Molecular Ecology (2007) **16**, 3801–3813 doi: 10.1111/j.1365-294X.2007.03448.x

Long distance pollen-mediated gene flow at a landscape level: the weed beet as a case study

STÉPHANE FÉNART,*‡ FRÉDÉRIC AUSTERLITZ,† JOËL CUGUEN* and JEAN-FRANÇOIS ARNAUD*
Laboratoire de Génétique et Évolution des Populations Végétales, UMR CNRS 8016, FR CNRS 1818, Bâtiment SN2, Université des Sciences et Technologies de Lille — Lille 1, F-59655 Villeneuve d'Ascq cedex, France, †Laboratoire Écologie, Systématique et Évolution, UMR CNRS-UPS-ENGREF 8079, Bâtiment 360, Université Paris-Sud, F-91405 Orsay cedex, France

Abstract

Gene flow is a crucial parameter that can affect the organization of genetic diversity in plant species. It has important implications in terms of conservation of genetic resources and of gene exchanges between crop to wild relatives and within crop species complex. In the *Beta vulgaris* complex, hybridization between crop and wild beets in seed production areas is well documented and the role of the ensuing hybrids, weed beets, as bridges towards wild forms in sugar beet production areas have been shown. Indeed, in contrast to cultivated beets that are bi-annual, weed beets can bolt, flower and reproduce in the same crop season. Nonetheless, the extent of pollen gene dispersal through weedy lineages remains unknown. In this study, the focus is directed towards weed-to-weed gene flow, and we report the results of a pollen-dispersal analysis within an agricultural landscape composed of five sugar beet fields with different levels of infestation by weed beets. Our results, based on paternity analysis of 3240 progenies from 135 maternal plants using 10 microsatellite loci, clearly demonstrate that even if weedy plants are mostly pollinated by individuals from the same field, some mating events occur between weed beets situated several kilometres apart (up to 9.6 km), with rates of interfield-detected paternities ranging from 11.3% to 17.5%. Moreover, we show that pollen flow appears to be more restricted when individuals are aggregated as most mating events occurred only for short-distance classes. The best-fit dispersal curves were fat-tailed geometric functions for populations exhibiting low densities of weed beets and thin-tailed Weibull function for fields with weed beet high densities. Thus, weed beet populations characterized by low density with geographically isolated individuals may be difficult to detect but are likely to act as pollen traps for pollen emitted by close and remote fields. Hence, it appears evident that interfield pollen-mediated gene flow between weed beets is almost unavoidable and could contribute to the diffusion of (*trans*)genes in the agricultural landscape.

Keywords: agricultural landscape, long-distance gene flow, microsatellites, parentage analysis, pollen dispersal, weed beets

Received 23 November 2006; revision received 14 March 2007; accepted 4 June 2007

Introduction

Genetic features of plant populations have important implications for the management and conservation of genetic resources. In particular, gene flow is a crucial parameter that can affect the organization of genetic diversity of cultivated or related landraces and wild relatives in various ways (Gepts & Papa 2003; Papa & Gepts 2003; Stewart *et al*. 2003). While gene flow between crops and related wild species has occurred since the beginning of agriculture (Ellstrand *et al*. 1999), more recent scientific attention to this process is motivated by concerns about the movement of transgenes (Ellstrand 2003a; Armstrong *et al*. 2005). To what extent the escape of

Correspondence: Jean-François Arnaud, Fax: +33 3 20 43 69 79; E-mail: jean-francois.arnaud@univ-lille1.fr
‡Present address: Laboratory for Plant Biology and Nature Management, Vrije Universiteit Brussel, Pleinlaan 2, B-1050 Brussels, Belgium.

genetically modified (GM) traits from crops to weedy and wild relatives will be a significant problem remains largely a matter of conjecture. Additional data are needed to evaluate rates and patterns of gene flow within crop and weed complexes, especially since several studies produced some evidences that crop-to-wild gene flow, even with conventional (i.e. non-GM) crops, has caused increased weediness in seven (and extinction of two) wild relatives of the world's 13 most important crops (see Ellstrand *et al*. 1999; Stewart *et al*. 2003 for reviews). In this context, studies dealing with transgene escape risk assessment have to consider cautiously the role played by such weedy hybrids in the spreading of transgenes in an agricultural landscape.

Plants in the *Beta vulgaris* complex are of immediate concern for the potential effects of gene flow between domesticated plants and wild relatives, because wild, crop and weed forms are all interfertile, wind pollinated and sympatric in Europe (Boudry *et al*. 1993; Letschert 1993; Gray & Raybould 1998; Bartsch *et al*. 1999). In addition, GM herbicide-tolerant or virus-resistant lines are already produced and potentially marketed (Mannerlöf *et al*. 1997). While sugar beet production fields are principally found in northern European countries, sugar beet seeds are mainly produced in Southwestern France or Northern Italy (Bartsch *et al*. 1999). In seeds' production fields, four lines of diploid male-sterile seed bearers are framed by two lines of pollen donors (classically tetraploid but increasingly diploid). To avoid contamination by wild pollen donors, all wild beets within 1000 m around the fields must be destroyed (European Council Directive 2002/54/EC). Despite these precautions, recurrent crop-wild hybridization events lead to the formation of a weedy form of beet which is then found in the seed lots sown in Northern Europe. Weed beets flower within the crop season and produce weed beet populations inside the production fields, unlike sugar beet cultivars which are biennial and need vernalization to flower. Weed beets have been identified as a serious agronomic problem for over 30 years in Europe (Longden 1989; Brants & Hermann 1998; Desplanque *et al*. 2002; Bartsch *et al*. 2003; Ellstrand 2003b). These weedy hybrid forms are now the focus of studies because of their potential to act as an indirect route for (*trans*)gene transfer from crops into natural populations of sea beet (Treu & Emberlin 2000; Arnaud *et al*. 2003; Bartsch *et al*. 2003; Viard *et al*. 2004) or towards other weed beet populations present in nearby fields (Viard *et al*. 2002).

Some evidence supports the existence of seed movements from crop fields to wild populations (e.g. Arnaud *et al*. 2003; Viard *et al*. 2004). Alternatively, natural history and empirical studies suggest that pollen movement is the predominant vehicle of gene flow in many plant species (Levin & Kerster 1974; Ellstrand 1992; Ellstrand 2003a). Consequently, studying contemporary gene flow mediated by pollen dispersal is of crucial importance, especially in the context of intercrop gene flow for crop species where engineered genes have been introduced (Ellstrand 2003a).

Model-based methods to estimate gene flow infer the effective number of migrants exchanged per generation (N_em) from the magnitude of genetic differentiation between populations (e.g. F_{ST} see Weir & Hill 2002 for a review). Several studies investigated patterns of gene flow of several crop species using patterns of genetic differentiation between cultivated and weedy plants (Whitton *et al*. 1997; Bartsch *et al*. 1999; Jenczewski *et al*. 1999; Papa & Gepts 2003; Viard *et al*. 2004). However, these methods can only provide long-term averages of gene flow estimates over a number of generations, and are inadequate for studies of contemporary gene movement (Sork *et al*. 1999). This is especially important in the light of the life-history features of weed beets. Indeed, weed beet populations originate from multiple contamination events and can be considered as a pool of spatially and temporally genetically distinct individuals coming from several sources within the seed bank of a cultivated beet field (Arnaud *et al*. 2003; Viard *et al*. 2004). In addition, weed beet lifespan is limited to one reproductive season and, depending on crop rotation and careless farming, only a subsample of this long-lived seed bank can germinate and flower (Viard *et al*. 2002). As such, an average gene flow (Nm) inferred from F-statistics may not apply for weed beet populations because of a complex dynamics and long-term history of weedy populations.

