plotted against the putative number of clusters K (ranking from $K = 1$ to $K = 15$ clusters) and standardized second-order rate of change of ln $P(X|K)$, ΔK, plotted against the putative number of clusters K, for the two study sites, MOR (a) and PAL (b).

of individuals were identified as originating from their own source patch (nonmigrants individuals), except for PB and PE, where slightly more important levels of incoming gene flow were detected (Table 4). This finding could be related to the relatively high occurrence of females within these two patches, which allow more incoming pollen flow events. PAL showed higher migration rates between patches compared to MOR (Table 4). We also detected asymmetrical migration rates for the pairs MA–MB, PD–PE and PB–PC. However, in the latter patch-pair, given the large overlap in CIs, it was not possible to conclude as to the statistical significance of the observed asymmetry in gene flow. Interestingly, those pairs corresponded to neighbouring geographical patches, suggesting that dispersal events occurred preferentially between adjacent patches, but not necessarily in strict symmetry.

Testing for isolation by distance

Several general observations can be made from the spatial genetic structure observed within the geographical patches. First, as shown by correlograms (Fig. 5), our results suggest a significant decline in genetic similarity with geographical distance regardless of the type of genetic marker (cytoplasmic or nuclear) within all geographical patches, except MD (nuclear data) and PB (nuclear and cytoplasmic data, see S_P statistics in Table 1). These two patches were also the smallest: the maximum distance between individuals is 35 m in MD and 10 m in PB, and these spatial scales are probably too small for the establishment of isolation by distance patterns. Second, there was considerable difference in the extent of spatial genetic structure between cytoplasmic and nuclear genetic variation. The S_P statistic for nuclear markers was on average 20-fold lower than for cytoplasmic haplotypes. These results confirmed the fact gene flow occurred predominantly through pollen dispersal within both study sites.

Discussion

The development of sex structure in gynodioecious populations

Sex ratio in a population is the combined result of CMS frequencies and restoration rate for each CMS type. In this study, we found that both CMS frequency and restoration rate varied greatly from one geographical patch to another (0–88.9% of individuals carrying a CMS gene and 42.5–100% restoration in CMS individuals). As a result, the sex ratio also showed pronounced variation among geographical patches (0–42.9% females). Although the spatial structure of sexual phenotypes has rarely been studied at fine geographical scales, our results confirm that local-scale sex ratio structure is a common feature in gynodioecious species (e.g. Manicacci *et al.* 1996; Laporte *et al.* 2001; Olson *et al.* 2006). Such striking spatial variation can be explained by the effects of natural selection, founder events, dispersal and genetic drift.

