MOLECULAR ECOLOGY

Molecular Ecology (2011) **20**, 46–55 doi: 10.1111/j.1365-294X.2010.04924.x

Influence of cryptic population structure on observed mating patterns in the wild progenitor of maize (*Zea mays* ssp. *parviglumis*)

MATTHEW B. HUFFORD, PAUL GEPTS and JEFFREY ROSS-IBARRA
Plant Sciences Department, University of California, One Shields Avenue, Davis, CA 95616, USA

Abstract

Indirect two-generation analysis of pollen flow has proven to be an effective alternative to exhaustive paternity analysis in numerous plant populations. In this investigation, the method is extended to an annual wild maize species, *Zea mays* ssp. *parviglumis* (*Poaceae*). Our analysis of mating system in *parviglumis* revealed high levels of outcrossing and higher biparental inbreeding than typically observed in grasses. Pollen dispersal analysis suggested low levels of long-distance dispersal. Given previous evidence for intrapopulation genetic structure in *parviglumis* populations, we explored the impact of cryptic population structure on estimates of mating system and pollen flow. Subpopulations inferred through spatially explicit Bayesian assignment showed markedly different values for both mating system parameters and pollen flow than the entire population. Finally, a novel method of pollen haplotype assignment revealed nonrandom mating consistent with intrapopulation structure. These results indicate *parviglumis* could be particularly susceptible to habitat fragmentation currently occurring throughout Mexico due to recent changes in land use.

Keywords: mating system, paternity analysis, POLDISP, pollen flow, teosinte, TWOGENER

Received 15 July 2010; revision received 9 September 2010; accepted 4 October 2010

Introduction

A clear picture of the evolutionary dynamics of plant populations requires an understanding of mating system and patterns of pollen flow. Mating system, a key life history trait, affects numerous population- and species-level processes (Barrett 1998, 2010; Goodwillie *et al.* 2005, 2010; Glemin & Bataillon 2009). Additionally, data on both mating system and pollen flow are required for accurate assessments of gene flow within and among populations. The development of molecular markers for numerous species has facilitated an explosion of studies characterizing both mating system and the spatial components of pollen flow in plant populations (Ashley 2010). However, marker-based estimates of mating system parameters in the absence of paternity inference have proven much more tractable than analyses of pollen flow based on paternity exclusion. Nevertheless, careful dissection of pollen flow through exclusion methods has yielded insights into patterns of mating, with implications for our understanding of long-distance dispersal (Nason *et al.* 1996; Robledo-Arnuncio & Gil 2005), population fragmentation (Aldrich & Hamrick 1998; White *et al.* 2002) and sex-specific fitness and selection (Broyles & Wyatt 1990; Hufford *et al.* 2009).

Pollen flow estimates from paternity exclusion methods require near-exhaustive sampling, limiting their application to relatively small populations (Jones *et al.* 2010). Over the last decade, the development of variance-component or two-generation methods (Smouse *et al.* 2001; Robledo-Arnuncio *et al.* 2006), in which maternal plants and progeny are subsampled, has made the estimation of pollen movement possible in even the largest of populations. However, to date these methods have been applied primarily to long-lived woody perennials (Sork *et al.* 2002; Oddou-Muratorio *et al.* 2005; De-Lucas *et al.* 2008; Klein *et al.* 2008; Larsen & Kjaer 2009), and our understanding of pollen movement in dense populations of annual species remains extremely limited.

Correspondence: Matthew B. Hufford, Fax: 530 752 4604;
E-mail: mbhufford@ucdavis.edu

Grasses are an excellent system in which to explore questions of pollen movement in dense annual populations. The census size in grass populations precludes exclusion methods and is much higher than that seen in most previous two-generation pollen flow studies. Moreover, because of their low stature, grasses may be particularly susceptible to landscape barriers to pollen dispersal, thus increasing the likelihood of cryptic genetic structure in populations. Despite the potential impacts of cryptic population structure on inference (Austerlitz & Smouse 2001; Dyer *et al.* 2004), this factor has generally been ignored in analyses of mating system.

Here, we present an analysis of mating system and pollen flow in the wild progenitor of maize, the annual grass *Zea mays* ssp. *parviglumis* (hereafter *parviglumis*). *Parviglumis* is found throughout southwest Mexico, often in large, extremely dense populations consisting of millions of individuals. Given its close relationship with maize, the population ecology of *parviglumis* is surprisingly understudied (Mondragon-Pichardo & Vibrans 2005); genetic structure has previously been observed at both regional (Moeller *et al.* 2007) and subpopulation scales (van Heerwaarden *et al.* 2010), but analyses of mating patterns and pollen movement are lacking. Preliminary inferences regarding pollen flow in *parviglumis*, however, can be drawn from investigations in maize fields. These studies suggest restricted dispersal in maize with very little outcrossing (<1%) occurring beyond 20–100 m (Ireland *et al.* 2006; Messeguer *et al.* 2006), while *parviglumis* pollen has physical characteristics that indicate a probability of dispersal over longer distances (e.g., more abundant, smaller pollen with a slower settling speed) (Aylor *et al.* 2005; Baltazar *et al.* 2005). This potential for longer-distance pollen dispersal may or may not be realized in the heterogeneous landscapes in which *parviglumis* occurs.