Recent developments of Bayesian approaches in population genetics may offer promising alternatives to precisely detect migrants, hybrids or admixed individuals using assignment tests (e.g. Kuroda *et al*. 2006; reviewed in Beaumont & Rannala 2004). Another useful way to study contemporary patterns of gene flow may also be best addressed through instantaneous methods based on parentage analysis, especially when several highly polymorphic markers are used (Burczyk & Chybicki 2004). Various methods of parentage analysis are now available, based on different assumptions or statistical methods (discussed in Jones & Ardren 2003). While paternity-analysis methods were initially developed to assess whether or not a male could be the father of an offspring, they are now widely used as methods of direct estimation of the distribution of pollen-dispersal distances (Devlin *et al*. 1988; Adams *et al*. 1992; Smouse & Meagher 1994; Burczyk et al. 2002; García *et al*. 2005; Robledo-Arnuncio & Gil 2005; Goto *et al*. 2006; Ishihama *et al*. 2006; Byrne et al. 2007). The simplest method is the simple-exclusion analysis, which consists in sampling seeds on various mother plants and studying for each of these seeds its genetic compatibility with the surrounding potential male parents (Streiff *et al*. 1999). This method has however, two main drawbacks: it cannot account for potential genotyping errors and the significance

of the attribution of paternity is not assessed. This yielded the development of maximum-likelihood methods, which along with a simulation procedure, allowed to assign categorically an offspring to a given father, accounting for genotyping error rate and providing a significance level (Marshall *et al.* 1998). Alternative to this categorical assignment method, the new mating model proposed by Oddou-Muratorio *et al.* (2005), derived from the neighbourhood model of Burczyk *et al.* (2002), is focused on the estimation of the pollen-dispersal distribution. This approach is based on a fractional parentage allocation method that computes for each offspring a probability of paternity for each candidate male parent (Devlin et al. 1988; Jones & Ardren 2003). Then, mean pollen dispersal is estimated by fitting families of dispersal curves to the observed paternity assignment that imply outcrossed mating events.

In an agricultural context, direct assessment of pollen dispersal has been classically performed into artificially designed experimental plots. Two serious limitations are typical of such experiments. (i) They are spatially and temporarily limited, and (ii) the longest observed dispersal distance corresponds to the size of the experimental field. Analyses of pollen dispersal within major crop species such as oilseed rape (Scheffler *et al*. 1993), potato (McPartlan & Dale 1994), cotton (Llewellyn & Fitt 1996), and sugar beet (Saeglitz *et al*. 2000) have been conducted on fields of about 1 ha or smaller. Extrapolation of the results to larger distances is possible through modelling, but accurate predictions are only possible when the population is continuous and pollen-dispersal parameters are spatially homogeneous (Lavigne *et al*. 1996; Tufto *et al*. 1997). However, modelling of pollen-mediated gene flow across a discontinuous landscape with roads, edges and fields grown with various crops seems (so far) to be poorly reliable (Lavigne *et al*. 2002). The few studies that have investigated the possibility of pollen-dispersal events at the landscape level for cultivated species have all reported very long-distance pollen flow (Kirkpatrick & Wilson 1988; Klinger *et al*. 1991; Watrud *et al*. 2004).

This study focused on paternity and pollen movement in an agricultural landscape of about 5000 ha, with five sugar beet fields having different levels of infestation by weed beets. Because the study area is not located close to the shoreline where wild sea beet populations are typically encountered (Viard *et al*. 2004; Fievet *et al*. 2007), the focus is only directed towards the study of weed-to-weed gene flow potentially taking place between sugar beet fields and does not concern the occurrence of gene flow from cultivated to wild sea beets. This sampling area was free from natural or anthropogenic obstacles to pollen flow and was then appropriate to perform a large-scale analysis of pollen dispersal in a real-life agricultural context, without being limited by the size of an experimental field. Moreover, these populations of weed beets showed different patterns of density and spatial structuring of individuals, thus allowing an estimation of the impact of different population features on pollen dispersal. We hypothesize that (i) given that *B. vulgaris* is a wind-pollinated species, pollen dispersal should occur over large geographical distances within and among sugar beet fields, and that (ii) strong clustering of weed beets within patches of high densities should reduce the scale over which pollen dispersal occurs. To address these issues, we first performed a large-scale paternity analysis to identify effective pollen dispersal using the classical maximum likelihood-based method of Marshall *et al*. (1998). In a second step, we applied the nonlinear statistical model recently developed by Oddou-Muratorio *et al*. (2005) to estimate the parameters of the outcrossed pollen-dispersal distribution under different models of dispersal curves, along with the selfing rate.

Materials and methods

Study sites and sampling design

Weed beet individuals were sampled during summer 2003 from five sugar beet fields (A, B, C, D and E), located in a sugar beet production area in Northern France and distant from around 70 km from the coastline (see Fig. 1). After thorough searching in the area indicated on the map, only these five sugar beet fields were characterized by the occurrence of weedy individuals. These five fields were located from 500 m to 10 km apart and we have identified neither natural nor anthropogenic obstacles to pollen flow.

Populations A and B were found in two neighbouring fields (located 400 m apart). They were composed of individuals spaced out across the field and showing low levels of infestation (0.05 weed beet individuals/m^2). We performed a quasi-exhaustive sampling of these two populations (40 individuals out of 50 in each population). A total of 40 weed beet individuals were sampled within population C, presenting a density of 0.2 individuals/m^2 for a total of 150 individuals and showing a slight micro-geographical clustering of individuals. For each of populations A, B and C, 20 individuals were randomly chosen out of 40 for mating-system analysis, and 24 offspring per individual were used for progeny analyses. Population D was the densest (0.3 individuals/m^2 for a population size of around 330 weed beet individuals) and showed a pronounced spatial clustering of individuals into five well-delimited patches (see Fig. 1). As such, a larger offspring sample size was chosen for population D. In this population, a total of 75 maternal plants and their progenies (24 seeds per individual) were sampled among the five patches (20 maternal plants in the largest patch, 25 in the second largest, and 10 within each of the three smaller patches). As we