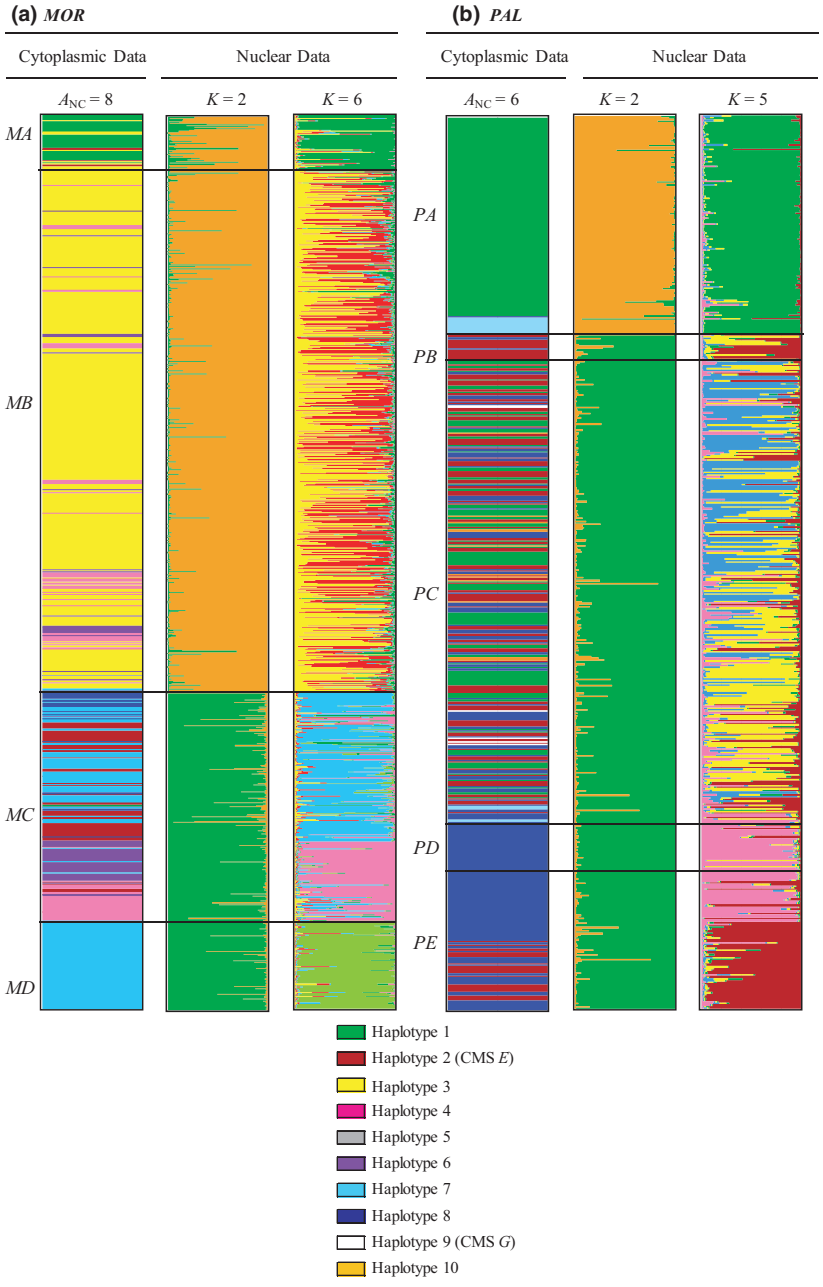

Fig. 4 Distribution of the different haplotypes in the two study sites, MOR (a) and PAL (b), and Bayesian assignment probabilities of membership based on nuclear multilocus genotypes for each individual in the inferred clusters, for the two modal *K* values (*K* = 2 and *K* = 6 in MOR and *K* = 2 and *K* = 5 in PAL). Each individual is represented by a thin horizontal line. For cytoplasmic data, each colour corresponds to a particular haplotype. A_{NC} is the number of haplotypes found in each site. For nuclear data, each individual is partitioned into *K* coloured segments that represent the individual's estimated membership fractions in *K* clusters. Black lines separate the different predefined geographical patches, labelled on the left. Individuals are sorted by increasing latitude.

Sex structure can arise as a result of recurrent selection within relatively persistent and well-established populations. This particular source of sex ratio variation has been extensively documented by studies conducted at a large spatial scale and showing relationships between female frequency and one or several environmental factor(s) (e.g. light, Van Etten & Chang 2009; soil moisture, Barr 2004; drought and temperature, Caruso & Case 2007; temperature and precipitation, Alonso *et al.* 2007). Although less obvious, the same processes can also take place at very local scale, among microsites within populations (Case & Barrett 2001; Van Etten & Chang 2009). Understanding whether and to what extent the sex structure reported here is attributable to differential selection acting on the sexual phenotypes within different micro-environments was not the aim of the current study, but constitutes an interesting perspective that would require a detailed characterization of the environment. Besides the action of selection, non-selective evolutionary forces are theoretically suspected to have a strong impact on the distribution of sexual phenotypes in space, but have received less empirical evidences. Analysing the relationship between spatial patterns in the distribution of neutral genetic diversity

Table 4 Mean (95% CI) posterior distribution for contemporary migration rates among *Beta vulgaris* ssp. *maritima* geographical patches within the two study sites (MOR and PAL)

	MOR			
	MA	MB	MC	MD
MA	**90.835 (86.386–94.878)**	8.409 (4.431–12.834)	0.445 (0.004–1.889)	0.311 (0.002–1.313)
MB	0.064 (0–0.218)	**99.883 (99.604–99.997)**	0.027 (0–0.141)	0.026 (0–0.125)
MC	0.308 (0.002–1.125)	0.127 (0.001–0.571)	**99.222 (97.931–99.938)**	0.342 (0.003–1.211)
MD	0.203 (0.001–0.989)	0.273 (0.001–1.185)	0.351 (0.004–1.353)	**99.172 (97.717–99.919)**

	PAL				
	PA	PB	PC	PD	PE
PA	**99.485 (98.538–99.941)**	0.130 (0.001–0.603)	0.145 (0–0.707)	0.126 (0–0.611)	0.113 (0–0.574)
PB	2.768 (0.013–10.067)	**75.777 (67.510–90.464)**	14.178 (2.598–24.080)	1.145 (0.004–5.257)	6.132 (0.044–19.592)
PC	0.192 (0.002–0.713)	5.561 (3.683–11.078)	**92.004 (86.142–94.792)**	2.037 (0.694–3.853)	0.204 (0.001–0.850)
PD	0.557 (0.001–2.769)	1.515 (0.123–5.264)	2.240 (0.139–6.385)	**94.915 (89.003–98.776)**	0.774 (0.001–3.735)
PE	0.381 (0.001–1.597)	0.283 (0.001–1.347)	0.397 (0.001–1.846)	13.214 (9.550–17.241)	**85.725 (81.673–89.445)**

Values along the diagonal (bold) are the percentage of individuals derived from the source patch (i.e. nonmigrants). Migration occurs from the geographical patch listed at the top of the column into geographical patches listed on the left.

and variation in local sex structure thus helped us to partly understand why sex structure can be as pronounced even at very local scale.

The mode of colonization has long been recognized as a major factor influencing the maintenance and the loss of genetic variation (Slatkin 1977; Wade & McCauley 1988). In our study sites, the high levels of neutral genetic structure suggested that founder events, along with restricted gene flow, might be responsible for the observed sex structure. Indeed, *Beta vulgaris* is generally patchily distributed along the sea shore, where seasonal storms cause frequent disturbance, sometimes leading to the extinction of local patches. It should be kept in mind that, while snapshot genetic data are useful to describe genetic structure, the processes underlying this structure can be diverse and potentially difficult to disentangle. Nonetheless, in our study, the occurrence of a spatial mosaic of distinct haplotypes, the highly significant genetic differentiation levels among geographical patches and haplotypes (as measured by pairwise F_{ST} estimates), and the results of assignment tests all showed that geographical groups of plants corresponded very closely to genetically distinct groups that have probably originated from different founder events.

