

ORIGINAL ARTICLE

Metapopulation structure and fine-scaled genetic structuring in crop-wild hybrid weed beets

J-F Arnaud, J Cuguen and S Fénart¹

Laboratoire de Génétique et Évolution des Populations Végétales, FRE CNRS 3268, Bâtiment SN2, Université des Sciences et Technologies de Lille—Lille 1, Villeneuve d'Ascq cedex, France

This study explores the microspatial and temporal genetic variation in crop-wild hybrid weed beets that emerged from the seed bank in a cultivated field surveyed over two successive years. We demonstrate the occurrence of demes highly genetically differentiated, kin-structured, characterized by moderate effective population sizes, differing in propensity for selfing, and arising from nonrandom genetic subsets of the seed bank. Only one deme identified in the first survey year significantly contributed to the weed beets that emerged in the second year. Spatial structuring appears to be primarily due to gravity seed dispersal and limited pollen flow among weed beet

demes. Within each genetic cluster identified by Bayesian assignments and multivariate analyses, F_{IS} estimates and level of biparental inbreeding—revealed by progeny analyses—dropped to non-significant values. This suggests that random mating occurs at the scale of genetically distinct demes over a very short scale. Our results highlight the need to carefully depict genetic discontinuities in weed species, when attempting to describe their local genetic neighborhoods within which genetic drift and selective processes occur.

Heredity (2011) **107**, 395–404; doi:10.1038/hdy.2011.23; published online 30 March 2011

Keywords: agricultural disturbances; *Beta vulgaris*; effective population size; mating system; spatial genetic structure; seed bank

Introduction

In plant populations, pollen or seed dispersal often governs the degree of fine-scale genetic structure and the size of local genetic neighborhoods within which genetic drift or selective processes will take place (Knowles *et al.*, 1992; Mimura *et al.*, 2009). Pollen and seed dispersal operate in sequence, but an overall predominance of gene flow through pollen is generally suggested in seed plants for outcrossing species (Levin, 1981; Vekemans and Hardy, 2004). Gene flow vary in space but also over time. In plant populations, seed dormancy leads to the build-up of seed banks, that is, seeds that can be buried and remain dormant in the soil. Seed banks are commonly found in populations that live in unpredictable or extreme environments. Plants can thereby survive unfavorable periods by persisting as dormant seeds in the soil, forming a reservoir of sleeping genes that accumulates and stores genetic diversity over many reproductive seasons, thus preventing the risk of extinction or competition with both relatives and non-relatives (Vitalis *et al.*, 2004). This kind of migration from the past can affect the metapopulation dynamics of plant

populations and can greatly influence the effective population size (N_e) and the spatial genetic structure within and among populations (Templeton and Levin, 1979; Shimono *et al.*, 2006).

A long-lived seed bank might therefore be of crucial importance in driving metapopulation dynamics of invasive weed species that have evolved from crop cultivation in agricultural and human-disturbed habitats. This is particularly true in the case of weed beets (*Beta vulgaris*) whose propensity for invasiveness stems from a mixture of life-history traits inherited during crop-wild hybridizations. The *B. vulgaris* species complex constitutes an evolving system of domesticated and wild relatives connected through shared ancestry and high potential for hybridization (Fénart *et al.*, 2008). Genetic analyses demonstrate that weed beets arise from accidental hybridization events between cultivated seed bearers and wild inland beets located in the vicinity of seed production areas designed for sugar beet production (Arnaud *et al.*, 2009). These F1 crop-wild hybrid seeds are indistinguishable from certified seeds and, consequently, are inadvertently mixed with cultivar seeds and sown in sugar beet fields. This results in weed beets that infest sugar beet fields by bolting and flowering during the crop year, contrary to cultivated beets that are biannual and harvested for roots before flowering. The subsequent invasiveness of weed beets can be attributed to their ability to flower the first year after seedling emergence and to their primary dormancy allowing weed beet seeds to survive several years buried in the soil (Van Dijk, 2004; Sester *et al.*, 2006). First-year flowering arises through the introgression of a wild genetic background into the crop gene pool, that is, the *B*

Correspondence: Dr 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, Villeneuve d'Ascq cedex, France. E-mail: jean-francois.arnaud@univ-lille1.fr

¹Current address: Laboratoire Stress Abiotiques et Différenciation des Végétaux Cultivés, UMR INRA 1281, Bâtiment SN2, Université des Sciences et Technologies de Lille—Lille 1, F-59655 Villeneuve d'Ascq cedex, France

Received 22 July 2010; revised 17 February 2011; accepted 22 February 2011; published online 30 March 2011

gene that cancels the vernalization requirement (Van Dijk, 2004; Arnaud *et al.*, 2010). In the wild, *B. vulgaris* has a purely outcrossing mating system that depends on wind pollination (De Cauwer *et al.*, 2010b). However, a mixed-mating system has been shown in weed beet populations with the occurrence of high selfing rates, up to 44%. Propensity for selfing may stem from the introduction of a crop-derived gene that offsets gametophytic self-incompatibility, the *Sf* gene (Owen, 1942). This may allow for some reproductive assurance when mate availability is limited in case of low flowering plant density (Arnaud *et al.*, 2010).

If initial F1 crop-wild hybrid beets that bolt and flower in sugar beet crops are not removed, they can shed large amounts of seeds, potentially giving rise to large problematic weed beet populations in the following years because of the dormant soil seed bank. Weed beet seeds can persist for several years buried in the soil, with a cyclic decrease of the seed bank due to natural mortality being 20% each year (Sester *et al.*, 2006).

Weed beets belong to the same species complex as cultivated beets and, therefore, cannot be eliminated by herbicides in sugar beet fields. The most efficient method for preventing the spread of weed beets is traditional, mechanical or hand weeding. In addition to causing yield losses and mechanical problems during harvest, the occurrence of weed beets within sugar beet fields is also problematic because weed and cultivated beets are fully cross-compatible. The management of weed beets is then of crucial importance with regard to the release of genetically modified (GM) herbicide-tolerant sugar beets, to ensure the coexistence of traditional and GM crops and to avoid the spread of herbicide-resistant weed beets (Ellstrand, 2003).

However, to date, no compelling data is available on the level and change in genetic diversity in weed beets due to metapopulation processes in agricultural-disturbed habitats. No clear knowledge is available on whether a weed beet population should be considered at the scale of a given studied field or whether demographic and genetic functional units occur at a smaller scale, that is, within geographically structured demes of genetically related weed beets, as hypothesized by Viard *et al.* (2002). This last hypothesis relies on the particular fruit structure in beets: up to six seeds are aggregated in corky seed balls, dispersed by gravity over short distances. This provides opportunity for kin-structured spatial genetic structure and high level of genetic differentiation (Whitlock and McCauley, 1990; Fievet *et al.*, 2007).