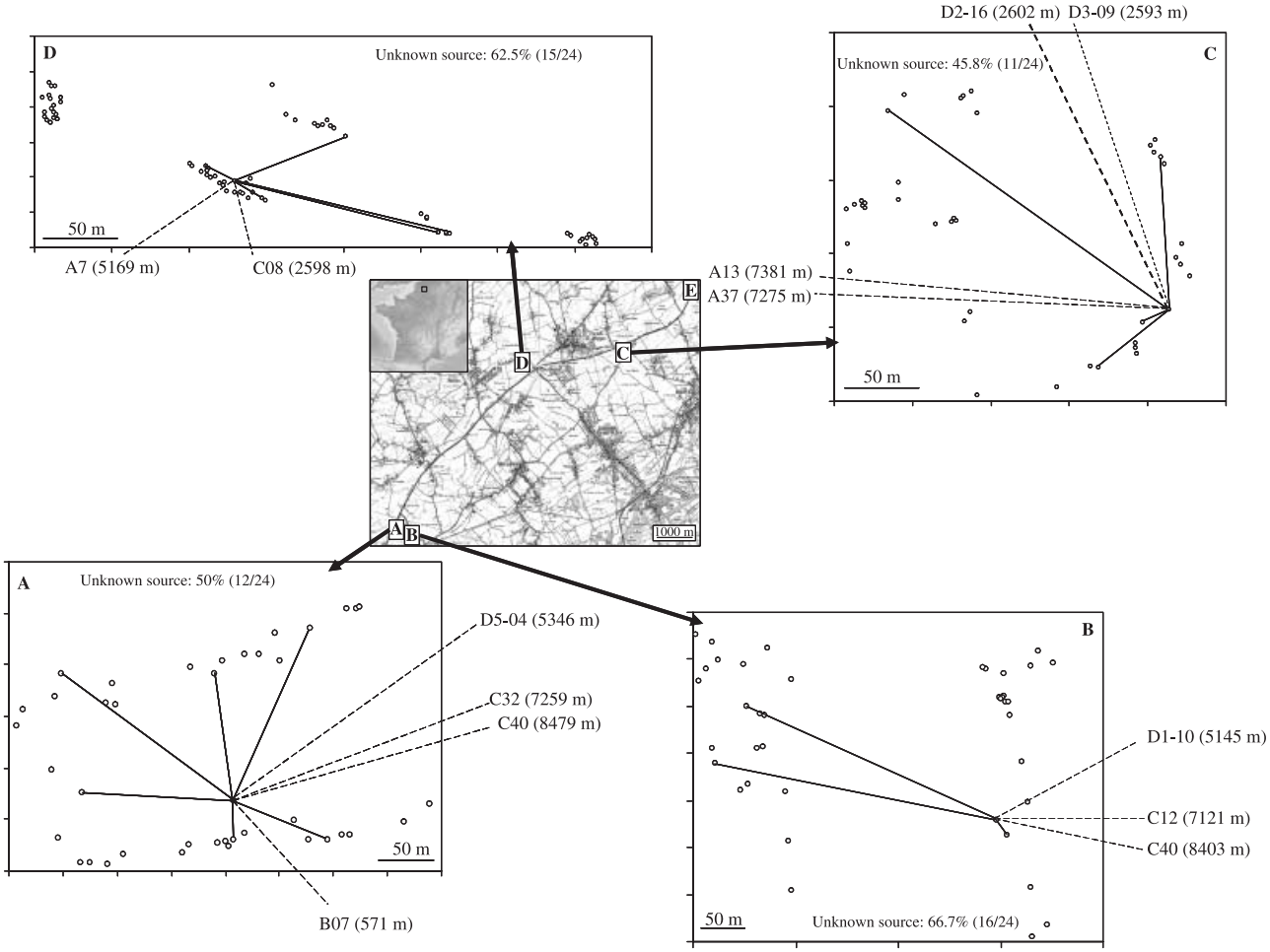

Fig. 1 Geographical location of sampled weed beets populations in the French sugar beet production area. Examples of mating events are also indicated for four seed parents (A33, B16, C10 and D2-16) sampled in population A, B, C and D, respectively. The width of the solid lines indicates the number of pollination events: from one (thinnest line) to three (thickest line) individuals from the same population; dashed lines indicate pollination events from identified pollen donors from another population. The proportion of progeny issued from mating with unknown pollen donors ('unknown source') is also indicated for each of the four seed parents. These unknown pollen donors can be either nonsampled adults from the studied populations or adults from nonsampled populations located outside the surveyed area.

only sampled seed-balls that were still attached to the female stems, the maternity of each seed was known. For the fifth population (E), 24 adults were sampled out of 100 located in a single patch along the field. Global positioning system (GPS) coordinates of all sampled adults were recorded in order to compute pairwise geographical distances between individuals.

Molecular typing

Extraction and purification of total DNA from the 219 sampled adults and the 3240 offspring were performed using a DNeasy 96 Plant Kit, following the standard protocol for isolation of DNA from dried plant leaf tissue, outlined in the DNeasy 96 Plant protocol handbook (QIAGEN Inc.).

All individuals were genotyped at 10 nuclear microsatellite loci: BVM3 (Mörchen *et al*. 1996), GAA1, GTT1, GCC1, CAA1 (Viard *et al*. 2002), SB04, SB06, SB07, SB15 (Richards *et al*. 2004) and FDSB1027 (McGrath *et al*. 2007). These 10 microsatellites were amplified in two multiplexed polymerase chain reactions (PCR). The first PCR multiplex was performed in a 10.5-μL volume mix as follows: 25 ng of DNA template, 1 μL of buffer 10× (PerkinElmer), 2.9 mM $MgCl_2$, 0.2 μg/μL of bovine serum albumin (BSA), 2% of dimethyl sulfoxide (DMSO), 0.1 μM of each primer for loci GTT1, BVM3, CAA1 and FDSB1027 and 0.05 μM of each primer for loci GCC1, 290 μM of each dNTP and 0.9 U/μL of hot start *Taq* polymerase (Ampli*Taq* Gold, PerkinElmer). The second multiplex was amplified in the same conditions but with 0.1 μM of each primer for loci GAA1 and SB04, and 0.05 μM of each primer for loci SB06, SB07 and SB15.

PCR was carried out on a 9700 thermal cycler (PerkinElmer) under the following conditions: 5 min denaturing at 94 °C, 45 s denaturing at 94 °C, 45 s annealing at 54 °C and 45 s extension at 72 °C and a final extension step at 72 °C for 10 min, after 40 cycles. After mixing 2.5 μL of PCR product in a mix with 10 μL of Hi-Di Formamide (Applied Biosystems) and 4% of GeneScan-500LIZ size standard (Applied Biosystems), PCR products were visualized on a 3100 Genetic Analyser (Applied Biosystems). Raw data of electrophoresis obtained were read using GENEMAPPER version 3.7 (Applied Biosystems). Individuals with doubtful genotypes (i.e. with missing data or when mismatches occurred between mothers and offspring) were genotyped a second time at all loci and were excluded from further analyses if the mismatch still occurred. Thus, this procedure strongly minimized the rate of genotyping scoring error.

Data analyses

Paternity assignment. Inferences based on exclusion probabilities sometimes can be misleading (Marshall *et al*. 1998), particularly when many potential sires are closely related to each other (Double *et al*. 1997). Therefore, paternity assignment was performed using the maximum likelihood-based method of Marshall *et al*. (1998), implemented in the program CERVUS version 2.0. As beets are hermaphroditic, all adults, including the sampled mothers, were considered as potential fathers for each offspring. A two-phased procedure (i.e. a first step of simulation and a second step of paternity assignment) was used to perform paternity analyses. A total of 20 000 offspring genotypes were simulated with an estimated sampling fraction of 50% of the entire adult population in this area, with a rate (e_g) of genotyping errors for the offspring set at 0.01. However, we used an error rate (e_l) of zero used to calculate the likelihoods in both the simulation and the analysis of the real data, that is that no mismatch between paternal and offspring genotypes were allowed. The paternity analyses were performed by using all loci and choosing the 95% confidence level. The multilocus logarithm of odds (LOD) scores for paternity of potential fathers were then calculated given the assumed genotyping error rate of zero (e_l). For this, CERVUS calculates a statistical test (Δ) from the differences between the highest and the second highest LOD scores and compares it to the critical value of Δ (Δ_{crit}) obtained from the simulation phase. Only the potential fathers with $\Delta > \Delta_{crit}$ are considered to be the true fathers of the analysed offspring (see Marshall *et al*. 1998 for details).

Using this approach, as recommended by Morrissey & Wilson (2005), consisting in using a lower error rate in calculating likelihoods on the real data set than in simulating offspring genotypes to estimate Δ_{crit}, one can expect to find a lower assignment rate than the one expected by simulations, but the performed assignments are likely to be more reliable. In other words, we promote quality instead of quantity by minimizing the risk of type I error, that is assignment of a wrong father to a given offspring, while the true father is either another sampled individual or a nonsampled individual (Oddou-Muratorio *et al*. 2003).

For each offspring, the likelihood-based parentage analysis produced three possible alternative outcomes. (i) Paternity could not be assigned to one of the sampled adults, either because the male parent was outside the study area or was one of the nonsampled adults within the fields. The cases where two or more adults were compatible with the offspring, but with a difference in LOD score too low to attribute paternity to the most likely parent ($\Delta < \Delta_{crit}$), were also included within this group. (ii) Paternity was attributed to the mother, allowing us to estimate the selfing rate of each mother. (iii) Paternity was most likely attributed to another sampled adult and it was then possible, knowing the position of both mother and assigned father, to calculate their physical distance. In this study, the focus is directed towards the assessment of gene exchanges. Consequently, the third category allowing the knowledge of father plants is the most informative because it provides a distribution of effective pollen dispersal occurring in different distance classes (e.g. Ishihama *et al*. 2006; Byrne *et al*. 2007). This distribution can be compared to the distribution of expected mating events under random-mating by calculating the rate of female–male pairs observed in each distance class (e.g. Hardy *et al*. 2004).