Interestingly, when comparing our results on patches separated by a few hundred metres or less with the results of a previous study performed at regional scale (33 sites sampled over several hundreds of kilometres, see Fievet *et al.* 2007), we found similar levels of nuclear genetic differentiation (0.107 in MOR, 0.082 in PAL and 0.089 at the regional scale). Regarding cytoplasmic data, the levels of differentiation were even higher at local scale than at regional scale (0.606 in MOR, 0.429 in PAL and 0.278 at the regional scale). This could be interpreted as an effect of metapopulation dynamics: if seeds colonizing a new site originate from several source sites or if there are multiple independent colonization events, a site, taken as a whole, can show high levels of genetic diversity (Slatkin 1977; Wade & McCauley 1988; Whitlock & McCauley 1990). Conversely, at smaller scales (within sites), genetic drift will tend to randomly fix different alleles in different local patches, leading to an increase in the genetic differentiation among local patches within sites, which could ultimately generate fine-scale sex ratio variation in the case of gynodioecious species. However, the consequences of such dynamics on sex structure largely depend on the degree of persistence of local patches, which is not directly quantified in our study. With low extinction rates, natural selection can substantially modify the initial allele frequencies within local patches, whereas strong extinction rates can render the effects of natural selection ephemeral.

Gene flow through seed and pollen dispersal

The levels of genetic structure depend not only on the number and the origin of seeds colonizing an unoccupied patch but also on the magnitude of migration between patches (Whitlock & McCauley 1990). Our results showed evidence of very restricted gene flow within both study sites. Indeed, Bayesian estimates of