To investigate pollen flow in *parviglumis*, we sampled over 1000 seeds from progeny arrays and pooled individuals collected at georeferenced sampling sites in a population from the state of Jalisco, Mexico. We genotyped each individual for nine polymorphic microsatellite markers, estimated mating system parameters and tested for the presence of cryptic population substructure. We then used two-generation methods to estimate pollen flow in the entire population as well as individual subpopulations. Finally, we implemented a novel approach to paternity analysis involving Bayesian assignment of pollen haplotypes to pools of individuals collected at sampling sites across the population. To our knowledge, our study is the first to assess patterns of pollen flow in a large, extremely dense population of annual plants, the first empirical analysis of the impact of cryptic structure on estimates of pollen flow and the first to provide evidence that subpopulation structure is attributed to nonrandom mating within the population.

Materials and methods

Study site

The *parviglumis* population characterized in this investigation (Ejutla A, hereafter EJUA) is one of the larger, more continuous and genetically diverse populations in central Jalisco (Hufford 2010). EJUA is found on the hillsides above the town of Ejutla and covers approximately 55 hectares of grazed and moderately forested terrain (Fig. 1). Based on estimates of plant density and the extent of the population, it is probable that the population consists of several million individuals. An escarpment approximately bisects the population, forming an upper area stretching along both sides of the El Grullo-Ejutla Highway and a lower area at the southwest edge of the town of Ejutla. The EJUA population is bounded by heavily forested areas that exclude heliophilic *parviglumis* and intensive agricultural fields where *parviglumis* is viewed as a noxious weed. The next-nearest *parviglumis* population is located approximately 5 km to the east of EJUA.

The EJUA population was extensively mapped during March 2007, and 52 sampling sites were randomly established along virtual transects (Fig. 1a). Seed was collected in early 2008 from a focal maternal individual at each sampling site and from a pooled sample of individuals within a 10-metre radius. At each site, several environmental measurements were recorded: (i) *parviglumis* density in a 1-m^2 quadrat; (ii) canopy cover, estimated with a convex spherical densiometer; (iii) slope, measured with a clinometer; and (iv) elevation and exposure, both determined using a GPS receiver. Qualitative descriptions of surrounding vegetation and land use were also recorded for each sampling site. Seed samples were transported to Davis, California, for genetic analysis. Appropriate documentation and approval were obtained from agricultural authorities for both the export of samples from Mexico (SEMARNAT) and the import of samples into the United States (USDA-APHIS).

DNA extraction, polymerase chain reaction and genotyping

Ten seeds from the focal maternal individual and ten seeds from the surrounding pooled sample at each site (a total of 1040 samples) were randomly selected for genetic analysis. Seeds were germinated and harvested at the five-leaf stage. Genomic DNA was extracted from ground lyophilized leaf material using a modified cetyl trimethylammonium bromide protocol based on the

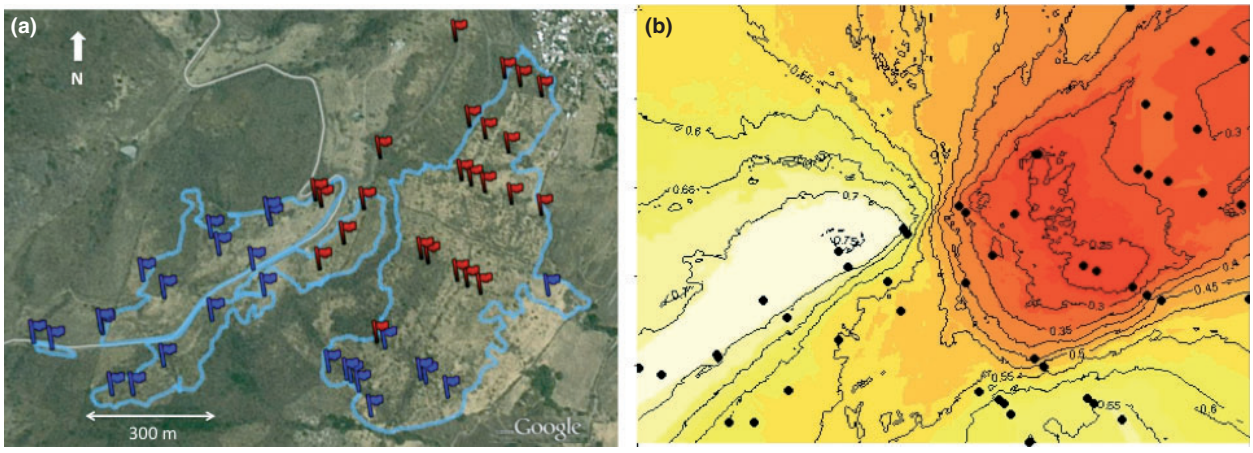

Fig. 1 EJUA population. (a) Sampling locations of maternal individuals and progeny. Patches of parviglumis are delineated with light blue lines, and blue (EJUA1) and red (EJUA2) flags indicate cryptic subpopulation membership. (b) Heat map from GENELAND analysis of probability of membership in the EJUA1 subpopulation. Areas in white have higher probability of membership in EJUA1 and contour lines correspond to 0.05% probability.

method of Saghai-Maroof *et al.* (1984). All individuals were genotyped at nine unlinked trinucleotide repeat microsatellite (SSR) markers (Table S1, Supporting information). The PCR protocol for amplification of SSR's was based on the method of Schuelke (2000), in which an M13 reverse sequence tail (TGTAAAACGACGGCCAGTATGC) is added to the 5′ end of forward primers and fluorescent labels. Each 20 μL reaction volume included 30 ng of genomic DNA, 0.2 μM dNTP (New England Biolabs), 8 pmol M13-labelled forward primer and 32 pmol reverse primer (Sigma Life Science), standard *Taq* buffer (New England Biolabs), 32 pmol M13-labelled 6-FAM, VIC or PET fluorescent dye (Sigma Life Science) and 0.5 unit of *Taq* polymerase (New England Biolabs). The PCR program consisted of 5 min at 94 °C, 30 cycles of 30 s at 94 °C, 45 s at 56 °C and 45 s at 72 °C followed by eight cycles of 30 s at 94 °C, 45 s at 53 °C and 45 s at 72 °C and a final extension at 72 °C for 10 min. Amplified fragments were diluted to 1:10 using sterile water and multiplexed according to size and fluorescent label. For SSR genotyping, 2 μL of diluted and multiplexed PCR products was added to 10 μL of Hi-Di™ formamide and 0.1 μL of GeneScan™ 500 LIZ® Size Standard (Applied Biosystems Inc.) and denatured for 3 min at 95 °C prior to analysis on an ABI 3730 instrument (Applied Biosystems Inc.). Sample genotypes were scored using the program GeneMarker® (version 1.85; SoftGenetics LLC, State College, PA, USA).