Pollen-dispersal distribution. Independently from the paternity analysis performed with CERVUS, we used the nonlinear statistical model recently developed by Oddou-Muratorio *et al*. (2005), which allows the estimation of the pollen-dispersal distribution. This method is a parametric method, which assumes that the allopollen disperses according to a given dispersal kernel characterized by one or several parameters. These parameters are estimated directly from the genotypic data on mothers, seeds and potential fathers, through a maximum-likelihood procedure that is based on fractional assignment of paternity. The parameters of the dispersal kernel are estimated jointly with the rate of unassigned paternity (u) (i.e. due to nongenotyped pollen donors located inside or outside the populations) and the selfing rate (s). With this method, several families of two-parameter dispersal functions can be tested. The families of dispersal curves tested here were the: (i) normal (ii) exponential (iii) exponential power (Clark 1998) (iv) geometric and (v) Weibull (Weibull 1951; Tufto *et al*. 1997) distributions. Details of mathematical expressions and behaviour of each family of curves can be found in Austerlitz *et al*. (2004). It must be emphasized

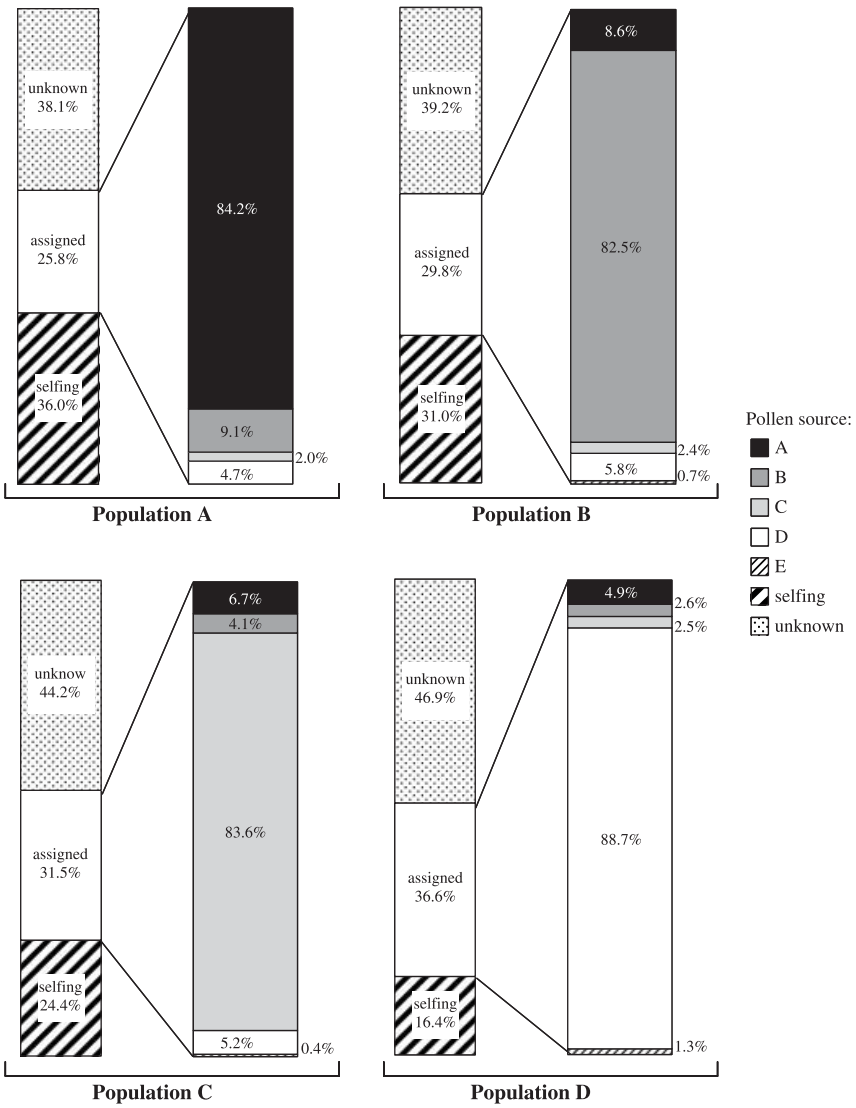

Fig. 2 Levels of pollen immigration into weedy populations A, B, C and D from pollen donors (assigned as fathers) located in each population. Rates of unassigned paternity (unknown) and of selfing are also presented for each population. Results are based on a total of 3240 progenies: 480 progenies from 20 maternal plants for populations A, B and C, 1800 progenies from 75 maternal plants for population D.

that the estimated dispersal kernel only accounts for the outcrossed mating events (see Oddou-Muratorio *et al*. 2005). Once dispersal kernel was established, a mean dispersal pollen distance (δ) was estimated from outcrossed events from equations described in Austerlitz *et al*. (2004); in a second step, the mean pollen-dispersal distance corrected for self-mating events was simply calculated as $(1–s)\delta$. Note also that the nonsampling of some nearby potential fathers will not affect the estimation of the dispersal kernel, since it is based only on the relative contribution of the different potential sampled fathers within the fraction of the pollen cloud of the mothers that is issued from these fathers. To estimate the goodness-of-fit of each family of curves to the data, we used the Akaike's information criterion (AIC, see Akaike 1974), which is more suitable than the log-likelihood when models being compared do not present the same number of parameters.

Results

Paternity analysis

A total of 117 alleles was detected and each adult individual possessed its own multilocus genotype. The cumulative exclusion probability (EP), corresponding to the probability to exclude a wrong father using the multilocus genetic information (Jamieson & Taylor 1997), was 0.997, suggesting that these loci were highly suitable for paternity analysis.

Some examples of detected mating events are shown in Fig. 1. Results of paternity analysis are summarized in Fig. 2 for each population. First, paternity remained unsolved for a large portion of progeny sampled (ranging from 38.1% to 46.9%), probably in part because of the highly conservative conditions that we assumed. Self-fertilization