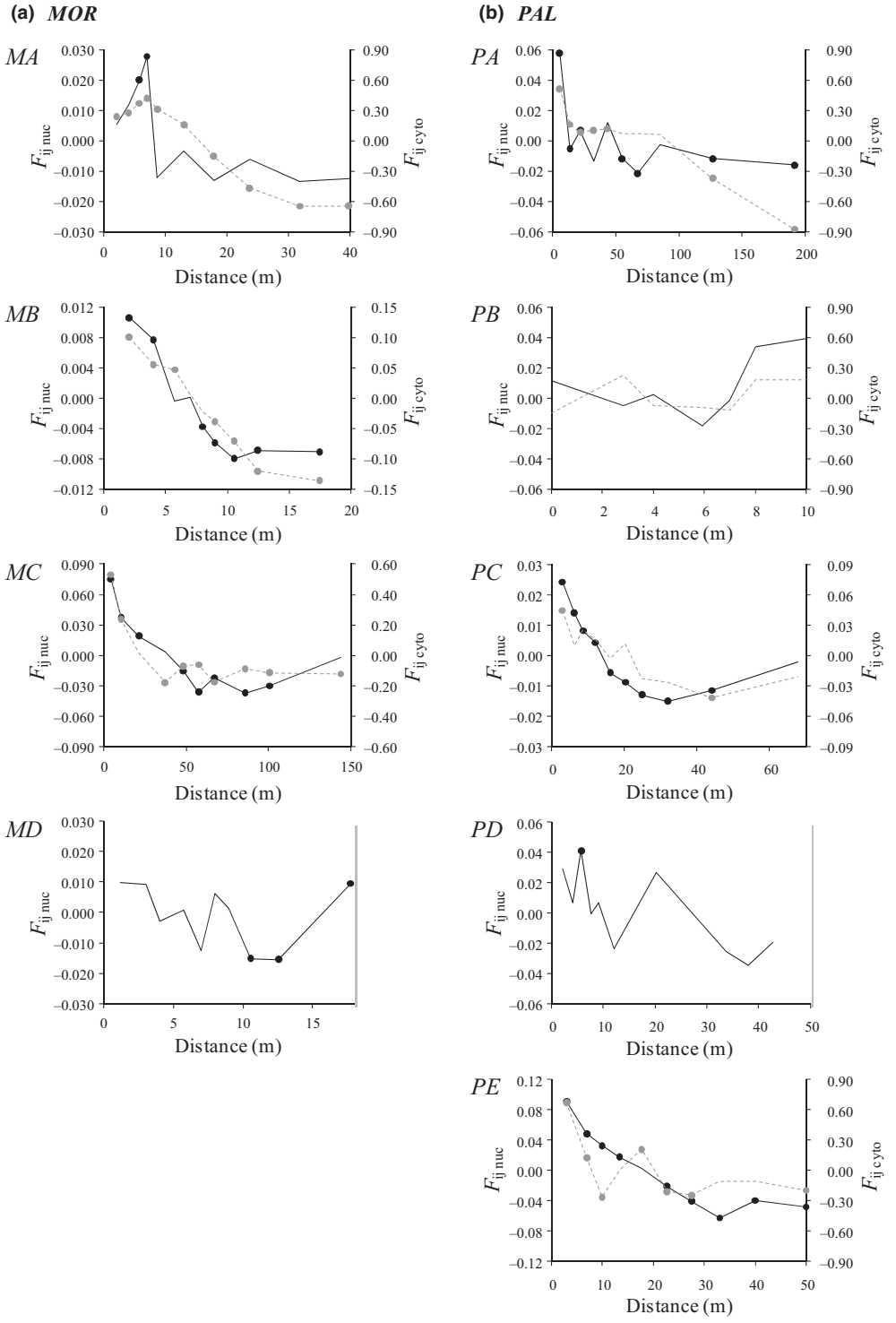

Fig. 5 Average pairwise kinship coefficient (F_{ij}, Loiselle *et al.* 1995) between individuals for nuclear microsatellites (black lines) and cytoplasmic haplotypes (grey dashed lines) in the two study sites, MOR (a) and PAL (b), plotted against the geographical distance within each predefined geographical patch. The significance of F_{ij} values was assessed by permutations of individuals over geographical locations: dots indicate significant F_{ij} values ($P < 0.05$ or less). MD and PD were monomorphic for cytoplasmic data.

contemporary gene flow (integrating both seed and pollen migration) between geographical patches were exceptionally low considering the spatial scale at which our study was conducted. Additionally, patterns of spatial genetic structure within each geographical patch, synthetically quantified by S_p statistics, suggested that

isolation by distance or abrupt genetic discontinuities can take place within a few tens of metres. Even if *B. vulgaris* is a wind-pollinated and self-incompatible plant, features that are likely to decrease genetic variance among demes (Loveless & Hamrick 1984; Vekemans & Hardy 2004), the patterns of gene flow within our study sites seemed limited enough to maintain a strong genetic structure probably initially generated by founder events. In addition, the spatial distribution of pollination events can also have direct consequences on the sex ratio in the next generations. If gene flow between individuals is mainly local, as in our study sites, inbreeding (keeping in mind that our species is largely self-incompatible, but see Arnaud *et al.* 2010, 2011) or biparental inbreeding may enhance the production of females within progenies (Emery & McCauley 2002; Bailey & McCauley 2005). However, this hypothesis was not verified in the study sites, as suggested by the sex ratios observed in progeny analyses (IDC, personal observation).

Additionally, our results showed strong differences in the extent of spatial genetic structure between nuclear and cytoplasmic loci, with the genetic structure at maternally inherited cytoplasmic markers being stronger than what was observed for nuclear loci, suggesting a predominance of pollen migration over seed migration within the two study sites. This was quantified through the estimates of the *r*-ratio (ratio of the amount of pollen migration over the amount of seed migration inferred from *F*-statistics) that was high within both study sites, as compared to other studies reporting wide-range cytonuclear spatial genetic structure (reviewed in Petit *et al.* 2005). It should be kept in mind that this estimate is based on assumptions of equal sex ratio and island model of population structure. Nonetheless, the S_P statistics—that do not rely on equilibrium assumptions—yielded consistent results and were also on average 20-fold lower for nuclear markers than for cytoplasmic haplotypes. The relative magnitude of gene flow through seed and pollen dispersal should be taken into account in theoretical models, as both processes are expected to have different consequences on the maintenance of cytonuclear gynodioecy. For instance, Dufay & Pannell (2010) theoretically showed that, in a subdivided population, while gynodioecy is systematically lost under drift alone, seed or pollen dispersal could maintain cytonuclear gynodioecy. More precisely, seed dispersal was shown to promote the maintenance of cytonuclear polymorphism at the level of the whole population as well as at the level of local demes, while pollen dispersal could not counter the loss of cytonuclear polymorphism at the level of the local deme, but promoted its maintenance at the level of the subdivided population taken as a whole.

In plant species that are often subject to catastrophic forms of disturbance, long-distance seed dispersal is essential for a metapopulation to persist (Cain *et al.* 2000). In our study, in addition to the fact that geographical patches showed significant levels of differentiation, our results also suggested some hierarchical substructuring owing to the presence of different haplotypes within geographical patches. Along with the observation of decreasing values of genetic linkage disequilibrium and F_{IS} estimators when dividing geographical patches according to haplotype identities, this strongly suggests that our study sites are dynamic entities, with recurrent and independent establishment of new haplotypes through seed migration, as previously suggested in other gynodioecious species (Olson & McCauley 2002). The fact that this signal was detected within both study sites suggests that (i) the arrival of different haplotypes is quite recent and/or (ii) reproduction may commonly involve individuals sharing the same haplotype, probably as a consequence of the patchiness of haplotypes within geographical patches and of spatially restricted pollen flow. Although seed migration is low compared to pollen migration in *B. vulgaris*, geographical patches where several haplotypes were found are probably the result of independent seed immigration events. As several seeds can be contained in one fruit, the opportunity for kin-structured foundation by groups of sibs is enhanced (e.g. Torimaru *et al.* 2007). In addition to the partitioning of cytonuclear diversity within geographical patches when several different haplotypes coexist, the recurrent arrival of new haplotypes also seems to increase the level of nuclear allelic richness within geographical patches (as shown by the significant correlation between $A_{R\ Nuc}$ and $A_{R\ Cyto}$).