This study focuses on the short-term change in spatial genetic structure and mating system in a population of weed beets. Among a set of weed beet populations previously described for genetic diversity, first-year flowering ability and mating system (Arnaud *et al.*, 2010), we studied one representative, well-established weed population for which plants from the seed bank arose the year following the first survey. We thus sought to answer the following questions: (1) is there inter-annual variation in the mating system and levels of genetic diversity between two successive years in this weed beet population? (2) Does this surveyed population show significant genetic discontinuities due to kin-structured demes of individuals? (3) If so, how does each deme contribute to the next generation and does

differential contribution translate into low N_e ? (4) Finally, is genetic diversity homogenized due to the mixing of seed cohorts from the seed bank, or is there a spatial genetic signature that reflects the establishment of genetically differentiated founding individuals?

Materials and methods

Sampling and molecular analyses

The studied weed beet population was located in a sugar beet field located within the French area for sugar beet production, located in northern France (2°52. 185' E; 50°34.799' N). For at least 15 years, sugar beets have been planted every 2–3 years in this field, depending on the crop rotation. This weed population was comprised in a set of weed beet populations previously surveyed for modeling the distribution of pollen dispersal, mating system and first-year flowering in non-vernalizing conditions (Fénart *et al.*, 2007; Arnaud *et al.*, 2010). In the first survey year, the sugar beet field was found to be locally infested by weed beets; 330 bolting plants were counted. This weed population was composed of back-crossed crop-wild hybrid weed beets germinating outside sowing lines. Therefore, this indicates that these weed beets arose from the soil seed bank. Five geographically distinct groups of individuals, labeled from 1 to 5, can be visualized in Figure 1. For this first

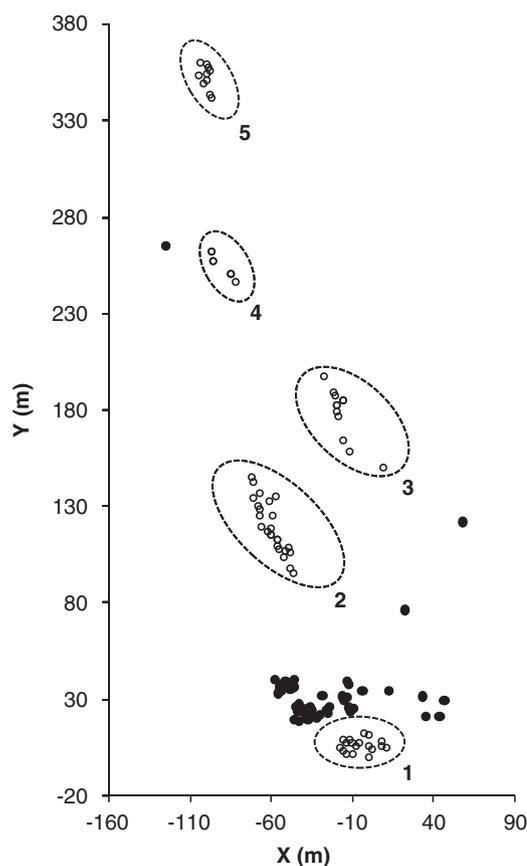


Figure 1 Spatial distribution of weed beets sampled in two successive years. Weed beets sampled in 2003 and 2004 are visualized by white circles and gray circles, respectively. Five geographical distinct clusters of weed beets occurred in 2003 and were labeled from 1 to 5.

survey year, leaf tissue from a total of 75 weed beets was sampled in summer 2003 and genotyped as described in Fénart *et al.* (2007). All collected weed beets were also examined for mating system, with 24 seeds per individual used for progeny analyses (Arnaud *et al.*, 2010). In the following year, the field was used for a potato crop, following the farmer's crop rotation schedule. However, despite the inability of weed beets to efficiently compete in most other crops of the rotation (Sester *et al.*, 2004), reappearance of weed beets still arose from the seed bank both in the field and at its margins, presumably owing to lax weeding practices (Figure 1). We exhaustively collected leaves from 66 weed beets in summer 2004, among which 16 were used as maternal plants with 24 seeds per individual, randomly chosen for progeny analyses. For both survey years, all adult plants were left in the field, and seeds used for progeny analyses were germinated in controlled conditions in a greenhouse until subsequent DNA extraction.

Extraction and purification of total DNA from the leaves of the 141 sampled individuals (75 and 66 weed beets in 2003 and 2004, respectively), together with their progeny arrays, was performed using the DNeasy 96 Plant Kit (Qiagen Inc., Hilden, Germany), as per the manufacturer's instructions. All individuals were examined for nuclear genetic variation using 10 microsatellite loci named GAA1, GTT1, GCC1, BVM3, CAA1, SB04, SB06, SB07, SB15 and FDSB1027 (Mörchen *et al.*, 1996; Viard *et al.*, 2002; Richards *et al.*, 2004). Allele sizing and PCR conditions are described in De Cauwer *et al.* (2010b). To determine cytoplasmic diversity and ensure that weed beets are the results of progenies from previous accidental hybridization events between cultivated seed bearers and wild pollinators, we checked for the presence of Owen cytoplasmic male sterility, named CMS Owen, and universally used in sugar beet breeding programs (Arnaud *et al.*, 2003). A diagnostic chloroplastic PCR-restriction-fragment length polymorphism marker was used as described in Ran and Michaelis (1995).

Data analysis

Interannual genetic variation was described by multilocus means of observed (H_o) and expected (H_e) heterozygosity, number of alleles (A_n), and allelic richness (A_r). Departures from Hardy-Weinberg (HW) expectations were quantified with the intrapopulation fixation index (F_{IS}), for which significant deviations from 0 were tested using 10 000 permutations of alleles among individuals within population samples. Pairwise population differentiation (F_{ST}) over years and among distinct clusters of individuals (see below) was assessed and tested for significance using a G-test with 10 000 runs of multilocus genotypes randomization among samples. These analyses were conducted using FSTAT version 2.9.3 (Goudet, 1995).

Multilocus population-level outcrossing rates (t_m), difference between multilocus and single-locus outcrossing rates ($t_m - t_s$), which provides an estimate of biparental inbreeding, and correlation of outcrossed paternity (r_p) were computed at the population level for both survey years following the numeric Newton-Raphson algorithm and population gene frequencies, using the MLTR version 3.2 software package (Ritland, 2002).

To identify genetic discontinuities across the focal area, we used the non-spatial, Bayesian, model-based algorithm described in Hubisz *et al.* (2009) to depict the pattern of fine-scale population structure and to further trace back the putative origin of individuals emerging from the seed bank one year later. Using STRUCTURE (Hubisz *et al.*, 2009) on the combined data set of individuals sampled in 2003 and 2004, we assessed the number of potential clusters (K) from 10 different runs along a range of K varying from 1 to 10. For each run, uninformed priors for population memberships were used, with a burn-in period of 10^4 and two 10^6 Markov chain Monte Carlo replications under a model of admixture ancestry with correlated frequencies. The *ad hoc* statistic ΔK , on the basis of rate of change in the log-likelihood of the values of K (Evanno *et al.*, 2005), was then calculated to determine the best-fit accurate level of structure. Once the most accurate K was estimated, we performed a new run of clustering, this time using information on the basis of sampling locations of individuals to partition them into each inferred clusters, following the procedure outlined in Hubisz *et al.* (2009). Sampling locations were coded as 1–5, which corresponded to the five geographical distinct clusters of weed beets (see Figure 1). This placed a previous weight on clustering outcomes that are correlated with the sampling locations and did not allow false inference when the ancestry of individuals was uncorrelated with the sampling locations.