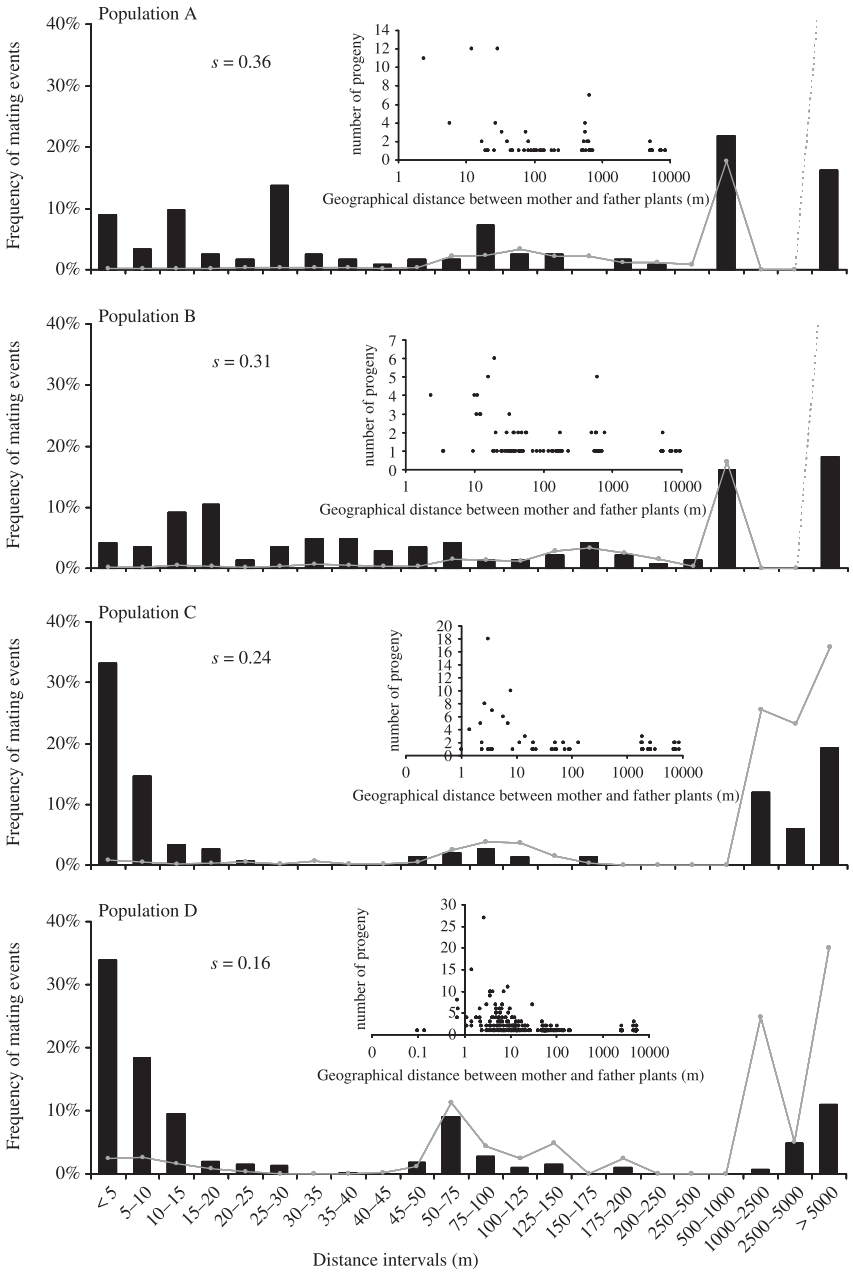

Fig. 3 Proportions of mating events assessed by paternity assignment for each class of distance between mates in each weedy population. Only results from outcrossing events are presented and correspond to the paternity assignment of 124 (25.8%) progenies in population A, 143 (29.8%) progenies in population B, 152 (31.5%) progenies in population c and 659 (36.6%) progenies in population d. Selfing rate (*s*) estimated by CERVUS is indicated for each surveyed population. Black columns show the frequencies of sires assigned for a given distance class, and grey lines refer to the frequencies of potential sires expected under random mating (i.e. the proportion of male–female pairs sampled in each distance class). In addition, a scatterplot of the number of progeny resulting from a specific mating combination against the pairwise distance (log-scale) between mother and father plants can also be visualized for each population.

was inferred in 16.4% to 36% of resolved paternities. Second, outcrossing events occurred principally between individuals from the same field (ranging from 82% to 89% of assigned fathers, see Fig. 2). Nonetheless, in each field, among the assigned paternities, a significant part of progeny appeared to have been sired by individuals from other fields. Populations A and B (located 400 m apart), presented the highest rate of pollen exchange (8.6% from A to B and 9.1% from B to A), but gene flow between more distant populations was also detected (Fig. 2). Overall, the cumulated rates of external assigned paternities ranged from 11.3% (population D) to 17.5% (population B).

The distribution of pollination distance (Fig. 3) also showed that pollen dispersal was more restricted in population C and D than in population A and B. We observed that mating events occurred mostly between nearby individuals in C and D but not in A and B as the distribution of assigned fathers spread over all distance intervals. Nonetheless, long-distance pollination occurred in all cases as all populations showed a substantial part of fatherhood assigned to pollen donors situated over 5000 m (> 10% of total assigned fathers). The maximum observed distance of pollination was 9.6 km and occurred between a maternal plant located in population B (individual B27) pollinated by a paternal plant in population E (individual E02).

Table 1 Best fits under all dispersal distribution families assumed (see Austerlitz *et al*. 2004; for the details on each family) for the four weed beet populations, separately and pooled together. The best-fitting family is indicated in bold in each case. Letters A, B, C and D refer to population name and 'All' means that the analysis was performed on the whole data set. Results are based on a total of 3240 genotyped seedlings: 480 progenies from 20 maternal plants for populations A, B and C, and 1800 progenies from 75 maternal seed plants for population D. AIC, Akaike's information criterion; *a* is the scale parameter that controls the extent of pollen dispersal and *b* is the shape parameter that affects the fatness of the tail of the dispersal distribution (Austerlitz *et al*. 2004)

Population	Curve	AIC	*a*	*b*	δ
A	gaussian	11374.07	47.79 (35.11–60.45)	2	42 (31–54)
	exponential	11351.95	26.18 (18.03–35.58)	1	52 (36–71)
	exp. power	11329.50	$\rightarrow 0$	0.049 (0.000–0.160)	901 (0- > 10^6)
	geometric	11329.30	$\rightarrow 0$	2.12	infinity*
	Weibull	11329.32	6.45×10^{-3}	0.09	328676
B	gaussian	12629.27	81.50 (66.89–106.55)	2	72 (59–94)
	exponential	12610.73	59.58 (45.27–78.62)	1	119 (90–157)
	exp. power	12611.03	38.01 (22.79–65.94)	0.78 (0.46–1.34)	137 (82–238)
	geometric	12609.01	170.65 (97.51–310.46)	4.74 (3.14–6.86)	197 (112–358)
	Weibull	12619.12	143.35 (91.28–247.06)	1.51 (1.11–1.89)	129 (82–223)
C	gaussian	12538.89	41.35 (35.76–47.81)	2	37 (32–42)
	exponential	12515.16	18.56 (14.76–23.32)	1	37 (30–47)
	exp. power	12509.26	1.22 (0.60–2.44)	0.43 (0.21–0.73)	62 (30–124)
	geometric	12510.16	16.64 (7.25–32.27)	2.80 (2.10–3.77)	infinity*
	Weibull	12509.03	44.13 (28.12–91.64)	1.20 (0.60–1.41)	41 (26–86)
D	gaussian	46028.21	41.73 (36.41–47.19)	2	37 (32–42)
	exponential	45994.72	22.16 (18.81–25.76)	1	44 (38–52)
	exp. power	45988.47	9.66 (7.12–12.90)	0.66 (0.51–0.87)	59 (44–79)
	geometric	45991.97	247.93 (122.97–391.54)	5.52 (4.54–6.82)	197 (98–310)
	Weibull	45986.94	57.81 (43.22–74.68)	1.57 (1.39–1.76)	52 (38–67)
All	gaussian	82768.25	48.41 (45.08–52.12)	2	43 (40–46)
	exponential	82673.75	25.03 (22.52–27.85)	1	50 (45–56)
	exp. power	82630.25	43.23 (33.77–55.28)	0.51 (0.42–0.61)	84 (66–108)
	geometric	82630.01	30.43 (23.36–39.25)	2.96 (2.63–3.36)	infinity*
	Weibull	82658.06	62.76 (51.25–78.86)	1.44 (1.28–1.58)	56 (47–72)

*The estimated *b* parameter of the geometric curve was below 3 which implies an infinite expected mean dispersal distance (δ) (see Austerlitz *et al*. 2004).

Pollen dispersal functions

Tables 1 and 2 show the best-fitting model for each population and over all populations, as well as the estimated mean (δ) and median dispersal distances, the rate of unassigned paternity (*u*), the rate of selfing (*s*) and the mean dispersal distance corrected for selfing rate. Detailed results of the parameter estimates obtained under the mating model for each family of dispersal curves are given in Table 1. The best-fitting dispersal curves were fat-tailed geometric curves for populations A and B, consistent with the long-distance pollen flow observed in the paternity analysis study for that case (Table 1). Mean pollen-dispersal distance (δ) was estimated as tending to infinity for population A and of 196 m for population B (Table 2). The estimated scale parameter (*a*) for population A tended towards zero, probably as a consequence of the limited sampling in this population. This does not affect the validity on the conclusion regarding the shape parameter (*b*) in this population. Indeed a recent study (Devaux *et al*. 2007) has shown that in cases where *a* could not be estimated, the estimation of *b* was still reliable. For populations C and D, the best-fitting dispersal curve was the Weibull distribution, with parameter *b* slightly above 1.0, implying a thin-tailed distribution and (as a consequence) few mating events between remote individuals (δ = 41.4 m and δ = 51.9 m for population C and D, respectively, see Table 2). However, when analysing pooled paternity assignments of all four weedy populations, the best-fitting dispersal pollen distribution was geometric, so quite fat-tailed, with a probability for a pollen grain moving more than 1000 m ($P_{>1000}$) of 6.48% (see Tables 1 and 2). It should be highlighted that this high value of $P_{>1000}$ was due to populations A and B, whose pollen dispersal followed geometric distributions, whereas $P_{>1000}$ values were very much smaller for populations C and D,