Finally, this particular functioning could also affect individual heterozygosity of CMS and non-CMS individuals differently. Female and restored CMS hermaphrodites showed significantly higher levels of individual heterozygosity compared to non-CMS hermaphrodites in both study sites. This may be a consequence of the fact that CMS and non-CMS seeds do not have the same mating opportunities when they become established in a patch: (i) when non-CMS seeds establish in a patch, they produce only hermaphrodites, and given the spatially limited pollen dispersal, a large part of mating events probably occurs between relatives, generating some biparental inbreeding; (ii) when CMS seeds establish in a patch, they produce either females or restored hermaphrodites, which have been shown to produce lower pollen quality than non-CMS individuals (Dufay *et al.* 2008; De Cauwer *et al.* 2011) and, as a consequence, a larger proportion of mating events must involve unrelated individuals, which were already occupying the patch.

Perspectives: the effects of sex structure on individual fitness

Understanding variation in local sex ratio is important because local sex ratio can in turn affect the mating success of individual plants. In sexually polymorphic species, the reproductive success of a given sexual phenotype often depends on its own frequency (McCauley & Bailey 2009). The pronounced spatial genetic structure and sex structure documented here could thus have important consequences for the dynamics of both CMS genes and restorer alleles. For instance, local sex ratio can affect female fitness of individuals, by decreasing seed production in female-biased patches through pollen limitation (e.g. Widen & Widen 1990; McCauley & Brock 1998; Graff 1999; De Cauwer *et al*. 2010a). Although less explored, the same process could impact male fitness. In *B. vulgaris*, restorer alleles have been found to not always perfectly restore pollen production (Dufay *et al*. 2008; De Cauwer *et al*. 2011). However, owing to spatial genetic structure for cytoplasms, restored hermaphrodites are more likely to be clumped with female than non-CMS hermaphrodites, which modifies their access to female plants and, ultimately, increases their male fitness compared to a situation with no genetic structure (De Cauwer *et al*. 2010b). An ongoing study of male siring success aims at investigating the variation in transmission success of restorer alleles according to the genetic properties of the patch of plants in which they occur.

Acknowledgements

We are grateful to C. Engel-Gautier, E. Petit and C. Lexer and two anonymous referees for very helpful comments on an earlier draft of the manuscript. We also thank N. Faure for her precious help in the field. This work was funded by a grant from the 'Agence Nationale de la Recherche' (ANR-06-JCJC-0074). Isabelle De Cauwer was supported by a CNRS/Région Nord-Pas-de-Calais fellowship. Jean-François Arnaud and Mathilde Dufay are grateful to the CNRS for supporting them as a full-time research scientist during the 2009–2010 and 2006–2007 academic years, respectively.

References

Alonso C, Mutikainen P, Herrera CM (2007) Ecological context of breeding system variation: sex, size and pollination in a (predominantly) gynodioecious shrub. *Annals of Botany*, **100**, 1547–1556.

Aparicio JM, Ortego J, Cordero PJ (2006) What should we weigh to estimate heterozygosity, alleles or loci? *Molecular Ecology*, **15**, 4659–4665.

Arnaud J-F, Viard F, Delescluse M, Cuguen J (2003) Evidence for gene flow via seed dispersal from crop to wild relatives in *Beta vulgaris* (Chenopodiaceae): consequences for the release of genetically modified crop species with weedy lineages. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **270**, 1565–1571.

Arnaud J-F, Fénart S, Cordellier M, Cuguen J (2010) Populations of weedy crop-wild hybrid beets show contrasting variation in mating system and population genetic structure. *Evolutionary Applications*, **3**, 305–318.

Arnaud J-F, Cuguen J, Fénart S (2011) Metapopulation structure and fine-scaled genetic structuring in crop-wild hybrid weed beets. *Heredity*, **107**, 395–404.

Ashman TL (2006) The evolution of separate sexes: a focus on the ecological context. In: *The Ecology and Evolution of Flowers* (eds Barrett SCH and Harder LD), pp. 419–465. Oxford University Press, Oxford.

Bailey MF, McCauley DE (2005) Offspring sex-ratio under inbreeding and outbreeding in a gynodioecious plant. *Evolution*, **59**, 287–295.

Barr CM (2004) Soil moisture and sex ratio in a plant with nuclear-cytoplasmic sex inheritance. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **271**, 1935–1939.

Boutin-Stadler V, Saumitou-Laprade P, Valero M, Jean R, Vernet P (1989) Spatio-temporal variation of male sterile frequencies in two natural populations of *Beta maritima*. *Heredity*, **63**, 395–400.

Cain ML, Milligan BG, Strand AE (2000) Long-distance seed dispersal in plant populations. *American Journal of Botany*, **87**, 1217–1227.

Caruso CM, Case AL (2007) Sex ratio variation in gynodioecious *Lobelia siphilitica*: effects of population size and geographic location. *Journal of Evolutionary Biology*, **20**, 1396–1405.