To complement the results from non-spatial Bayesian clustering, we also analyzed the data set sampled in 2003 by assigning individuals from multilocus genotypes into a number K of clusters, using the spatial model of Bayesian clustering implemented in the R package GENELAND and described in Guillot *et al.* (2005). We allowed K to vary from 1 to 10, with 10 independent runs assuming correlated allele frequencies and running 2×10^6 Markov chain Monte Carlo iterations. Once the modal K was inferred, a detailed map of posterior probability assignments of group membership was estimated using a burn-in of 10 000 iterations.

However, Bayesian clustering approaches are on the basis of genetic assumptions that rely on HW equilibrium and linkage equilibrium between loci. These prerequisites are often violated in natural populations and Bayesian assignments should be complemented by different approaches (for example, see Rutledge *et al.*, 2010). In this respect, multivariate ordination analyses are explanatory methods that aim at identifying the patterns of genetic variability without requiring strong assumptions about the genetic model (reviewed in Jombart *et al.*, 2009). Therefore, to corroborate the genetic structure inferred from Bayesian clustering, we used a recently developed multivariate analysis that makes no assumptions regarding the underlying data structure and population genetic model; the discriminant analysis of principal components introduced by Jombart *et al.* (2010). This multivariate method consists of a two-step procedure, as follows: (i) genetic data are first centered and scaled using principal component analysis as a previous step; (ii) principal components of principal component analysis are then submitted to a linear discriminant analysis. This approach has been showed to extract useful information from genetic data in subdivided populations and to outperform Bayesian

clustering in case of hierarchical structure or clinal variation due to isolation by distance (Jombart *et al.*, 2010). In all, 50 principal components of principal component analysis were retained in the data transformation step and all analyses were carried out using the ADEGENET package (Jombart, 2008) implemented in R (R Development Core Team, 2009).

To assess the intensity of spatial genetic structure and its associated significance, we used the S_p statistic on the basis of pairwise kinship coefficients as described in Vekemans and Hardy (2004). To further depict the scale and the direction over which a spatial genetic structure could occur through short-range seed dispersal after germination from the seed bank, we also conducted a two-dimensional Mantel correlogram analysis by using the methodology described in Oden and Sokal (1986). If different mixed seed bank cohorts germinate together, there should be no significant trends in spatial genetic structure. In contrast, germination of kin-structured seed bank cohorts should leave a genetic signature suggestive of isolation by distance among different founding groups. For the matrix of pairwise genetic distances among weed beets, we used the Cavalli-Sforza and Edwards' (1967) chord distance as a measure of genetic divergence. Results were only reported for distance/direction sectors containing at least 20 pairwise comparisons to ensure statistical relevance of normalized Mantel r_z correlation coefficients (Smouse *et al.*, 1986). Significance of r_z was tested in each distance-direction class using Mantel tests with 10 000 randomizations.

As proposed by Saarinen *et al.* (2010), we estimated the N_e both through the temporal method on the basis of magnitude of random changes in allele frequency over time and through the single-sample method on the basis of linkage disequilibrium (LD) information among alleles at different loci caused by genetic drift in finite populations. We first used a moment-based temporal estimator of N_e described in Jorde & Ryman (2007) and implemented in the TEMPOFS software, which provides estimates less sensitive to small sample sizes and skewed allele frequencies. Temporal estimates of N_e were calculated according to sample plan 1, that is, when individuals are either sampled after reproduction or non-destructively sampled and returned back into the population before reproduction (Waples, 1989). We then used the Burrows' Δ as a single-sample estimator on the basis of LD and implemented in the LNDE software (Waples and Do, 2010). This method corrects for biases associated with small sample sizes, does not depend on the assumption of random mating and yields unbiased estimates of LD (\hat{r}^2) from which N_e can be derived. To ensure that no bias was introduced due to low-frequency alleles, N_e estimates were calculated after excluding alleles with frequencies less than 5%.

Last, we performed a paternity analysis for data set sampled in the first survey year (2003). Data were reanalyzed on the basis of paternity study by Fénart *et al.* (2007) that aimed at modeling pollen-dispersal kernels in weed beets. Paternity assignment was performed using the maximum likelihood-based method described in Marshall *et al.* (1998). To determine whether the paternity of offspring could significantly be assigned to sampled flowering adults with the highest paternity likelihood, we used the difference in likelihood score between the most likely parents (Δ LOD) with a critical value (ΔC) of

Δ , below which paternity could not be attributed at 95% confidence level. Proportions of siring events within and among demes were on the basis of 358 successfully assigned seedlings at a 95% confidence level.

Results

Level of genetic diversity and genetic differentiation over years

In contrast to nuclear genetic data, cytoplasmic genetic diversity was monomorphic, that is: all weed beets exhibited the CMS Owen cytoplasm, indicating a common cultivated maternal origin. Among the 66 weed beets sampled in the second survey year, 10 were found to be triploid at some nuclear microsatellite loci and were subsequently removed from the following population genetics analyses. No triploid individuals were found in the first survey year. Genetic differentiation was highly significant across survey years, with an associated F_{ST} value of 0.045 ($P < 10^{-4}$). Although a trend of lower levels of genetic variability can be seen in Table 1, neither mean H_o and H_e , nor the mean A_n and A_r , varied significantly between survey years (all at $P > 0.05$, two-tailed t -test). In contrast, genotypic structure expected under HW equilibrium varied significantly over years; mean multilocus F_{IS} estimates were significantly greater than zero in the first survey year ($F_{IS} = 0.033$, $P < 0.01$), but did not show significant departures from random mating in the following year ($F_{IS} = 0.018$, $P = 0.084$). Likewise, whereas a mixed-mating system was observed in weed beets collected in the first survey year, t_m was not significantly different from complete outcrossing in progeny arrays analyzed in the following year, which also showed a very low level of biparental inbreeding ($t_m = 0.984 \pm 0.026$, $t_m - t_s = 0.019 \pm 0.030$, see Table 1). Levels of r_p were, however, of the same order of magnitude in both survey years.