Mating system, genetic diversity and cryptic population structure

Progeny arrays from each sampling site were used to reconstruct the most likely maternal genotypes using the software MLTR 3.2 (Ritland 2002). Mating system parameters, including rates of single- (t_s) and multilocus (t_m) outcrossing, biparental inbreeding ($t_m - t_s$) and correlated paternity ($r_{p(m)}$, MLTR; $2\Phi_{FT}$, TWOGENER) were also calculated with MLTR and TWOGENER (as implemented in the software POLDISP1.0c) based on progeny arrays.

Genetic diversity in parent and offspring generations was estimated using reconstructed maternal genotypes and a single randomly chosen individual from each progeny array. The program GENALEX 6 (Peakall & Smouse 2006) was used to calculate observed (H_O) and expected (H_E) heterozygosity, the number of alleles per locus (A) and the fixation index (F).

To explore the potential interaction of genetic structure and mating system in the EJUA population, spatially explicit Bayesian assignment was carried out with maternal genotypes using the software GENELAND (Guillot *et al.* 2005a,b). The spatial prior implemented in GENELAND particularly improves assignment in instances where cryptic subpopulations are modestly diverged ($F_{ST} < 2\%$), as would be expected in continuous, dense populations of annuals. The number of subpopulations in the analysis, K, was allowed to vary between 1 and 10 and treated as a simulated variable in a correlated alleles model during Markov Chain Monte Carlo (MCMC) analysis. The MCMC chain was run for 10^5 iterations, saving 1% of iterations and discarding the first 200 as burn-in. Ten independent runs were conducted to confirm the consistency of results.

Pollen dispersal and effective population density

Pollen dispersal and population density parameters were estimated using the two-generation approaches of

Austerlitz & Smouse (2001, 2002) and Robledo-Arnuncio *et al.* (2006), as implemented in the POLDISP1.0c software package (Robledo-Arnuncio *et al.* 2007). During this process, the parameters of five pollen dispersal distributions (Gaussian, exponential, exponential power, geometric and Student's 2D*t*) were fit to the data, allowing the estimation of average effective pollen dispersal distance (δ). The fit of each model was assessed by least-square residuals, and the best model was chosen using Akaike's information criterion (AIC). Dispersal distribution parameters and population allele frequencies based on genotypes of pooled samples from sites were utilized when calculating effective density of *parviglumis* (d_e). The ratio of d_e to the average census density of *parviglumis* in 1-m^2 quadrats (d) was calculated for both the entire population and inferred subpopulations as a measure of heterogeneity in reproductive contribution of males (Oddou-Muratorio *et al.* 2005; Mimura *et al.* 2009).

Bayesian assignment of pollen haplotypes

In addition to the analysis of pollen movement via two-generation methods, we investigated the most likely origin of individual pollen haplotypes by assigning pollen to subpopulations and individual sampling sites with the program STRUCTURE (Pritchard *et al.* 2000). Assignment was based on the genotypes of pooled individuals from each sampling site. The existence of the GENELAND subpopulations identified in maternal individuals was confirmed in sampling site pools by setting $K = 2$ in STRUCTURE and implementing a location prior as described by Hubisz *et al.* (2009) (data not shown). Pollen haplotypes were generated by subtracting maternal alleles from individuals in progeny arrays using custom perl scripts. In cases where paternal alleles could not be unambiguously assigned, the allele was chosen using a random method that incorporated allele frequencies from the entire EJUA population. Pollen haplotypes were assigned to subpopulations ($K = 2$) or sampling sites ($K = 52$) that were fixed with the 'USEPOPINFO' option in STRUCTURE and run for 10^6 iterations, discarding the first 10^5 as burn-in. An additional 1000 runs confirmed that random choice of paternal alleles had no qualitative effect on the outcome (data not shown).

Results

Genetic diversity analysis and identification of cryptic population structure

Genetic diversity estimates in both the maternal and progeny generations in the entire EJUA population (Table 1) were quite similar and within the range reported in a meta-analysis of wild plant populations with similar life history (Nybom 2004). Spatially explicit Bayesian assignment revealed two cryptic subpopulations within EJUA (Fig. 1b) with fully consistent results across all runs. The number of populations along the Markov chain was plotted, and the density was always substantially greater for $K = 2$ than any other number of subpopulations (data not shown). The run with the highest likelihood value was chosen for assignment of sampling sites to the two subpopulations (EJUA1 and EJUA2). However, the assignment of sampling sites was nearly identical across each run; fewer than three sampling sites changed assignment in runs with lower likelihoods. Each subpopulation contained 26 of the 52 sampling sites, and pairwise F_{ST} between them was calculated as 1.9%. EJUA2 had higher genetic diversity and a lower fixation index than EJUA1 in both maternal and progeny generations (Table 1).