Case AL, Barrett SCH (2001) Ecological differentiation of combined and separate sexes of *Wurmbea dioica* (Colchicaceae). *Ecology*, **82**, 2601–2616.

Coulon A (2009) GENHET: an easy-to-use R function to estimate individual heterozygosity. *Molecular Ecology Ressources*, **10**, 167–169.

Cuguen J, Wattier R, Saumitou-Laprade P *et al*. (1994) Gynodioecy and mitochondrial DNA polymorphism in natural populations of *Beta vulgaris* ssp. *maritima*. *Genetics, Selection, Evolution*, **26**, S87–S101.

Cureton AN, Burns MJ, Ford-Lloyd BV, Newbury HJ (2002) Development of simple sequence repeat (SSR) markers for the assessment of gene flow between sea beet (*Beta vulgaris* ssp. *maritima*) populations. *Molecular Ecology Notes*, **2**, 402–403.

De Cauwer I, Arnaud J-F, Schmitt E, Dufay M (2010a) Pollen limitation of female reproductive success at fine spatial scale in a gynodioecious and wind-pollinated species, *Beta vulgaris* ssp. *maritima*. *Journal of Evolutionary Biology*, **23**, 2636–2647.

De Cauwer I, Dufay M, Cuguen J, Arnaud J-F (2010b) Effects of fine-scale genetic structure on male mating success in gynodioecious *Beta vulgaris* ssp. *maritima*. *Molecular Ecology*, **19**, 1540–1558.

De Cauwer I, Arnaud J-F, Courseaux A, Dufay M (2011) Gender-specific fitness variation in gynodioecious *Beta vulgaris* ssp. *maritima*: do empirical observations fit theoretical predictions? *Journal of Evolutionary Biology*, **24**, 2456–2472.

Delph LF, Touzet P, Bailey MF (2007) Merging theory and mechanism in studies of gynodioecy. *Trends in Ecology & Evolution*, **22**, 17–24.

Desplanque B, Viard F, Bernard J *et al.* (2000) The linkage disequilibrium between chloroplast DNA and mitochondrial DNA haplotypes in *Beta vulgaris* ssp. *maritima* (L.): the usefulness of both genomes for population genetic studies. *Molecular Ecology*, **9**, 141–154.

Dommée B, Assouad MW, Valdeyron G (1987) Natural selection and gynodioecy in *Thymus vulgaris*. *Biological Journal of the Linnean Society*, **77**, 17–28.

Dufay M, Pannell JR (2010) The effect of pollen versus seed flow on the maintenance of nuclear-cytoplasmic gynodioecy. *Evolution*, **64**, 772–784.

Dufay M, Touzet P, Maurice S, Cuguen J (2007) Modelling the maintenance of a male fertile cytoplasm in a gynodioecious population. *Heredity*, **99**, 349–356.

Dufay M, Vaudey V, De Cauwer I *et al.* (2008) Variation in pollen production and pollen viability in natural populations of gynodioecious *Beta vulgaris* spp. *maritima*: evidence for a cost of restoration? *Journal of Evolutionary Biology*, **21**, 202–212.

Dufay M, Cuguen J, Arnaud J-F, Touzet P (2009) Sex ratio variation among gynodioecious populations of sea beet: can it be explained by negative frequency-dependent selection? *Evolution*, **63**, 1483–1497.

Durand J, Jay F, Gaggiotti OE, François O (2009) Spatial inferences of admixture proportions and secondary contact zones. *Molecular Biology and Evolution*, **26**, 1963–1973.

El Mousadik A, Petit RJ (1996) High level of genetic differentiation for allelic richness among population of the argan tree [*Argania spinosa* (L.) Skeels] endemic to Morocco. *Theoretical and Applied Genetics*, **92**, 832–839.

Emery SN, McCauley DE (2002) Consequences of inbreeding for offspring fitness and gender in *Silene vulgaris*, a gynodioecious plant. *Journal of Evolutionary Biology*, **15**, 1057–1066.

Ennos RA (1994) Estimating the relative rates of pollen and seed migration among plant populations. *Heredity*, **72**, 250–259.

Ennos RA (2001) Inferences about spatial processes in plant populations from the analysis of molecular markers. In: *Integrating Ecology and Evolution in a Spatial Context* (eds Silvertown J and Antonovics J), pp. 45–71. Blackwell Science Ltd, Oxford, UK.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, **164**, 1567–1587.

Fénart S, Touzet P, Arnaud J-F, Cuguen J (2006) Emergence of gynodioecy in wild beet (*Beta vulgaris* ssp. *maritima* L.): a genealogical approach using chloroplastic nucleotide sequences. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **273**, 1391–1398.

Fénart S, Austerlitz F, Cuguen J, Arnaud J-F (2007) Long distance pollen-mediated gene flow at a landscape level: the weed beet as a case study. *Molecular Ecology*, **16**, 3801–3813.

Fénart S, Arnaud J-F, De Cauwer I, Cuguen J (2008) Nuclear and cytoplasmic genetic diversity in weed beet and sugar beet accessions compared to wild relatives: new insights into the genetic relationships within the *Beta vulgaris* complex species. *Theoretical and Applied Genetics*, **116**, 1063–1077.