Genetic structuring and cluster membership

When using the non-spatially explicit clustering algorithm of Hubisz *et al.* (2009), complemented with the *ad hoc* statistic of Evanno *et al.* (2005), the ΔK statistic showed a strong mode at $K = 3$, whether the data set included individuals sampled in 2004 or not (Figure 2a). Average genetic differentiation among the three inferred clusters was high, with a mean F_{ST} value of 0.10 (95% CI: 0.062–0.139; $P < 10^{-4}$) and mean pairwise F_{ST} estimates ranging from 0.095 to 0.103, all at $P < 0.01$ after Bonferroni correction. As a result of removing the spatial Wahlund effect, mean within-cluster F_{IS} values fell to nonsignificant values of -0.099 , -0.058 and 0.007 for clusters 1, 2 and 3, respectively, all at $P > 0.05$. Further evidence for genetic equilibrium within clusters was revealed in estimates of biparental inbreeding ($t_m - t_s$), which decreased to values close to zero within each cluster (Table 1). Overall, the genetic distinctiveness among clusters was high and generally congruent with the spatial locations of weed beets (Figure 2b). Individuals belonging to geographical patch 4 were found to have admixed genotypes from both genetic clusters 2 and 3. Only one individual sampled within cluster 2 had a strong probability ($q > 0.60$) of belonging to another inferred cluster, cluster 3 (Figure 2b). As shown by S_p statistics (see Table 1), no significant spatial genetic

Table 1 Mean estimates of nuclear genetic diversity and mating system parameters calculated over 10 microsatellite loci for a weed beet population sampled in 2003 and in 2004

Survey year	2003				2004
Sample size	75				56
<i>Mean genetic diversity parameters</i>					
<i>Cluster 1</i>	<i>Cluster 2</i>	<i>Cluster 3</i>	<i>All</i>		
A_n	4.200	3.700	5.300	6.400	5.400
A_r	3.607	4.856	4.959	6.010	5.380
H_e	0.506	0.507	0.614	0.590	0.563
H_o	0.555	0.536	0.610	0.571	0.553
F_{IS}	-0.099 ^{NS}	-0.058 ^{NS}	0.007 ^{NS}	0.033 ^{**}	0.018 ^{NS}
S_p	0.004 ^{NS}	0.006 ^{NS}	0.036 ^{***}	0.042 ^{***}	0.016 ^{***}
<i>Mating system parameters</i>					
N_f	20	25	30	75	16
N_s	480	600	720	1800	384
t_m	0.922 (0.012)	0.776 (0.011)	0.782 (0.073)	0.809 (0.006)	0.984 (0.026)
$t_m - t_s$	0.026 (0.014)	0.045 (0.013)	0.033 (0.039)	0.114 (0.008)	0.019 (0.030)
r_p	0.091 (0.018)	0.199 (0.028)	0.203 (0.037)	0.262 (0.015)	0.213 (0.033)
<i>Effective population size</i>					
$N_{e(LD)}$ CI (95%)	17 (8.5–47.4)	9.6 (5.1–17.5)	36.5 (20.3–55.8)	14.4 (11–18.8)	23.8 (15.6–38)
$N_{e(FS)}$ CI (95%)	—	—	—	5 (3–17)	—

Abbreviations: A_n , mean number of alleles; A_r , allelic richness; CI, confidence intervals; F_{IS} , intrapopulation fixation index, a measure of departure from panmixia and its associated significance; H_e , gene diversity; H_o , observed heterozygosity; $N_{e(FS)}$, temporal method; $N_{e(LD)}$, linkage disequilibrium; N_f , number of mother plants for which progeny arrays were analyzed; N_s , total number of progeny surveyed; r_p , multilocus correlated paternity within maternal sibships; S_p , statistic quantifying the strength of spatial genetic structure; t_m , multilocus estimates of outcrossing rate; ($t_m - t_s$), estimation of biparental inbreeding.

Clusters 1–3 denote the genetically distinct groups of weed beets depicted using Bayesian clustering for survey year 2003. Standard errors of t_m , ($t_m - t_s$), and r_p were estimated using 1000 bootstraps on progeny arrays and are indicated in parentheses. Also provided are multilocus estimates of effective population size on the basis of $N_{e(LD)}$ or on the basis of $N_{e(FS)}$. Their associated 95% confidence intervals were estimated by a jackknifing procedure over the 10 loci.

** $P < 0.01$; *** $P < 0.001$; NS, not significant.

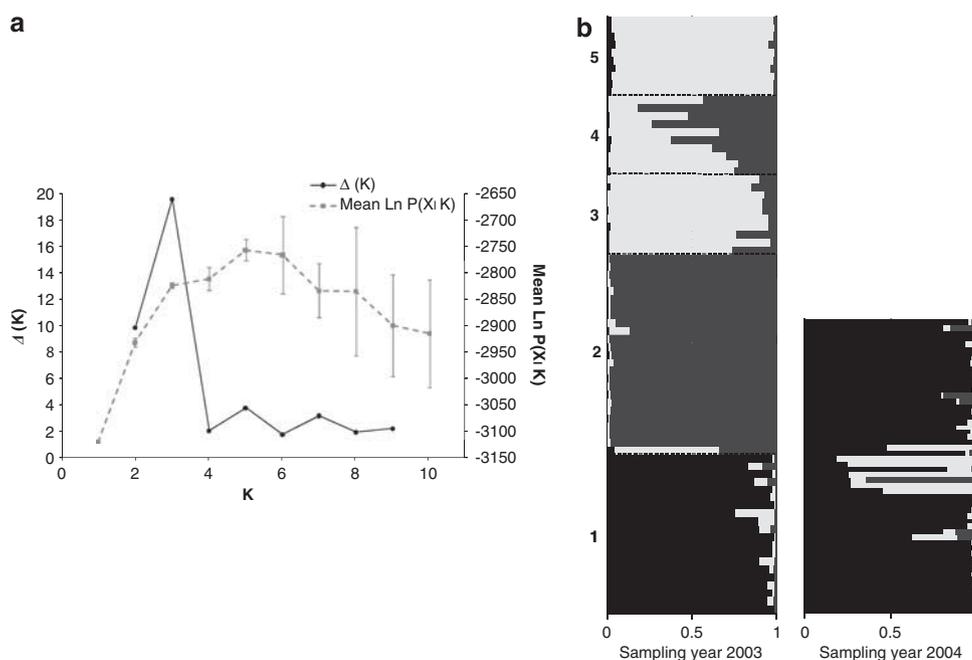


Figure 2 Bayesian assignment results from the non-spatial clustering algorithm described in Hubisz *et al.* (2009) and applied on the whole data set comprising the two temporal samples. (a) Mean (\pm s.d.) probabilities of the data $\text{Ln Pr}(X|K)$ over 10 replicated runs plotted as a function of the putative number of clusters K (gray square) and the standardized second-order rate of change of $\text{Ln Pr}(X|K)$, ΔK , as a function of K (black circle). (b) Bayesian assignment probabilities of membership for the whole data set of weed beets sampled in 2003 and 2004 into the three inferred clusters. Each individual is represented by a thin horizontal line (y axis) partitioned into three colored segments that represented the individual's estimated membership fractions depicted in $K = 3$ clusters (x axis); black: cluster 1, dark gray: cluster 2; light gray: cluster 3.

structuring occurred within clusters, except for cluster 3. Running the clustering algorithm on both survey years further showed that almost all weed beets sampled in