Environmental measurements

Average measurements of per cent slope, per cent canopy cover, density and elevation were calculated for both the entire EJUA population and the EJUA1 and EJUA2 subpopulations. Various differences were found between the two subpopulations: EJUA1 was situated at higher average elevation than EJUA2 (EJUA1: 1258 m, EJUA2: 1236 m; d.f. = 44.2; $P = 0.2821$) and had a signif-

Table 1 Genetic diversity of maternal and progeny generations in the EJUA population and the EJUA1 and EJUA2 subpopulations

	N	*A*	H_O	H_E	*F*
Maternal EJUA	52	6.111	0.660	0.679	0.034^{AB}
Maternal: EJUA1 subpopulation	26	6.222	0.577	0.662	0.135^{A}
Maternal: EJUA2 subpopulation	26	6.000	0.744	0.696	-0.067^{B}
Progeny EJUA	52	5.833	0.627	0.668	0.063^{AB}
Progeny: EJUA1 subpopulation	26	5.778	0.597	0.669	0.112^{A}
Progeny: EJUA2 subpopulation	26	5.889	0.658	0.667	0.014^{AB}

N, number of sampling sites; *A*, average alleles per locus; A_E, effective alleles per locus; H_O, average observed heterozygosity; H_E, average gene diversity; *F*, fixation index. Superscripted letters indicate statistically significant groups based on a Tukey–Kramer HSD test ($\alpha = 0.05$).

icantly lower per cent slope (EJUA1: 14.54%, EJUA2: 20.19%; d.f. = 49.4; $P = 0.0306$) and marginally lower *parviglumis* density (EJUA1: 38 plants/m^2, EJUA2: 106 plants/m^2; d.f. = 27.617; $P = 0.0704$).

Mating system parameters

Parviglumis outcrossed at very high rates (Table 2), but biparental inbreeding rates were higher than previously estimated in grass species (Goodwillie *et al.* 2005). Estimates of correlated paternity based on Ritland's method ($r_{p(m)}$) (2002) were lower than calculations following Austerlitz and Smouse ($2\Phi_{FT}$) (2001, 2002), consistent with previous observations indicating $r_{p(m)}$ underestimates correlated paternity at the family level (Hardy *et al.* 2004). Both biparental inbreeding and estimates of correlated paternity were higher in the full population than either of the individual subpopulations.

Pollen dispersal and effective population density

Among sibship correlated paternity was shown to decrease as intermaternal separation distance increased: a Spearman's rank correlation (ρ) provided evidence of the theoretically expected negative correlation over the distance range sampled in the EJUA population (ρ = −0.157; d.f. = 1324; $P < 0.0001$), the EJUA1 subpopulation (ρ = −0.235; d.f. = 323; $P < 0.0001$) and the EJUA2 subpopulation (ρ = −0.137; d.f. = 323; $P = 0.0134$), indicating that pollen dispersal could be estimated with two-generation methods in all three areas.

Estimates of pollen dispersal distributions differed between the full population and the subpopulations analysed individually. When the EJUA population was analysed in its entirety, the one-parameter Gaussian and the two-parameter exponential power distributions showed the best fit to the data based on least-square residuals and AIC (Table 3). However, the Gaussian distribution indicated a much shorter average pollen dispersal distance (δ = 4 m) than that calculated with the exponential power distribution (δ = 35 m), and Gaussian pollen dispersal distributions are atypical of the leptokurtic shape often seen in natural plant populations (Austerlitz *et al.* 2004). In contrast, when the EJUA population was divided into two subpopulations the least-square residuals and AIC of the two-parameter models (exponential power, geometric and Student's 2D*t*) were much lower. Of these, only the exponential power distribution provided realistic results (Table 3), as the geometric and Student's 2D*t* distributions estimated an infinite value for δ. The geometric and Student's 2D*t* distributions were not further considered

Table 2 Mating system parameters in EJUA and subpopulations. Bootstrap-based standard deviations are in parentheses

Parameter	EJUA	EJUA1	EJUA2
Single-locus outcrossing rate, t_s	0.829 (0.01)	0.839 (0.033)	0.835 (0.030)
Multilocus outcrossing rate, t_m	0.969 (0.019)	0.967 (0.014)	0.970 (0.036)
Biparental inbreeding rate, $t_m - t_s$	0.141 (0.016)	0.127 (0.024)	0.135 (0.039)
Single-locus correlated paternity, $r_{p(s)}$	0.119 (0.030)	0.107 (0.047)	0.102 (0.046)
Multilocus correlated paternity, $r_{p(m)}$	0.089 (0.011)	0.075 (0.017)	0.079 (0.013)
Effective number of pollen donors, $1/r_{p(m)}$	11.236	13.333	12.650
Global Φ_{FT}	0.0719	0.0770	0.0639
Effective number of pollen donors, $1/2\Phi_{FT}$	6.955	6.494	7.830

Table 3 Descriptive parameters of best-fit pollen dispersal kernels from KINDIST analysis of the EJUA population and the EJUA1 and EJUA2 subpopulations. The effective density (d_e) per square metre was estimated with TWOGENER and subsequently the ratio to census density (d) was calculated

Population	Model	Least-square residuals	AIC*	Scale (a)	Shape (b)	δ—Average dispersal (m)	d	d_e/d
EJUA	Gaussian	226.31	−2342.39	0.000045	–	4	81	1.14
EJUA	Exponential	1352.59	28.33	4872.52	–	$\sim 10^8$	$\sim 10^{11}$	–
EJUA	Exponential power	226.96	−2338.59	0.00052	5.36	35	$\sim 10^{11}$	–
EJUA1	Gaussian	295.13	−29.33	14142.14	–	$\sim 10^8$	10	0.27
EJUA1	Exponential	295.13	−29.33	4872.34	–	$\sim 10^8$	$\sim 10^{16}$	–
EJUA1	Exponential power	47.96	−617.87	0.00130	5.42	87	11	0.29
EJUA2	Gaussian	364.60	39.37	14135.06	–	$\sim 10^8$	10	0.09
EJUA2	Exponential	364.60	39.37	4872.52	–	$\sim 10^8$	$\sim 10^{15}$	–
EJUA2	Exponential power	63.07	−528.86	0.00052	2.80	39	10	0.09