Table 2 Parameters of pollen dispersal obtained from the mating model analyses in each weed beet population. δ, average pollen-dispersal distance (in metres); *median* (*m*), distance over which 50% of pollen can disperse; $P_{>100}$ and $P_{>1000}$, probability of pollen dispersal over 100 m and 1000 m, respectively; *u*, rate of unassigned fathers; *s*, selfing rate. The δ (*m*) corrected for self-mating events was calculated as (1 – *s*)δ

	All	Population A	Population B	Population C	Population D
Number of families	135	20	20	20	75
Number of progeny	3240	480	480	480	1800
Mean distance between mothers	4773.74	191.23	212.42	131.35	136.72
Best-fitted model	geometric	geometric	geometric	Weibull	Weibull
δ (*m*)	Infinity*	Infinity*	196.59	41.46	51.91
median (*m*)	77.326	N.C.†	118.85	32.56	45.79
δ (*m*) *corrected for self–mating events*	Infinity	Infinity	123.85	29.60	42.20
$P_{>100}$	42.74%	N.C.†	56.93%	6.85%	9.36%
$P_{>1000}$	6.48%	N.C.†	1.72%	< 0.01%	< 0.01%
u	0.481	0.384	0.441	0.489	0.512
s	0.266	0.440	0.370	0.286	0.187

*the estimated *b* parameter of the geometric curve was below 3 (see Table 1) which implies an infinite expected mean dispersal distance (δ) (see Austerlitz *et al*. 2004); †not calculable: the *a* parameter of the geometric curve could not be estimated (the algorithm converged towards infinitely low values) thus median, $P_{>100}$ and $P_{>1000}$ could not be computed.

both characterized by higher densities of individuals. Note also that the differences in AIC values between the different curves are minimal in populations A and C, so we cannot conclude firmly on the best-fitting family of curves in these populations. Finally, the estimated proportion of seeds with unassigned fathers (*u*) and selfing rate (*s*) were similar to the results obtained using CERVUS.

Discussion

In this paper, we report the first genetic study dealing with pollen-dispersal analysis of a weedy crop-wild hybrid within a large agricultural landscape. Our study area was composed of five sugar beet fields presenting different level of infestation by weed beets. We could hence investigate pollen-mediated gene flow in weed beet within and among sugar beet fields and thus provide accurate information about pollen-dispersal distances, an important point on the ongoing debate on the issue of not only crop-to-wild, but also crop-to-weed and thereafter weed-to-weed gene flow. Indeed, in the context of co-existence of non-GM and GM crops, interfields gene flow is worthy of being studied (Ellstrand 2003a).

Crop-wild hybridization during seed production is no longer 'still to be proven' in the *Beta vulgaris* complex (Boudry *et al*. 1993; Desplanque *et al*. 1999; Alibert *et al*. 2005). The ensuing hybrid, weed beet, has already been sampled for population genetics studies dealing with crop–weed interactions (Viard *et al*. 2002) or crop-to-wild gene flow (Arnaud *et al*. 2003; Viard *et al*. 2004). Those studies have reported evidence for historical gene flow in the *Beta vulgaris* complex, in which weed beets were demonstrated to play a key role. Experimental studies dealing with pollen dispersal in beets have mainly focused on pollen flow from sugar beet seed production fields (reviewed in Treu & Emberlin 2000). The dispersal distances observed in such studies presented a large range of values (from several hundreds of meters to 2 km), and they have always been consistent with a leptokurtic decline of pollen concentrations. However, related questions to contemporary pollen flow within and among sugar beet fields presenting weed beets had not been investigated until now.

Paternity analyses were performed under highly conservative conditions and revealed a large fraction of paternity that remains unresolved for each population. More precisely, the fraction of unassigned paternity obtained with our assignment approach (Fig. 2) were larger than the one obtained with a simple paternity exclusion analyses, for which the proportions of unassigned paternity were 27.5%, 35.4%, 37.5% and 36.9% for population A, B, C and D, respectively. However, this does not constitute a drawback since we aimed at assigning paternities with the best possible confidence rather than assigning paternity for a high proportion of progeny with the possibility of a higher risk of assignment error. As a consequence, although being possible, the probability to have false-positives is quite limited for two main reasons: (i) each assignation was performed at a 95% confidence level for each individual, and (ii) each adult individual is characterized by its own multilocus genotype, excluding any possible confusion in individual identity. In particular, the inferred long-distance pollination events are not likely to be artefacts, and our results clearly showed that a significant fraction of mating events can occur between individuals located several kilometres apart. Indeed, all fields exchanged genes via long-distance pollen flow (with an observed maximum

distance between mates of 9.6 km). More precisely, the results of the paternity analysis clearly highlighted that within populations A, B and C, many mating events occurred in the distance classes between 500 m and 1000 m and over 5000 m. These observations are consistent with those of Meier & Artschwager (1938) who found viable pollen grains at an altitude of 5000 feet vertically over seeds production fields in USA, supporting that beet pollen grain is particularly resistant to desiccation and can, as a consequence, cover large distances by wind dispersal.

The mating model analysis confirmed the tendency towards long-distance dispersal, highlighting also the differences among weedy populations. Populations A and B showed the greatest likelihood of long-distance pollen dispersal, following a fat-tailed geometric distribution. In contrast, populations C and D showed thin-tailed pollen-dispersal distributions, with a higher proportion of mating events occurring between neighbouring individuals. Note that the differences in AIC values are very minimal in populations A and C, so we cannot conclude firmly on the best-fitting family of curves in these populations, while the differences are stronger in populations B and D. To the best of our knowledge, no theoretical analyses have been performed in order to state how far apart AIC values have to be in order to be significantly different and we should keep in mind that results based on difference of only 0.02 in AIC values for geometric or exponential power dispersal kernel must be cautiously interpreted (see Table 1). Nonetheless, if we compare the geometric curve, which is the most fat-tailed family of curves, with the others, the AIC values for the geometric distribution are always clearly above the AIC values of both Weibull and exponential power distributions in populations C and D. In contrast, in population A, AIC values for these three curve families are nearly identical but indicated a general trend for fat-tailed dispersal because of very low b values. Moreover, even if the accuracy of the estimates (as shown by the width of the confidence intervals) is lower in populations A and B than in populations C and D because of a lower seedling sampling effort, the fact that the best-fitting dispersal curve is geometric in both populations A and B and the congruence with the CERVUS estimates are strong indications that these differences are not just an artefact. The high difference in AIC between the geometric curve and the others in population B is also a strong indication in that direction.

From a methodological point of view, the fact that different families of curves performed better depending on the considered population highlights the necessity of considering different families of dispersal curves for a given study. Indeed having focused only on one family (e.g. the exponential power) would have made us miss the difference between populations A and B, on one side, and C and D, on the other side, since the differences are much less obvious within a given family. Also the fact that populations C and D were not exhaustively sampled is not likely to create biases for the estimates of the dispersal parameters. Indeed, the estimation method accounts for the unsampled proportion of males and the fact that these males are inside or outside the population does not have any impact on the estimation of the dispersal parameters.

These differences between weedy populations in terms of pollen dispersal may be related to a density-level effect. Indeed, populations C and D, which were presumably older (in terms of establishment time) than populations A and B, were denser and showed a clear spatial clustering of individuals. In such a dense patchy structure, mothers may be saturated by pollen from neighbouring weed beets, whereas in open areas (population A and B), the possibility for a mother receiving pollen from distant pollen donors is clearly higher. In the same way, a pollen grain emitted by a plant located in a patch may have a greater chance of being rapidly grabbed by a nearby mother than a pollen grain emitted by an isolated plant. As such, the adult density effect is probably a key factor in determining mating patterns within neighbourhoods since it conditions the amount of long-distance dispersal. In this respect, numerous studies in natural populations have documented that the spatial arrangement and local density of individual adults shape pollen-dispersal distributions by promoting short intermate distances in high-density situations where individuals are clumped (e.g. García *et al*. 2005; reviewed in Ghazoul 2005; Ishihama *et al*. 2006). Moreover, theoretical studies have shown that a strong clumping of individuals decreases the effective pollen pool size and the amount of pollen dispersal (Robledo-Arnuncio & Austerlitz 2006). In wind-pollinated species, evidences for large quantities of pollen that originated from outside a large sampling scheme were also ascribed as fat-tailed pollen flow distributions (Robledo-Arnuncio & Gil 2005; Goto *et al*. 2006; and references therein).