2004 were assigned with very high posterior probability of membership to only one cluster sampled in 2003, cluster 1. Only 7 out of 56 weed beets (8.9%) were admixed

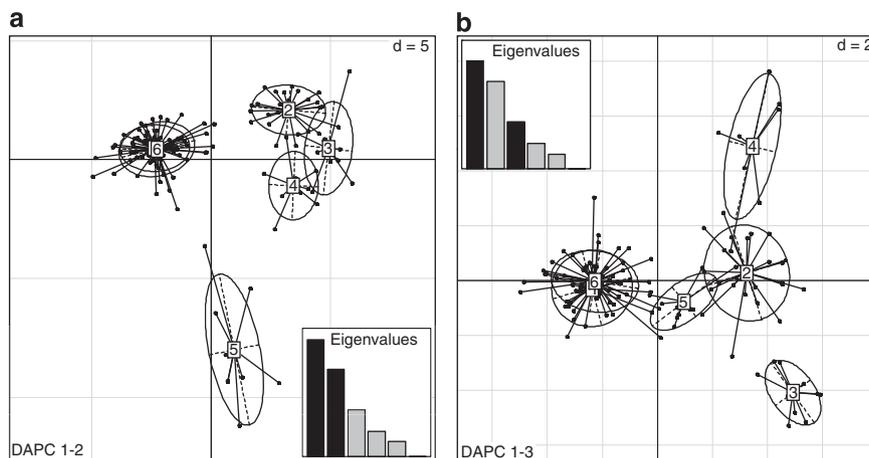


Figure 3 Typology of weed beets and representation of genetic relatedness between geographical clusters of weed beets obtained by discriminant analysis of principal components (DAPC). Scatterplots show the first two principal components (a) and the first and third principal components (b) of the DAPC of weed beets genotypes. Eigenvalues corresponding to the represented components are filled in black. Points represent genotypes and geographical groups of weed beets are labeled inside their 95% inertia ellipse. Labels from 1 to 5: weed beets sampled in 2003; label numbered 6: weed beets sampled in 2004.

individuals, with probability of membership higher than 50% for belonging to other inferred clusters (Figure 2b). The genetic differentiation over time for individuals belonging to cluster 1 was moderate but significant, with a F_{ST} value of 0.022 (CI: 0.004–0.043, $P < 0.01$).

Using the spatially explicit algorithm of Guillot *et al.* (2005) for the 2003 data set, posterior density distributions of the estimated K displayed consistent results across the 10 replicated runs. As in the non-spatial explicit Bayesian clustering analysis, we found a clear mode at $K=3$. Identical genetic discontinuities were depicted; with some admixed individuals located in the geographical patch numbered 4 (see Supplementary Figure S1).

The multivariate discriminant analysis of principal components analysis strongly confirmed the genetic distinctiveness of geographical patches of weed beets. Visual inspection of the scatterplots of the first principal components of discriminant analysis of principal components showed that five groups of genetically closer clusters could be identified (Figure 3). The geographical patch of weed beets labeled 1 was clearly differentiated from the others by the first three principal components. However, depending on the principal components, the remaining geographical patches of weed beets were less structured, with some mixing of individuals in accordance with previous Bayesian assignments. As for Bayesian analyses, the most striking feature was the great genetic proximity of weed beets sampled in the second year survey with the geographical patch labeled 1 in the first year survey, as shown by visual assessment of the scatterplots that did not allow distinguishing these two groups of weed beets (Figure 3).

Spatial genetic structure in the weeds that emerge from the seed bank

To further test whether seed dispersal is constrained over short distances, we tested for genetic structuring through isolation by distance in individuals that emerged from the seed bank in the second survey year. Along with a significant S_p statistic (see Table 1), neighboring

individuals showed strong genetic similarity within the first four distance/direction classes, declining thereafter with significantly negative r_z values found for the last distance classes along an east/west gradient (Figure 4). This long-distance differentiation does not correspond to a classic continuous isolation-by-distance pattern. Instead, it exemplifies the germination of two spatially and genetically distinct founding groups of individuals, as depicted in Supplementary Figure S2, by using the spatially explicit clustering method (Guillot *et al.*, 2005). High genetic differentiation occurred between the two clusters ($F_{ST}=0.13$, $P < 0.001$), which were respectively composed of 49 and 7 individuals, with probability of membership that matched previous results on the basis of non-spatial genetic clustering applied to the whole data (see above).

Estimates of N_e

Single-sample N_e estimated on the basis of LD (Burrow's Δ) found in the first and second survey years were 14.4 (95% CI: 11–18.8) and 23.8 (95% CI: 15.6–38) (see Table 1). When accounting for the effect of population genetic discontinuities within the field, N_e estimates were of the same order of magnitude with overlapping confidence intervals for the three inferred genetic clusters, with N_e of 17, 9.6 and 36.5 for clusters 1, 2 and 3, respectively. The temporal N_e estimates on the basis of changes of allele frequencies over survey years were lower ($N_e=5$; 95% CI: 3–17; Table 1). Because Bayesian assignment tests indicated that only genetic cluster 1 seemed to contribute to most of the weed beet emergence from the seed bank, we re-examined the data set by excluding the seven individuals previously assigned to cluster 2 and 3 from the analysis, that is: N_e estimates were then even closer to the single-sample estimate with $N_e=8$ (95% CI: 4–26).

Pollen dispersal and siring events

By reanalyzing the parentage data set used in Fénart *et al.* (2007) for modeling the dispersal of pollen flow, we showed that most siring events involved intra-patch pollen flow (Figure 5). Using the three clusters

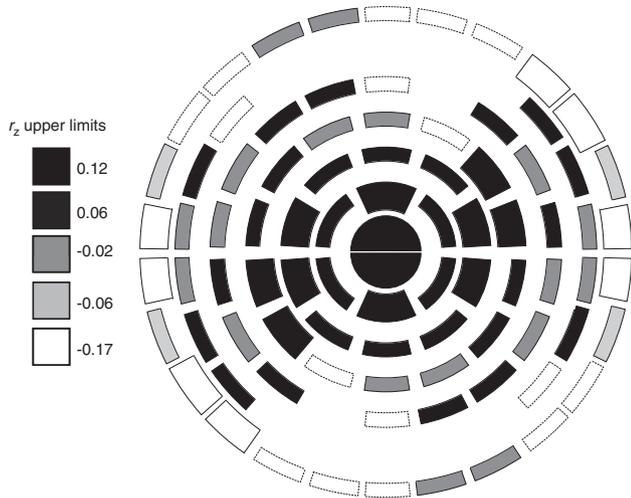


Figure 4 Two-dimensional Mantel correlogram summarizing the spatial genetic variation among weed beets that emerged from the seed bank the second survey year (2004). This windrose correlogram is on the basis of comparison of pairwise genetic distances against geographical distances between individuals. The seven circular and successive annuli represent interval distance classes with upper limits of 4, 7, 12, 19, 28, 39 and 52 m, respectively. Full-width boxes correspond to significant r_z values at $P < 0.05$ after 10 000 permutations, whereas half-width boxes represent nonsignificant correlation coefficients. Boxes with dashed outlines are distance-direction classes with insufficient number of pairwise comparisons (< 20) to ensure relevant r_z estimates. Shading represents the range of values covered by the normalized Mantel statistic r_z .

previously inferred by Bayesian clustering as geographical units for weed beets, we found that mating events ranged from 73 to 85% within clusters, whereas inter-cluster pollen flow was much lower among the three genetic clusters, ranging from 1.4 to 9.6% (see Figure 5).