*Akaike's information criterion (AIC) for each model; a lower AIC indicates a better fit.

because they provided little information regarding dispersal distance. The two subpopulations differed primarily in their estimates of δ, with a longer estimate of mean pollen dispersal distances in EJUA1 (87 m) than in EJUA2 (39 m). The shape of the pollen dispersal curve was also more platykurtic in EJUA1, indicating a greater amount of medium-distance dispersal. However, all of the dispersal distributions had very thin tails, suggesting a paucity of long-distance dispersal.

Effective male population density (d_e) was estimated to be much higher in the full EJUA population (81 individuals/m^2; Table 3) than in individual subpopulations (~10 individuals/m^2 in both EJUA1 and EJUA2). Likewise, the ratio of d_e to the census population density (d), a measure of heterogeneity in pollen donors, was much higher in the full population (1.14 in EJUA vs. 0.29 in EJUA1 and 0.09 in EJUA2). The higher d_e/d ratio in EJUA1 compared with EJUA2 may be attributable to the longer realized dispersal distance, a factor of lower plant density in this subpopulation, and outcrossing between individuals from different demes. Lower estimates of correlated paternity and biparental inbreeding in EJUA1 also support this result.

Bayesian assignment of pollen haplotypes

Assignment of pollen haplotypes to the EJUA1 and EJUA2 subpopulations revealed genetic structure qualitatively similar to that found in maternal individuals. When pooling probabilities of assignment to subpopulation of origin, *parviglumis* plants in EJUA2 were more likely to receive pollen from EJUA2 (probability of 0.69), but *parviglumis* in EJUA1 were equally likely to receive pollen from both subpopulations (probability of 0.48 of assignment to EJUA1). However, spatial analysis of pollen assignment across sampling sites clearly showed the groupings found in maternal individuals (Fig. 2) and revealed a high level of admixture occurring at the interface of the two subpopulations. Assignment of pollen haplotypes to the individual 52 sampling sites confirmed the low level of long-distance dispersal evident in pollen dispersal curves: when probability of assignment was averaged across pollen from a given site, 14 of 52 pollen groups had the highest probability of assignment to their site of origin. Additionally, the site of pollen origin was in the top five highest probability sites of pollen assignment in a total of 33 of 52 pollen groups (Table S2, Supporting information).

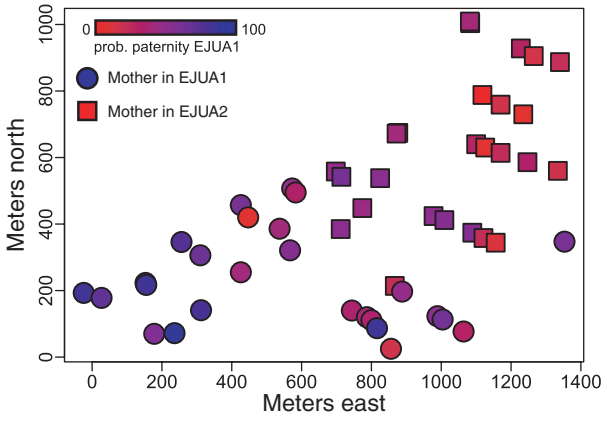

Fig. 2 Probability of assignment of pollen groups from each sampling site to subpopulation EJUA1.

Discussion

The variance-component or two-generation method for evaluating patterns of pollen flow has often been applied to populations with a census size that prohibited direct paternity analysis (e.g., De-Lucas *et al.* 2008; Mimura *et al.* 2009). In the current investigation, we have employed this type of analysis in an annual grass species, *parviglumis*, with an average metre-scale density greater than that seen in hectares of many other species similarly assessed. Our analyses with *parviglumis* have provided an opportunity to test the applicability of two-generation techniques to dense stands of annuals and assess the effects of cryptic subpopulation structure on inferences of mating system and pollen flow. Furthermore, we have implemented a novel combination of sampling technique and pollen haplotype assignment that can detect differential mating patterns across populations of high-density plants.

Assessing mating system and pollen flow in dense stands of annuals

Several interesting comparisons to previous work can be made regarding mating system based on the results of the current study. Our estimate of the outcrossing rate in *parviglumis* (0.97) is similar to those found in other wind-pollinated grass species (Ennos 1985; Macdonald & Lieffers 1991; Bush & Barrett 1993) but substantially higher than a previous approximation (0.84) in *parviglumis* based on the equilibrium selfing rate estimated from F_{IS} (van Heerwaarden *et al.* 2010). This latter disparity may be attributed to population-level differences in outcrossing rates owing to varying topography or climatic conditions but may also be attributable to a downward bias in F_{IS}-based estimates of outcrossing rates in highly outbred species (Jarne & Auld 2006; David *et al.* 2007). Rates of biparental inbreeding in this population of *parviglumis* are higher than typically observed in other grasses—nearly double the highest rate of biparental inbreeding (0.073) observed across other grass species in a recent meta-analysis (Goodwillie *et al.* 2005). This is consistent with

our low estimates of pollen dispersal distance as well as the large size of *parviglumis* seeds, which may limit long-distance dispersal of maternal alleles.