It is worth noting that the difference in the extent of long-distance pollen dispersal between low-density populations and high-density populations may to some extent be counterbalanced by differences in level of selfing events. Indeed, weedy populations A and B were characterized by a higher propensity to self-fertilization, as shown by selfing rate of 36% and 31%, respectively (Fig. 2). As a consequence, higher level of selfing rate is also likely to reduce the extent of pollen dispersal, even if plants are not clustered in dense patches of individuals, as is it the case for populations A and B. However, it is clear that for outcrossing events, the extent of external pollination events remains higher for low-density populations. A quite similar result was found in a bird-pollinated shrub where extensive long-distance dispersal was depicted, but more frequent internal-population mating events and less selfing occurred in the denser population (Byrne *et al*.

2007). A last point may be related to the inclusion of selfing events in the estimation of dispersal kernel. The populations with the flattest curves showed the highest selfing rate. But accounting for self-mating events in the estimation of the dispersal kernel is not likely to modify our conclusion that the dispersal curves are more fat-tailed for population A and B because the inclusion of self-seedlings would only inflate the extent of very short-distance dispersal. As a consequence, the inferred curves would become more leptokurtic, that is the likelihood of intermediate distance dispersal events is lower, and so the likelihood of long-distance events would not be decreased.

Finally, since we show here that long-distance pollination events occur in *B. vulgaris* as in many other plant species (e.g. Kirkpatrick & Wilson 1988; Klinger *et al*. 1991; Watrud *et al*. 2004; Robledo-Arnuncio & Gil 2005; Goto *et al*. 2006), a (*trans*)gene would likely be spread over long distances via pollen flow in any agro-ecosystem composed of cultivated, weedy and/or wild beets, unless a strong biocontainment strategy is designed. Indeed, as previously highlighted by Desplanque *et al*. (2002), the eventual formation of transgenic weed beets is not only possible, but almost unavoidable, whatever the molecular strategy used for genetic engineering. Regarding the sugar beet production areas we investigated, our study shows that isolated weedy plants like those found in weed populations A and B are able to receive pollen from very large distances. Thus, such isolated plants, which are quite difficult to detect and eliminate, could function quite well as pollen traps from fields located several kilometres around, and, as such, should be quite efficient in installing transgenic herbicide tolerant weed beet populations.

Acknowledgements

We are very grateful to Victoria L. Sork, Peter E. Smouse and three anonymous referees for providing helpful comments and suggestions on previous versions of the manuscript, and to Etienne K. Klein for providing us his Mathematica notebook for the estimation of the dispersal curve. We also wish to express our gratitude to Mathilde Cordellier for her technical assistance and to Licia Touzet for her linguistic help. 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.

References

Adams WT, Griffin AR, Moran GF (1992) Using paternity analysis to measure effective pollen dispersal in plant populations. *American Naturalist*, **140**, 762–780.

Akaike H (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**, 716–723.

Alibert B, Sellier H, Souvré A (2005) A combined method to study gene flow from cultivated sugar beet to ruderal beets in the glasshouse and open field. *European Journal of Agronomy*, **23**, 195–208.

Armstrong TT, Fitzjohn RG, Newstrom LE, Wilton AD, Lee WG (2005) Transgene escape: what potential for crop-wild hybridization? *Molecular Ecology*, **14**, 2111–2132.

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.

Austerlitz F, Dick CW, Dutech C *et al*. (2004) Using genetic markers to estimate the pollen dispersal curve. *Molecular Ecology*, **13**, 937–954.

Bartsch D, Lehnen M, Clegg J *et al*. (1999) Impact of gene flow from cultivated beet on diversity of wild sea beet populations. *Molecular Ecology*, **8**, 1733–1741.

Bartsch D, Cuguen J, Biancardi E, Sweet J (2003) Environmental implications of gene flow from sugar beet to wild beet—current status and future research needs. *Environmental Biosafety Research*, **2**, 105–115.

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

Boudry P, Mörchen M, Saumitou-Laprade P, Vernet P, Van Dijk H (1993) The origin and evolution of weed beets: consequences for the breeding and release of herbicide-resistant transgenic sugar beets. *Theoretical and Applied Genetics*, **87**, 471–478.

Brants I, Hermann H (1998) Herbicide tolerant sugar beet. In: *Proceedings of the 61st IIRB Congress* (ed. IIRB), pp. 195–204. Brussels, Belgium.

Burczyk J, Chybicki IJ (2004) Cautions on direct gene flow estimation in plant population. *Evolution*, **58**, 956–963.

Burczyk J, Adams WT, Moran GF, Griffin AR (2002) Complex patterns of mating revealed in a *Eucalyptus regnans* seed orchard using allozyme markers and the neighbourhood model. *Molecular Ecology*, **11**, 2379–2391.

Byrne M, Elliot CP, Yates C, Coates DJ (2007) Extensive pollen dispersal in a bird-pollinated shrub, *Calathamnus quadrifidus*, in fragmented landscape. *Molecular Ecology*, **16**, 1303–1314.

Clark JS (1998) Why trees migrate so fast: confronting theory with dispersal biology and the paleorecord. *American Naturalist*, **152**, 204–224.

Desplanque B, Boudry P, Broomberg K *et al*. (1999) Genetic diversity and gene flow between wild, cultivated and weedy forms of *Beta vulgaris* L. (Chenopodiaceae), assessed by RFLP and PCR-based methods. *Theoretical and Applied Genetics*, **98**, 1194–1201.

Desplanque B, Hautekèete N-C, Van Dijk H (2002) Transgenic weed beets: possible, probable, avoidable? *Journal of Applied Ecology*, **39**, 561–571.

Devaux C, Lavigne C, Austerlitz F, Klein EK (2007) Modelling and estimating pollen movement in oilseed rape (*Brassica napus*) at the landscape scale using genetic markers. *Molecular Ecology*, **16**, 487–499.

Devlin B, Roeder K, Ellstrand NC (1988) Fractional paternity assignment: theoretical development and comparison to other methods. *Theoretical and Applied Genetics*, **76**, 369–380.

Double MC, Cockburn A, Barry SC, Smouse PE (1997) Exclusion probabilities for single-locus paternity analysis when related males compete for matings. *Molecular Ecology*, **6**, 1155–1166.

Ellstrand NC (1992) Gene flow by pollen—implications for plant conservation genetics. *Oikos*, **63**, 77–86.

Ellstrand NC (2003a) Current knowledge of gene flow in plants: implications for transgene flow. *Philosophical Transactions of the*

Royal Society of London. Series B, Biological Sciences, **358**, 1163–1170.

Ellstrand NC (2003b) *Dangerous Liaisons? When Cultivated Plants Mate with Their Wild Relatives*. Johns Hopkins University Press, Baltimore and London.

Ellstrand NC, Prentice HC, Hancock JF (1999) Gene flow and introgression from domesticated plants into their wild relatives. *Annual Review of Ecology and Systematics*, **30**, 539–563.

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.

García C, Arroyo JM, Godoy JA, Jordano P (2005) Mating patterns, pollen dispersal, and the ecological maternal neighbourhood in a *Prunus mahaleb* L. population. *Molecular Ecology*, **14**, 1821–1830.

Gepts P, Papa R (2003) Possible effects of (trans) gene flow from crops on the genetic diversity from landraces and wild relatives. *Environmental Biosafety Research*, **2**, 89–103.

Ghazoul J (2005) Pollen and seed dispersal among dispersed plants. *Biology Review*, **80**, 413–443.

Goto S, Shimatani K, Yoshimaru H, Takahashi Y (2006) Fat-tailed gene flow in the dioecious canopy tree species *Fraxinus mandshurica* var. *japonica* revealed by microsatellites. *Molecular Ecology*, **15**, 2985–2996.