Discussion

Patterns of fine-scaled spatial genetic structure

Despite agricultural practices that lead to extinctions of weed beet populations each year during harvest, a persistent weed beet seed bank may (i) prevent the increase and maintenance of significant genetic structure by pooling the reproductive output of several weed generations and averaging out the effects of each generation's dispersal patterns and (ii) allow further opportunities for secondary dispersal events through soil disturbance due to plowing, harrowing or other agricultural practices, which could weaken the spatial genetic structure (for example, see Cabin, 1996). However, we found substantial genetic discontinuities within the surveyed field, which can be attributed to nonrandom genetic subsets of the soil seed bank. Three genetically distinct and highly differentiated clusters of individuals were depicted in the first survey year, of which only one seemed to significantly contribute to the weed beets that emerged the second year. The spatial genetic structure we observed is then consistent with gravity-driven seed fall that generates a seed bank comprised of kin-structured clusters of progenies in the vicinity of the previous generation's adults, as often observed in weed species in which seed banks are spatially aggregated (for

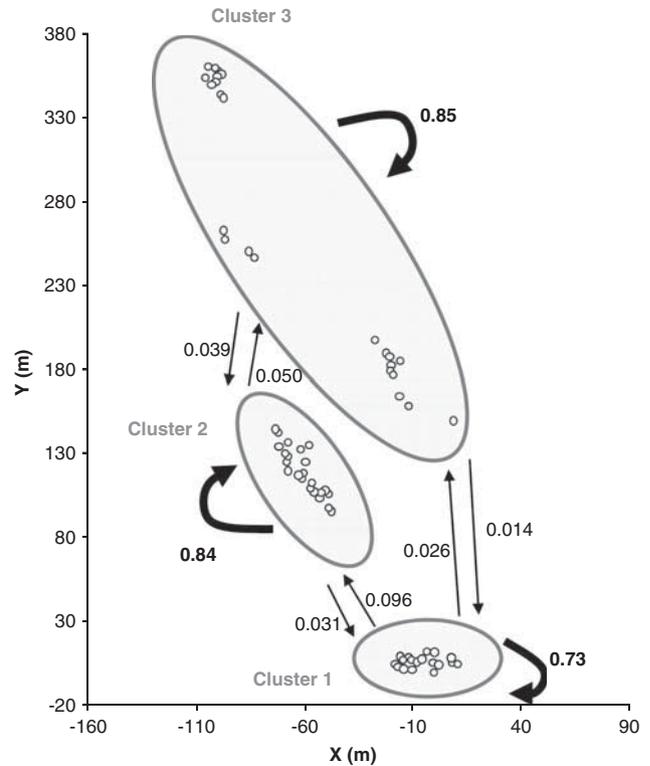


Figure 5 Graphical representation of siring event proportions mediated through pollen dispersal within and among the three genetic clusters of weed beets inferred by Bayesian assignments. Arrows indicate the proportion and the orientation of pollen flow events. Pollen flow proportions involving a mother plant sired by pollen donors located outside the studied sugar beet field are equal to 0.14, 0.07 and 0.09 for clusters 1, 2 and 3, respectively. These pollen donors are weed beets located in nearby fields surveyed by Fénart *et al.* (2007).

example, see Lortie *et al.*, 2010). Furthermore, beet seeds are aggregated in a seed ball, simply gravity dispersed only over short distances (Fievet *et al.*, 2007; De Cauwer *et al.*, 2010b). This type of fruit structure enhances the likelihood that founders are close relatives (siblings) and, as a consequence, strongly increases the genetic differentiation among demes compared with an island model of population structure with founders drawn at random from diverse sources (Whitlock and McCauley, 1990; Ingvarsson and Olsson, 1997). Along with spatially restricted pollen dispersal, kin-structured founders may thus explain the highly significant differentiation ($F_{ST} = 0.10$; $P < 10^{-4}$) observed at very fine geographical scale among the three genetically distinct clusters of weed beets.

Spatial scale of functional genetic units

At the scale of the whole surveyed weed beet population, significant departure from HW expectations occurred along with significant spatial genetic structure and—as revealed by progeny analyses—a significant level of biparental inbreeding ($t_m - t_s = 0.114$). However, within each cluster identified by Bayesian assignments and discriminant analysis of principal components analyses, all F_{IS} estimates and level of biparental inbreeding dropped to values that were not significantly different

from zero, suggesting that random mating occurs at the scale of genetically distinct demes, over very restricted geographical areas. Likewise, low or no significant spatial genetic structure occurred within clusters, as shown by S_p statistics. The same holds for weed beets that primarily emerged from one of the genetically distinct clusters one year later. The high observed genetic differentiation, in concert with nonsignificant F_{IS} values within clusters, corresponds to a metapopulation equilibrium theoretically expected in an annual, randomly mating population with a seed bank, a low seed dispersal rate, a large number of flowering plants, and a high germination rate (Vitalis *et al.*, 2004). Our findings suggest that (i) once a seed bank has been successfully constituted, the spatial distribution of genetic diversity is primarily determined by the origin of founders, where the founders act as distinct seed sources and reinforce significant spatial genetic structure among genetically differentiated seed banks, (ii) random mating occurs within local genetic neighborhoods generated by the seed bank, (iii) local mixing of the seed bank at a very fine scale, due to the close spacing of individuals producing progeny dispersed over very short distances with overlap in seed shadows, results in random distribution of genetic diversity when a given local seed bank germinates (for example, see Knowles *et al.*, 1992; Cabin, 1996). Our study thus highlights the need for carefully considering fine-scaled genetic discontinuities when attempting to describe patterns of mating system and genetic structure in a weedy species at small geographical scales.

Seed- and pollen-dispersal events

Conversely, if seed dispersal is restricted, extensive pollen flow may mitigate the impact of seed dispersal on genetic structure by diluting the genetic differentiation among seed bank cohorts. Nonetheless, the high level of r_p suggested that only a few pollen donors sire a given plant, which can be attributed to density effects that constrain pollen flow to very short distances. Although this pattern is unexpected in a wind-pollinated species (Levin, 1981; Friedman and Barrett, 2009), high local individual density may saturate the pollen cloud of a given plant by pollen from close neighboring individuals, and this effect has been shown to affect mating patterns within neighborhoods in weed and wild beet populations (Fénart *et al.*, 2007; De Cauwer *et al.*, 2010b). The high nuclear genetic differentiation observed in this study thus appeared to persist among demes because of limited pollen exchanges. Indeed, we found that most of mating events were spatially restricted and occurred within geographical clusters of individuals, whereas intercluster pollen flow was much lower (Figure 5).