Previous studies of pollen-mediated gene flow in grasses have had to rely on measures of pollen deposition (Davis *et al.* 2004; Huang *et al.* 2004; Vogler *et al.* 2009) or the use of molecular markers (e.g., transgenes or homozygous allozymes) in which flow from donor to recipient individuals could be detected (Nurminiemi *et al.* 1998; Rognli *et al.* 2000; Wang *et al.* 2006). These techniques are obviously not optimal for quantitative analysis of pollen flow in natural populations. Our application of two-generation methods and a novel pollen haplotype assignment method to study mating system and pollen flow in *parviglumis* therefore represent a step forward in grass population ecology. Both methods infer a predominance of short-distance pollen movement, suggesting that populations of *parviglumis* may be particularly susceptible to nonrandom mating and genetic isolation owing to landscape barriers to pollen-mediated gene flow. Such dynamics were indeed observed within the EJUA population and provided an excellent opportunity to evaluate the impact of cryptic subpopulation structure on estimates of mating system parameters and pollen flow.

The importance of identifying cryptic population structure

Analysis of cryptic population structure in this study proved to be essential for understanding patterns of genetic diversity, mating system and pollen flow in the EJUA population. The markedly different results obtained when considering the population as a whole vs. those seen in the two subpopulations revealed the effects of landscape and land use on these parameters and confirmed previous concerns regarding the effects of genetic structure in two-generation methods (Austerlitz & Smouse 2001).

Our pollen haplotype assignment approach proved effective at assessing the relationship between structure in pollen and structure in maternal plants. While evidence of population substructure alone is suggestive of barriers to dispersal or mating, other factors such as selection, founder events or gene flow from nearby maize could also potentially account for observed patterns. Here, we show that pollen haplotype assignment correlates well with maternal genetic structure, implicating restricted dispersal—likely due to landscape features that limit pollen movement—as the most likely determinant of the observed structure.

The two subpopulations (EJUA1 and EJUA2) distinguished based on multilocus genetic data also differed substantially in terms of land use and landscape. Whereas EJUA2 was situated on steep slopes with little human-mediated disturbance, EJUA1 was bisected by the El Grullo-Ejutla Highway and contained two cattle corrals. The extensive grazing taking place in EJUA1 is likely the causal factor for the lower *parviglumis* density and may contribute to the higher level of inbreeding observed in this subpopulation. Moreover, a common garden experiment with *parviglumis* from this region has shown significant fitness costs in inbred material (Hufford 2010). Taken together, these data are consistent with previous assessments that grazing is a contributor to *parviglumis* population decline (Wilkes 2007). EJUA1 and EJUA2 were further distinguished by exposure: EJUA1 was found on south-facing slopes, and EJUA2 was situated primarily on east-facing slopes. The influence of prevailing east–west winds on pollen dispersal may therefore contribute to genetic differentiation between EJUA1 and EJUA2. Notably, STRUCTURE runs showed that the subpopulations identified using GENELAND were more likely than subpopulations that might be defined *a priori* based on the escarpment that bisects the population (data not shown), suggesting that visual analysis of the population alone may not always be sufficient to identify cryptic structure.

Despite their considerable genetic and physical differentiation, EJUA1 and EJUA2 were quite similar in terms of mating system parameters: outcrossing, biparental and correlated paternity rates were nearly indistinguishable between the two subpopulations, and the biparental inbreeding rate was only slightly elevated in the more dense EJUA2 subpopulation. Conversely, substantial differences were revealed in our two-generation analysis of pollen flow, with longer-distance dispersal occurring in EJUA1. These data may indicate that generalities of mating system in this species are maintained despite variation in land use or landscape, yet specifics of pollen flow dynamics may be profoundly changed.

The most striking effects of cryptic population structure in our study were observed when results from the full population and subpopulations were compared. Both the rates of biparental inbreeding and correlated paternity were inflated in the entire population relative to subpopulations. Moreover, the average pollen dispersal distance was considerably underestimated in the full population, and a biologically unrealistic dispersal model provided the best fit to the data. These differences are a result of the nonrandom mating observed when EJUA is considered as a whole. Such results should be cautionary for investigations into mating system and pollen flow in fragmented landscapes or in species such as *parviglumis* that are short statured and thus more susceptible to subtle variation in the landscape.

Implications for parviglumis conservation

In conclusion, several lessons can be learned from this study regarding the conservation of *parviglumis*. We have shown that cryptic subpopulation structure can occur in dense *parviglumis* stands even on the scale of hundred of metres and that this structure is reflected in patterns of mating. These results indicate that *parviglumis* may be particularly susceptible to the effects of habitat fragmentation currently occurring throughout Mexico owing to grazing. For instance, inbreeding and failure to purge genetic load in isolated, grazed populations such as EJUA1 could eventually lead to instances of local extinction. Populations of *parviglumis* in central Jalisco are quite isolated, so recolonization or 'rescue' by nearby populations seems unlikely.

Conservation of *parviglumis* in Mexico will, in all probability, occur within a 'working' landscape largely commensal to agriculture, as is the case in the Sierra de Manantlán Biosphere Reserve, designed to protect a related species of wild *Zea* (Jardel *et al.* 1996). With this in mind, it is necessary to understand how agricultural activities impact the genetic diversity and ecology of *parviglumis*. Our finding that grazing and variation in plant density appear to affect patterns of pollen flow should thus be an important consideration in the management of *parviglumis* populations. Finally, the likelihood of cryptic structure in low-statured species such as *parviglumis* predicates a need for wide collections across populations (particularly across environmental gradients and varied land use) when conducting *ex situ* conservation.