Gray AJ, Raybould AF (1998) Reducing transgene escape routes. *Nature*, **392**, 653–654.

Hardy OJ, Gonzalez-Martinez SC, Fréville H *et al*. (2004) Fine-scale genetic structure and gene dispersal in *Centhaurea corymbosa* (Azsteraceae) I. Pattern of pollen dispersal. *Journal of Evolutionary Biology*, **17**, 795–806.

Ishihama F, Ueno S, Tsumura Y, Washitani I (2006) Effects of density and floral morph on pollen flow and seed reproduction of an endangered heterostylous herb, *Primula sieboldii*. *Journal of Ecology*, **94**, 846–855.

Jamieson A, Taylor SCS (1997) Comparisons of three probability formulae for parentage exclusion. *Animal Genetics*, **28**, 397–400.

Jenczewski E, Prosperi JM, Ronfort J (1999) Evidence for gene flow between wild and culivated *Medicago sativa* (Leguminosae) based on allozyme markers and quantitative traits. *American Journal of Botany*, **86**, 677–687.

Jones AG, Ardren WR (2003) Methods of parentage analysis in natural populations. *Molecular Ecology*, **12**, 2511–2523.

Kirkpatrick MJ, Wilson HD (1988) Interspecific gene flow in Cucurbita: *C. texana* vs. *C. pepo*. *American Journal of Botany*, **75**, 519–527.

Klinger T, Elam DR, Ellstrand NC (1991) Radish as a model system for the study of engineered gene escape rates via crop-weed mating. *Conservation Biology*, **5**, 531–535.

Kuroda Y, Kaga A, Tomooka N, Vaughan DA (2006) Population genetic structure of Japanese wild soybean (*Glycine soja*) based on microsatellite variation. *Molecular Ecology*, **15**, 959–974.

Lavigne C, Godelle B, Reboud X, Gouyon P-H (1996) A method to determine the mean pollen dispersal of individual plants growing within a large pollen source. *Theoretical and Applied Genetics*, **93**, 1319–1326.

Lavigne C, Klein EK, Couvet D (2002) Using seed purity data to estimate an average pollen mediated gene flow from crops to wild relatives. *Theoretical and Applied Genetics*, **104**, 139–145.

Letschert JPW (1993) *Beta Section Beta: Biogeographical Patterns of Variation and Taxonomy*. PhD Thesis, Agricultural University of Wageningen, The Netherlands.

Levin DA, Kerster HW (1974) Gene flow in seed plants. *Evolutionary Biology*, **7**, 139–220.

Llewellyn D, Fitt G (1996) Pollen dispersal from two field trials of transgenic cotton in the Namoi Valley, Australia. *Molecular Breeding*, **2**, 157–166.

Longden PC (1989) Effects of increasing weed-beet density on sugar-beet yield and quality. *Annals of Applied Biology*, **114**, 527–532.

Mannerlöf M, Tuveson S, Steen P, Tenning P (1997) Transgenic sugar beet tolerant to glyphosate. *Euphytica*, **94**, 83–91.

Marshall TC, Slate J, Kruuk LE, Pemberton JM (1998) Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology*, **7**, 639–655.

McGrath JM, Trebbi D, Fenwick A *et al*. (2007) An open-source first-generation molecular genetic map from a sugarbeet x table beet cross and its extension to physical mapping. *Crop Science*, **47** (S1), 27–44.

McPartlan HC, Dale PJ (1994) An assessment of gene transfer by pollen from field-grown transgenic potatoes to non-transgenic potatoes and related species. *Transgenic Research*, **3**, 216–225.

Meier FC, Artschwager E (1938) Airplane collections of sugar-beet pollen. *Science*, **88**, 507–508.

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

Morrissey MB, Wilson AJ (2005) The potential costs of accounting for genotypic errors in molecular parentage analyses. *Molecular Ecology*, **14**, 4111–4121.

Oddou-Muratorio S, Houot M-L, Demesure-Mush B, Austerlitz F (2003) Pollen flow in the wildservice tree, *Sorbus torminalis* (L.) Crantz. I. Evaluating the paternity analysis procedure in continuous populations. *Molecular Ecology*, **12**, 3427–3439.

Oddou-Muratorio S, Klein EK, Austerlitz F (2005) Pollen flow in the wildservice tree, *Sorbus torminalis* (L.) Crantz. II. Pollen dispersal and heterogeneity in mating success inferred from parent-offspring analysis. *Molecular Ecology*, **14**, 4441–4452.

Papa R, Gepts P (2003) Asymmetry of gene flow and differential geographical structure of molecular diversity in wild and domesticated common bean (*Phaseolus vulgaris* L.) from Mesoamerica. *Theoretical and Applied Genetics*, **106**, 239–250.

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.

Robledo-Arnuncio JJ, Austerlitz F (2006) Pollen dispersal in spatially aggregated populations. *American Naturalist*, **168**, 500–511.

Robledo-Arnuncio JJ, Gil L (2005) Patterns of pollen dispersal in a small population of *Pinus sylvestris* L. revealed by total-exclusion paternity analysis. *Heredity*, **94**, 13–22.

Saeglitz C, Pohl M, Bartsch D (2000) Monitoring gene flow from transgenic sugar beet using cytoplasmic male-sterile bait plants. *Molecular Ecology*, **9**, 2035–2040.

Scheffler J, Parkinson R, Dale P (1993) Frequency and distance of pollen dispersal from transgenic oilseed rape (*Brassica napus*). *Transgenic Research*, **2**, 356–364.

Smouse PE, Meagher TR (1994) Genetic analysis of male reproductive contribution in *Chamaelirium luteum* (L.) Gray (Liliaceae). *Genetics*, **136**, 313–322.

Sork VL, Nason J, Campbell DR, Fernandez JF (1999) Landscape approaches to historical and contemporary gene flow in plants. *Trends in Ecology & Evolution*, **14**, 219–224.

Stewart CN Jr, Halfhill MD, Warwick SI (2003) Transgene introgression from genetically modified crops to their wild relatives. *Nature Reviews Genetics*, **4**, 804–816.

Streiff R, Ducousso A, Lexer C *et al*. (1999) Pollen dispersal inferred from paternity analysis in a mixed oak stand of *Quercus robur* L. & *Q. petraea* (Matt.) Liebl. *Molecular Ecology*, **8**, 831–841.

Treu R, Emberlin J (2000) *Pollen Dispersal in the Crops Maize (*Zea mays*), Oil Seed Rape (*Brassica Napus *ssp*. oleifera*), Potatoes (*Solanum tuberosum*), Sugar Beet (*Beta vulgaris *ssp*. Vulgaris*) and Wheat (*Triticum Aestivum*)*, p. 57. Soil Association, Bristol, UK.

Tufto J, Engen S, Hindar K (1997) Stochastic dispersal processes in plant populations. *Theoretical Population Biology*, **52**, 16–26.

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 *vs*. hot spots of hybridization over a regional scale. *Molecular Ecology*, **13**, 1357–1364.

Watrud LS, Lee EH, Fairbrother A *et al*. (2004) Evidence for landscape-level, pollen-mediated gene flow from genetically modified creeping bentgrass with *CP4 EPSPE* as a marker. *Proceedings of the National Academy of Sciences, USA*, **101**, 14533–14538.

Weibull W (1951) A statistical distribution function of wide applicability. *Journal of Applied Mechanics*, **18**, 293–297.

Weir BS, Hill WG (2002) Estimating *F*-statistics. *Annual Review of Genetics*, **36**, 721–750.

Whitton J, Wolf DE, Arias DM, Snow AA, Rieseberg LH (1997) The persistence of cultivar alleles in wild population of sunflowers five generations after hybridization. *Theoretical and Applied Genetics*, **95**, 33–40.

This work was part of the PhD thesis of S. Fénart that deals with population dynamics and genetic structure of weed beet populations. F. Austerlitz has interest in population genetics and in modelling and theoretical work on gene flow. J. Cuguen and J.-F. Arnaud are interested on the evolution of breeding system in plants and on its consequences on patterns of spatial genetic structure.