It should also be noted that seed dispersal was not strictly restricted within local genetic clusters. Despite the lack of cytoplasmic polymorphism found in our study, Bayesian clustering of individuals, based on nuclear data, revealed rare but effective seed dispersal events among clusters. One first-generation migrant from cluster 3 was clearly identified within cluster 2, implying a seed movement of at least 50 meters within the sugar beet field. Plants that emerged from the seed bank the following year further clearly revealed weed beets that had high probabilities of being members of

two out of three distinct clusters identified the previous year, one of them being located 150 meters away. Beyond pollen dispersal, these findings thus indicate that agricultural practices may allow some movement of seeds across a field.

Variation in mating system among clusters

During the course of accidental crop-wild hybridizations in the seed production area, the introduction of the crop-derived *Sf* gene in weedy beets may provide opportunities to reproduce when mate availability is limited, that is, reproductive assurance (Baker, 1974). Indeed it has been shown that, although being a wind-pollinated species, pollen limitation of reproduction occurs over very short scales in *B. vulgaris* (De Cauwer *et al.*, 2010a). Mates may then be limited during the initial stage of infestation, when the density of F1 crop-wild hybrids is very low (Van Dijk, 2004; Arnaud *et al.*, 2010). In this study, weed beet clusters slightly differed in their propensity for selfing with estimates of t_m ranging from 0.77 to 0.92, suggesting that the distribution of the self-fertility factor varied within the sugar beet field. Compared with the established, first-year weed beet populations composed of F1 crop-wild hybrids (see Arnaud *et al.*, 2010); an overall predominantly outcrossing mating system was observed in this study. Moreover, we showed that plants that emerged from the seed bank one year later yielded a t_m not significantly different from pure outcrossing, as commonly found in wild accessions (De Cauwer *et al.*, 2010b). Altogether, there was no conclusive evidence that agricultural-like disturbances affect mating patterns toward increased selfing rate in well-established weed beet populations, presumably due to the high local density of conspecifics found in this weed population.

N_e in weed beets

In spatially subdivided populations, although extinction/recolonization dynamics are thought to reduce N_e , a seed bank may increase the global N_e by acting as a reservoir of neutral genetic diversity that decreases the efficiency of random genetic drift (Honnay *et al.*, 2008). We found relatively consistent and overlapping values of N_e when using the single-sample estimator on the basis of LD, with N_e estimates ranging from 9 to 36, depending on the genetically distinct cluster considered. The temporal method yielded a slightly lower N_e estimate that was similar in value. Nonetheless, single-sample and temporal N_e estimates must be interpreted with caution; they assume that genetic drift is the only evolutionary force acting, and that mutation, selection and migration are negligible. Because we showed that seed and pollen flow among demes of weed beets were very low, we could reasonably assume that our estimates of N_e were not biased by immigration effects.

Another potential bias that could lead to erroneous estimates of N_e is the assumption of non-overlapping generations, an assumption that could be violated in our study population. Indeed, it is difficult to distinguish between transient seed banks, composed of seeds that germinate immediately the next reproductive year, and persistent seed banks, where seeds remain viable in the soil during several years (for example, see Shimono *et al.*, 2006). In our study, the occurrence of 10 aneuploid

individuals in the weed population probably resulted from a residual weed beet seed bank established when tetraploid cultivar lines were used to pollinate diploid seed bearers to produce triploid sugar beet cultivars, a breeding process that was progressively abandoned as of 2000 and definitively abolished since 2006 (unpublished report from the 'Institut Technique de la Betterave', Paris). Beyond this resurgence of old crop-weed hybrids, the overlap of different generations may be low because the recruitment of seeds from different cohorts should generate significant positive F_{IS} values within demes due to a temporal Wahlund effect (Vitalis *et al.*, 2004; Honnay *et al.*, 2008). As (i) the genotypic structure we observed within clusters did not deviate from HW expectations and (ii) clusters were highly genetically differentiated as expected for age-structured demes (see Vitalis *et al.*, 2004), mixtures of different cohorts are unlikely to occur and bias our N_e estimates.

Implications for weed management in beet crops

This study showed a predominantly outcrossing mating system, which favors the dispersal of genes in weed beets, and quite substantial N_e , which decreases the efficiency of genetic drift. In addition, we found some movements of seeds across the study field, which means that it may be difficult to confine weed beets to a restricted geographical area. These processes are likely to increase the spread of selectively advantageous genes, which is of crucial concern in the case of GM genes that confer herbicide-resistance ability. Managing and controlling weed beet populations inside a field where GM crop beets are cultivated and avoiding their spread to neighboring non-GM fields thus requires the development of an optimal cropping system, such as the combination of various tillage and burial practices, and strict regulations, such as isolation measures or time intervals between successive sugar beet crops on the same field (Ellstrand, 2003; Sester *et al.*, 2007). Careful attention should be directed to sugar beet, potato and pea crops in priority because they provide a large amount of light, leave more free soil due to larger inter-row spacing and result in a less shady crop canopy (Sester *et al.*, 2004). In our study, the greater light exposure in a potato field may indeed explain the successful reappearance of weed beets one year later, especially at the field margin where weeding efforts were apparently lower. Finally, predictive models that aim at preventing crop-to-weed gene flow among fields through efficient cropping system should also take into account the occurrence, within a field, of fine-scale structuring into distinct genetic neighborhoods that are highly genetically differentiated and that can differ in mating system.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

We wish to express our gratitude to Fabrice Roux, Pascal Touzet, Carolyn R Engel, Isabelle De Cauwer and two anonymous reviewers for providing insightful comments on a previous version of the paper. This work was funded by the 'Contrat de Plan État/Région Nord-

Pas-de-Calais'. Stéphane Fénart was supported by an INRA/Région Nord-Pas-de-Calais fellowship. Jean-François Arnaud is grateful to the CNRS for supporting him as a full-time research scientist during the 2009–2010 academic year.