Fievet V, Touzet P, Arnaud J-F, Cuguen J (2007) Spatial analysis of nuclear and cytoplasmic DNA diversity in wild sea beet (*Beta vulgaris* ssp. *maritima*) populations: do marine currents shape the genetic structure? *Molecular Ecology*, **16**, 1847–1864.

Forcioli D, Saumitou-Laprade P, Valero M, Vernet P, Cuguen J (1998) Distribution of chloroplast DNA diversity within and among populations in gynodioecious *Beta vulgaris* ssp. *maritima* (Chenopodiaceae). *Molecular Ecology*, **7**, 1193–1204.

Friar EA, Cruse-Sanders JM, McGlaughlin ME (2007) Gene flow in *Dubautia arborea* and *D. ciliolata*: the roles of ecology and isolation by distance in maintaining species boundaries despite ongoing hybridization. *Molecular Ecology*, **16**, 4028–4038.

Goudet J (1995) FSTAT (Version 1.2). A computer program to calculate F-Statistics. *Journal of Heredity*, **86**, 485–486.

Goudet J (2005) HIERFSTAT, a package for R to compute and test hierarchical *F*-statistics. *Molecular Ecology Notes*, **5**, 184–186.

Goudet J, Raymond M, De Meeüs T, Rousset F (1996) Testing differentiation in diploid populations. *Genetics*, **144**, 1933–1940.

Gouyon P-H, Vichot F, van Damme JMM (1991) Nuclear-cytoplasmic male-sterility – single-point equilibria versus limit-cycles. *The American Naturalist*, **137**, 498–514.

Graff A (1999) Population structure and reproductive fitness in gynodioecious *Sidalcea malviflora malviflora* (Malvaceae). *Evolution*, **53**, 1714–1722.

Hamrick JL, Nason JD (1996) Consequences of dispersal in plants. In: *Population Dynamics in Ecological Space and Time* (eds Rhodes OE, Chesser RK and Smith MH), pp. 203–236. University of Chicago Press, Chicago.

Hardy OJ, Vekemans X (2002) SPAGEDI: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*, **2**, 618–620.

Holsinger KE, Weir BS (2009) Genetics in geographically structured populations: defining, estimating and interpreting F_{ST}. *Nature Reviews Genetics*, **10**, 639–650.

Koelewijn HP, van Damme JMM (1995) Genetics of male sterility in gynodioecious *Plantago coronopus*. I. Cytoplasmic variation. *Genetics*, **139**, 1749–1758.

Laporte V, Viard F, Bena G, Valero M, Cuguen J (2001) The spatial structure of sexual and cytonuclear polymorphism in the gynodioecious *Beta vulgaris* ssp. *maritima*: I/ at a local scale. *Genetics*, **157**, 1699–1710.

Larsen K (1977) Self-incompatibility in *Beta vulgaris* L. I. Four gametophytic, complementary S-loci in sugar beet. *Hereditas*, **85**, 227–248.

Letschert JPW (1993) *Beta* section *Beta*: biogeographical patterns of variation and taxonomy. *Wageningen Agricultural University Papers*, **93**, 1–137.

Loiselle BA, Sork VL, Nason J, Graham C (1995) Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). *American Journal of Botany*, **82**, 1420–1425.

Loveless MD, Hamrick JL (1984) Ecological determinants of genetic structure in plant populations. *Annual Review of Ecology & Systematics*, **15**, 65–95.

Manicacci D, Couvet D, Belhassen E, Gouyon P-H, Atlan A (1996) Founder effects and sex ratio in the gynodioecious *Thymus vulgaris* L. *Molecular Ecology*, **5**, 63–72.

McCauley DE (1998) The genetic structure of a gynodioecious plant: nuclear and cytoplasmic genes. *Evolution*, **52**, 255–260.

McCauley DE, Bailey MF (2009) Recent advances in the study of gynodioecy: the interface of theory and empirism. *Annals of Botany (London)*, **104**, 611–620.

McCauley DE, Brock MT (1998) Frequency-dependent fitness in *Silene vulgaris*, a gynodioecious plant. *Evolution*, **52**, 30–36.

McGrath JM, Trebbi D, Fenwick A *et al.* (2007) An open-source first-generation molecular genetic map from a sugarbeet × table beet cross and its extension to physical mapping. *Crop Science*, **47**(Suppl. 1), 27–44.

Medrano M, Alonso C, Herrera CM (2005) Mating system, sex ratio, and persistence of females in the gynodioecious shrub *Daphne laureola* L. (Thymelaeaceae). *Heredity*, **94**, 37–43.

Mörchen M, Cuguen J, Michaelis G, Hanni C, Saumitou-Laprade P (1996) Abundance and lenght polymorphism of microsatellite repeats in *Beta vulgaris* L. *Theoretical and Applied Genetics*, **92**, 326–333.

Munkacsi AB, Stoxen S, May G (2008) *Ustilago maydis* populations tracked maize through domestication and cultivation in the Americas. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **275**, 1037–1046.

Nilsson E, Agren J (2006) Population size, female fecundity, and sex ratio variation in gynodioecious *Plantago maritima*. *Journal of Evolutionary Biology*, **19**, 825–833.