Acknowledgements

We thank Marie Jasieniuk and Kevin Rice for their advice throughout this project. We also thank Roberto Miranda-Medrano, Josar Medina-Garcia and Mariano Corona for assistance in Mexico, Irina Kirgiz and Amy Ohe for assistance in the laboratory and Joost van Heerwaarden and Andrew Eckert for providing feedback during data analysis. Helpful comments on a previous draft were provided by Preston Aldrich and two anonymous reviewers. This work was supported by the UC Davis IGERT for Biological Invasions (NSFDGE#0114432), the Jastro Shields Research Fellowship, the UC Davis Humanities Research Award and UC MEXUS grants to M.B.H. and by a grant from United States Department of Agriculture (2009-01864) to J.R-I.

References

Aldrich PR, Hamrick JL (1998) Reproductive dominance of pasture trees in a fragmented tropical forest mosaic. *Science*, **281**, 103–105.

Ashley MV (2010) Plant parentage, pollination, and dispersal: how DNA microsatellites have altered the landscape. *Critical Reviews in Plant Sciences*, **29**, 148–161.

Austerlitz F, Smouse PE (2001) Two-generation analysis of pollen flow across a landscape. III. Impact of adult population structure. *Genetical Research*, **78**, 271–280.

Austerlitz F, Smouse PE (2002) Two-generation analysis of pollen flow across a landscape. IV. Estimating the dispersal parameter. *Genetics*, **161**, 355–363.

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

Aylor DE, Baltazar BM, Schoper JB (2005) Some physical properties of teosinte (*Zea mays* subsp. *parviglumis*) pollen. *Journal of Experimental Botany*, **56**, 2401–2407.

Baltazar BM, Sanchez-Gonzalez JD, de la Cruz-Larios L, Schoper JB (2005) Pollination between maize and teosinte: an important determinant of gene flow in Mexico. *Theoretical and Applied Genetics*, **110**, 519–526.

Barrett SCH (1998) The evolution of mating strategies in flowering plants. *Trends in Plant Science*, **3**, 335–341.

Barrett SCH (2010) Understanding plant reproductive diversity. *Philosophical Transactions of the Royal Society B-Biological Sciences*, **365**, 99–109.

Broyles SB, Wyatt R (1990) Paternity analysis in a natural population of *Asclepias exaltata*: multiple paternity, functional gender, and the pollen-donation hypothesis. *Evolution*, **44**, 1454–1468.

Bush EJ, Barrett SCH (1993) Genetics of mine invasions by *Deschampsia cespitosa (Poaceae)*. *Canadian Journal of Botany-Revue Canadienne De Botanique*, **71**, 1336–1348.

David P, Benoit P, Viard F, Castella V, Goudet J (2007) Reliable selfing rate estimates from imperfect population genetic data. *Molecular Ecology*, **16**, 2474–2487.

Davis HG, Taylor CM, Lambrinos JG, Strong DR (2004) Pollen limitation causes an Allee effect in a wind-pollinated invasive grass (*Spartina alterniflora*). *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 13804–13807.

De-Lucas AI, Robledo-Arnuncio JJ, Hidalgo E, Gonzalez-Martinez SC (2008) Mating system and pollen gene flow in Mediterranean maritime pine. *Heredity*, **100**, 390–399.

Dyer RJ, Westfall RD, Sork VL, Smouse PE (2004) Two-generation analysis of pollen flow across a landscape V: a stepwise approach for extracting factors contributing to pollen structure. *Heredity*, **92**, 204–211.

Ennos RA (1985) The mating system and genetic structure in a perennial grass, *Cynosurus cristatus* L. *Heredity*, **55**, 121–126.

Glemin S, Bataillon T (2009) A comparative view of the evolution of grasses under domestication. *New Phytologist*, **183**, 273–290.

Goodwillie C, Kalisz S, Eckert CG (2005) The evolutionary enigma of mixed mating systems in plants: occurrence, theoretical explanations, and empirical evidence. *Annual Review of Ecology Evolution and Systematics*, **36**, 47–79.

Goodwillie C, Sargent RD, Eckert CG *et al.* (2010) Correlated evolution of mating system and floral display traits in flowering plants and its implications for the distribution of mating system variation. *New Phytologist*, **185**, 311–321.

Guillot G, Estoup A, Mortier F, Cosson JF (2005a) A spatial statistical model for landscape genetics. *Genetics*, **170**, 1261–1280.

Guillot G, Mortier F, Estoup A (2005b) GENELAND: a computer package for landscape genetics. *Molecular Ecology Notes*, **5**, 712–715.

Hardy OJ, Gonzalez-Martinez SC, Colas B *et al.* (2004) Fine-scale genetic structure and gene dispersal in *Centaurea corymbosa* (*Asteraceae*). II. Correlated paternity within and among sibships. *Genetics*, **168**, 1601–1614.

van Heerwaarden J, Ross-Ibarra J, Doebley J *et al.* (2010) Fine scale genetic structure in the wild ancestor of maize (*Zea mays* ssp. *parviglumis*). *Molecular Ecology*, **19**, 1162–1173.

Huang ZH, Zhu JM, Mu XJ, Lin JX (2004) Pollen dispersion, pollen viability and pistil receptivity in *Leymus chinensis*. *Annals of Botany*, **93**, 295–301.

Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, **9**, 1322–1332.

Hufford MB (2010) Genetic and ecological approaches to guide conservation of teosinte (*Zea mays* ssp. *parviglumis*), the wild progenitor of maize, PhD thesis, University of California.

Hufford KM, Hamrick JL, Rathbun SL (2009) Male reproductive success at three early life stages in the tropical tree *Platypodium elegans*. *International Journal of Plant Sciences*, **170**, 724–734.