References

- 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. *Evol Appl* **3**: 305–318.
- Arnaud J-F, Fénart S, Godé C, Deledicque S, Touzet P, Cuguen J (2009). Fine-scale geographical structure of genetic diversity in inland wild beet populations. *Mol Ecol* **18**: 3201–3215.
- 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. *Proc R Soc Lond B* **270**: 1565–1571.
- Baker HG (1974). The evolution of weeds. *Ann Rev Ecol Syst* **5**: 1–24.
- Cabin RJ (1996). Genetic comparisons of seed bank and seedling populations of a perennial desert mustard, *Lesquerella fendleri*. *Evolution* **50**: 1830–1841.
- Cavalli-Sforza LL, Edwards AWF (1967). Phylogenetic analysis: models and estimation procedures. *Am J Hum Genet* **19**: 233–257.
- 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*. *J Evol Biol* **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*. *Mol Ecol* **19**: 1540–1558.
- Ellstrand NC (2003). Current knowledge of gene flow in plants: implications for transgene flow. *Philos Trans R Soc Lond B* **358**: 1163–1170.
- Evanno G, Regnaut S, Goudet J (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* **14**: 2611–2620.
- 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. *Theor Appl Genet* **116**: 1063–1077.
- 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. *Mol Ecol* **16**: 3801–3813.
- 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? *Mol Ecol* **16**: 1847–1864.
- Friedman J, Barrett SCH (2009). Wind of change: new insights on the ecology and evolution of pollination and mating in wind-pollinated plants. *Ann Bot (Lond)* **103**: 1515–1527.
- Goudet J (1995). FSTAT (Version 1.2). A computer program to calculate F-Statistics. *J Hered* **86**: 485–486.
- Guillot G, Mortier F, Estoup A (2005). GENELAND: a computer package for landscape genetics. *Mol Ecol Notes* **5**: 712–715.
- Honnay O, Bossuyt B, Jacquemyn H, Shimono A, Uchiyama K (2008). Can a seed bank maintain the genetic variation in the above ground plant population? *Oikos* **117**: 1–5.
- Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009). Inferring weak population structure with the assistance of sample group information. *Mol Ecol Res* **9**: 1322–1332.
- Ingvarsson PK, Olsson K (1997). Hierarchical genetic structure and effective population sizes in *Phalacrus substriatus*. *Heredity* **79**: 153–161.

- Jombart T (2008). ADEGENET: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**: 1403–1405.
- Jombart T, Devillard S, Balloux F (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet* **11**: 94.
- Jombart T, Pontier D, Dufour A-B (2009). Genetic markers in the playground of multivariate analysis. *Heredity* **102**: 330–341.
- Jorde PE, Ryman N (2007). Unbiased estimator for genetic drift and effective population size. *Genetics* **177**: 927–935.
- Knowles P, Perry DJ, Foster HA (1992). Spatial genetic structure in two tamarack [*Larix laricina* (Du Roi) K. Koch] populations with differing establishment histories. *Evolution* **46**: 572–576.
- Levin DA (1981). Dispersal versus gene flow in plants. *Ann Missouri Bot Gard* **68**: 233–253.
- Lortie CJ, Munshaw M, DiTomaso J, Hierro JL (2010). The small-scale spatiotemporal pattern of the seedbank and vegetation of a highly invasive weed, *Centaurea solstitialis*: strength in numbers. *Oikos* **119**: 428–436.
- Marshall TC, Slate J, Kruuk LEB, Pemberton JM (1998). Statistical confidence for likelihood-based paternity inference in natural populations. *Mol Ecol* **7**: 639–655.
- Mimura M, Barbour RC, Potts BM, Vaillancourt RE, Watanabe KN (2009). Comparison of contemporary mating patterns in continuous and fragmented *Eucalyptus globulus* native forests. *Mol Ecol* **18**: 4180–4192.
- Mörchen M, Cuguen J, Michaelis G, Hanni C, Saumitou-Laprade P (1996). Abundance and length polymorphism of microsatellite repeats in *Beta vulgaris* L. *Theor Appl Genet* **92**: 326–333.
- Oden NL, Sokal RR (1986). Directional autocorrelation: an extension of spatial correlograms to two dimensions. *Syst Zool* **35**: 608–617.
- Owen FV (1942). Inheritance of cross- and self-sterility and self-fertility in *Beta vulgaris*. *J Agric Res* **64**: 679–698.
- R Development Core Team (2009). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. <http://www.R-project.org/>; ISBN: 3-900051-07-0.
- Ran Z, Michaelis G (1995). Mapping of a chloroplast RFLP marker associated with the CMS cytoplasm of sugar beet (*Beta vulgaris*). *Theor Appl Genet* **91**: 836–840.
- 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*). *Mol Ecol Notes* **4**: 243–245.
- Ritland K (2002). Extensions of models for the estimation of mating systems using n independent loci. *Heredity* **88**: 221–228.
- Rutledge LY, Garroway CJ, Loveless KM, Patterson BR (2010). Genetic differentiation of eastern wolves in Algonquin Park despite bridging gene flow between coyotes and grey wolves. *Heredity* **105**: 520–531.
- Saarinen EV, Austin JD, Daniels JC (2010). Genetic estimates of contemporary effective population size in an endangered butterfly indicate a possible role for genetic compensation. *Evol Appl* **3**: 28–39.
- Sester M, Delanoy M, Colbach N, Darmency H (2004). Crop and density effects on weed beet growth and reproduction. *Weed Res* **44**: 50–59.
- Sester M, Dürr C, Darmency H, Colbach N (2006). Evolution of weed beet (*Beta vulgaris* L.) seed bank: quantification of seed survival, dormancy, germination and pre-emergence growth. *Eur J Agron* **24**: 19–25.
- Sester M, Dürr C, Darmency H, Colbach N (2007). Modelling the effects of cropping systems on the seed bank dynamics and the emergence of weed beet. *Ecol Model* **204**: 47–58.
- Shimono A, Ueno S, Tsumura Y, Washitani I (2006). Spatial genetic structure links between soil seed banks and above-ground populations of *Primula modesta* in subalpine grassland. *J Ecol* **94**: 77–86.
- Smouse PE, Long JC, Sokal RR (1986). Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Syst Zool* **35**: 627–632.
- Templeton AR, Levin DA (1979). Evolutionary consequences of seed pools. *Am Nat* **114**: 232–249.
- Van Dijk H (2004). Gene exchange between wild and crop in *Beta vulgaris*: how easy is hybridization and what will happen in later generations?. In: Den Nijs HCM, Bartsch D and Sweet J (eds). *Introgression from Genetically Modified Plants into Wild Relatives and its Consequences*. CABI publishers, Inc.: Oxfordshire, UK, pp 53–69.
- Vekemans X, Hardy OJ (2004). New insights from fine-scale spatial genetic structure analyses in plant populations. *Mol Ecol* **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. *Theor Appl Genet* **104**: 688–697.
- Vitalis R, Glémin S, Olivieri I (2004). When genes go to sleep: the population genetic consequences of seed dormancy and monocarpic perenniality. *Am Nat* **163**: 295–311.
- Waples RS (1989). A generalized approach for estimating effective population size from temporal changes in allele frequency. *Genetics* **121**: 379–391.
- Waples RS, Do C (2010). Linkage disequilibrium estimates of contemporary N_e using highly variable genetic markers: a largely untapped resource for applied conservation and evolution. *Evol Appl* **3**: 244–262.
- Whitlock MC, McCauley DE (1990). Some population genetic consequences of colony formation and extinction: genetic correlations within founding groups. *Evolution* **44**: 1717–1724.

Supplementary Information accompanies the paper on Heredity website (<http://www.nature.com/hdy>)