Nishizawa S, Kubo T, Mikami T (2000) Variable number of tandem repeat loci in the mitochondrial genomes of beets. *Current Genetics*, **37**, 34–38.

Olson MS, McCauley DE (2002) Mitochondrial DNA diversity, population structure, and gender association in the gynodioecious plant *Silene vulgaris*. *Evolution*, **56**, 253–262.

Olson MS, Graf AV, Niles KR (2006) Fine scale spatial structuring of sex and mitochondria in *Silene vulgaris*. *Journal of Evolutionary Biology*, **19**, 1190–1201.

Owen FV (1942) Inheritance of cross- and self-sterility and self-fertility in *Beta vulgaris*. *Journal of Agricultural Research*, **64**, 679–698.

Petit RJ, Duminil J, Fineschi S *et al.* (2005) Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. *Molecular Ecology*, **14**, 689–701.

Polato NR, Concepcion GT, Toonen RJ, Baums IB (2010) Isolation by distance across the Hawaiian Archipelago in the reef-building coral *Porites lobata*. *Molecular Ecology*, **19**, 4661–4677.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Ran Z, Michaelis G (1995) Mapping of a chloroplast RFLP marker associated with the CMS cytoplasm of sugar beet (*Beta vulgaris*). *Theoretical and Applied Genetics*, **91**, 836–840.

Raymond M, Rousset F (1995) GENEPOP (Version 1.2): a population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248–249.

Rice WR (1989) Analyzing tables of statistical tests. *Evolution*, **43**, 223–225.

Richards CM, Brownson M, Mitchell SE, Kresovich S, Panella L (2004) Polymorphic microsatellite markers for inferring diversity in wild and domesticated sugar beet (*Beta vulgaris*). *Molecular Ecology Notes*, **4**, 243–245.

Shykoff JA, Kolokotronis SO, Collin CL, Lopez-Villavicencio M (2003) Effects of male sterility on reproductive traits in gynodioecious plants: a meta-analysis. *Oecologia*, **135**, 1–9.

Slatkin M (1977) Gene flow and genetic drift in a species subject to frequent local extinctions. *Theoretical Population Biology*, **12**, 253–262.

Torimaru T, Tani N, Tsumura Y, Nishimura N, Tomaru N (2007) Effects of kin-structured seed dispersal on the genetic structure of the clonal dioecious shrub *Ilex leucoclada*. *Evolution*, **61**, 1289–1300.

Van Etten ML, Chang SM (2009) Effects of environmental heterogeneity on the distribution of sexes within and among populations in a gynodioecious species, *Geranium maculatum*. *New Phytologist*, **183**, 649–660.

Vaughton G, Ramsey M (2004) Dry environments promote the establishment of females in monomorphic populations of *Wurmbea biglandulosa* (Colchicaceae). *Evolutionary Ecology*, **18**, 323–341.

Vekemans X, Hardy OJ (2004) New insights from fine-scale spatial genetic structure analyses in plant populations. *Molecular Ecology*, **13**, 921–935.

Viard F, Bernard J, Desplanque B (2002) Crop-weed interactions in the *Beta vulgaris* complex at a local scale: allelic diversity and gene flow within sugar beet fields. *Theoretical and Applied Genetics*, **104**, 688–697.

Viard F, Arnaud J-F, Delescluse M, Cuguen J (2004) Tracing back seed and pollen flow within the crop-wild *Beta vulgaris* complex: genetic distinctiveness versus hot spots of hybridization over a regional scale. *Molecular Ecology*, **13**, 1357–1364.

Wade MJ, McCauley DE (1988) Extinction and recolonization: their effects on the genetic differentiation of local populations. *Evolution*, **42**, 995–1005.

Weir BS (1996) *Genetic Data Analysis II: Methods for Discrete Population Genetic Data*. Sinauer Associates, Inc., Sunderland, MA.

Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.

Whitlock MC, McCauley DE (1990) Some population genetic consequences of colony formation and extinction: genetic correlations within founding groups. *Evolution*, **44**, 1717–1724.

Widen B, Widen M (1990) Pollen limitation and distance-dependent fecundity in females of the clonal gynodioecious herb *Glechoma hederacea* (Lamiaceae). *Oecologia*, **83**, 191–196.

Wilson GA, Rannala B (2003) Bayesian inference of recent migration rates using multilocus genotypes. *Genetics*, **163**, 1177–1191.

Yang RC (1998) Estimating hierarchical *F*-statistics. *Evolution*, **52**, 950–956.

This work was a part of I.D.C.'s PhD thesis in population genetics and dynamics of gynodioecy in wild *Beta vulgaris* ssp. *maritima* populations. All authors are interested in ecology and evolution of plant reproductive systems, as well as in the interactions among sexual polymorphisms, fitness and spatial genetic structuring within and among populations. More information about the activities of the 'Laboratoire de Génétique et Evolution des Populations Végétales' can be found on the following website: http://gepv.univ-lille1.fr/.

Data accessibility

Sample locations, sexual phenotypes, cytoplasmic and nuclear genetic data are available in DRYAD: doi: 10.5061/dryad.gf8m3k8v.