Ireland DS, Wilson DO, Westgate ME, Burris JS, Lauer MJ (2006) Managing reproductive isolation in hybrid seed corn production. *Crop Science*, **46**, 1445–1455.

Jardel EJ, Santana E, Graf S (1996) The Sierra de Manantlán Biosphere Reserve: conservation and regional sustainable development. *Parks*, **6**, 14–22.

Jarne P, Auld JR (2006) Animals mix it up too: the distribution of self-fertilization among hermaphroditic animals. *Evolution*, **60**, 1816–1824.

Jones AG, Small CM, Paczolt KA, Ratterman NL (2010) A practical guide to methods of parentage analysis. *Molecular Ecology Resources*, **10**, 6–30.

Klein EK, Desassis N, Oddou-Muratorio S (2008) Pollen flow in the wildservice tree, *Sorbus torminalis* (L.) Crantz. IV. Whole interindividual variance of male fecundity estimated jointly with the dispersal kernel. *Molecular Ecology*, **17**, 3323–3336.

Larsen AS, Kjaer ED (2009) Pollen mediated gene flow in a native population of *Malus sylvestris* and its implications for contemporary gene conservation management. *Conservation Genetics*, **10**, 1637–1646.

Macdonald SE, Lieffers VJ (1991) Population variation, outcrossing, and colonization of distrubed areas by *Calamagrostis canadensis*: evidence from allozyme analysis. *American Journal of Botany*, **78**, 1123–1129.

Messeguer J, Penas G, Ballester J *et al.* (2006) Pollen-mediated gene flow in maize in real situations of coexistence. *Plant Biotechnology Journal*, **4**, 633–645.

Mimura M, Barbour RC, Potts BM, Vaillancourt RE, Watanabe KN (2009) Comparison of contemporary mating patterns in continuous and fragmented *Eucalyptus globulus* native forests. *Molecular Ecology*, **18**, 4180–4192.

Moeller DA, Tenaillon MI, Tiffin P (2007) Population structure and its effects on patterns of nucleotide polymorphism in teosinte (*Zea mays* ssp. *parviglumis*). *Genetics*, **176**, 1799–1809.

Mondragon-Pichardo J, Vibrans H (2005) Ethnobotany of the Balsas teosinte (*Zea mays* ssp. *parviglumis*). *Maydica*, **50**, 123–128.

Nason JD, Herre EA, Hamrick JL (1996) Paternity analysis of the breeding structure of strangler fig populations: evidence for substantial long-distance wasp dispersal. *Journal of Biogeography*, **23**, 501–512.

Nurminiemi M, Tufto J, Nilsson NO, Rognli OA (1998) Spatial models of pollen dispersal in the forage grass meadow fescue. *Evolutionary Ecology*, **12**, 487–502.

Nybom H (2004) Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. *Molecular Ecology*, **13**, 1143–1155.

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.

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, **6**, 288–295.

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

Ritland K (2002) Extensions of models for the estimation of mating systems using n independent loci. *Heredity*, **88**, 221–228.

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.

Robledo-Arnuncio JJ, Austerlitz F, Smouse PE (2006) A new method of estimating the pollen dispersal curve independently of effective density. *Genetics*, **173**, 1033–1045.

Robledo-Arnuncio JJ, Austerlitz F, Smouse PE (2007) POLDISP: a software package for indirect estimation of contemporary pollen dispersal. *Molecular Ecology Notes*, **7**, 763–766.

Rognli OA, Nilsson NO, Nurminiemi M (2000) Effects of distance and pollen competition on gene flow in the wind-pollinated grass *Festuca pratensis* Huds. *Heredity*, **85**, 550–560.

Saghai-Maroof MA, Soliman KM, Jorgensen RA, Allard RW (1984) Ribosomal DNA spacer-length polymorphisms in barley—Mendelian inheritance, chromosomal location, and population-dynamics. *Proceedings of the National Academy of Sciences of the United States of America*, **81**, 8014–8018.

Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology*, **18**, 233–234.

Smouse PE, Dyer RJ, Westfall RD, Sork VL (2001) Two-generation analysis of pollen flow across a landscape. I. Male gamete heterogeneity among females. *Evolution*, **55**, 260–271.

Sork VL, Davis FW, Smouse PE *et al.* (2002) Pollen movement in declining populations of California Valley oak, *Quercus lobata*: where have all the fathers gone? *Molecular Ecology*, **11**, 1657–1668.

Vogler A, Wettstein-Battig M, Aulinger-Leipner I, Stamp P (2009) The airborne pollen flow of maize (*Zea mays* L.) in a multi-crop designed field plot. *Agricultural and Forest Meteorology*, **149**, 1776–1780.

Wang F, Yuan QH, Shi L *et al.* (2006) A large-scale field study of transgene flow from cultivated rice (*Oryza sativa*) to common wild rice (*O-rufipogon*) and barnyard grass (*Echinochloa crusgalli*). *Plant Biotechnology Journal*, **4**, 667–676.

White GM, Boshier DH, Powell W (2002) Increased pollen flow counteracts fragmentation in a tropical dry forest: an example from *Swietenia humilis* Zuccarini. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 2038–2042.

Wilkes G (2007) Urgent notice to all maize researchers: disappearance and extinction of the last wild teosinte population is more than half completed. A modest proposal for teosinte evolution and conservation in situ: the Balsas, Guerrero, Mexico. *Maydica*, **52**, 49–65.

Supporting information

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

Table S1 Microsatellite markers used for genotyping *parviglumis* samples

Table S2 Matrix of probability of assignment of pollen haplotype groups to pooled samples from each sampling site

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